SMART 2020/694 | D2

Impact Assessment on enhancing the use of data in Europe

Report on Task 1 – Data governance

Internal identification

Contract number: SMART 2020/694
VIGIE number: 2020-0694

SMART 2020/694 | D2

Impact Assessment on enhancing the use of data in Europe

Report on Task 1 – Data governance

Manuscript completed in November 2020

First edition

Contents

1 Introduction

This chapter illustrates the purpose of the document and briefly explains which data collection tools were used to gather the evidence underpinning the findings and conclusions of this assignment.

1.1 Purpose of the document

This Report (D2) is the second interim report of the study to support an Impact Assessment on enhancing the use of data in Europe. The purpose of this report is to report on the **final results of Task 1**.

Task 1 of this assignment comprises four separate sub-tasks:

- **Subtask 1.1** relates to measures facilitating secondary use of sensitive data held by the public sector;
- **Subtask 1.2** relates to establishing a certification scheme for data altruism mechanisms;
- **Subtask 1.3** relates to establishing a European structure for governance aspects of data sharing; and
- **Subtask 1.4** relates to establishing a certification framework for data intermediaries.

The study follows an iterative approach to data collection and analysis and these subtasks will be reported on and improved for the subsequent deliverables.

The document is structured as follows:

- **Chapter 1 : Introduction** on the purpose and the structure of the document
- **Chapter 2** : **Background and problem assessment** for each subtask
- **Chapter 3 : Policy objectives and policy options** for each subtask
- **Chapter 4 : Assessment of the impacts of the policy options** for each subtask
- **Chapter 5 : Comparison of the options** for each subtask
- **Chapter 6 : Assessment of macro-economic impacts**
- **Chapter 7 : Conclusion** for this phase of the study

The document contains the following annexes:

- **Annex A** – Subtask 1.1 | Case studies
- **Annex B** – Subtask 1.4 | One pagers
- **Annex C –** Cost benefit analysis per subtask
- **Annex D** – Macro economic analysis | Top-down
- **Annex E** – Macro economic analysis | Bottom-up

1.2 Scope of this study

The study sustains the development of policy measures concerning the areas defined by the Communication on a European Strategy for data[1], which was published in February and constitutes the mandate for the new Commission. The study focuses in particular on eight key issues which are outlined in the strategy, divided in two tasks. This current D2 report focusses on topics under **Task 1**. Subsequent reports will address the topics under Task 2.

---

[1] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, A European strategy for data, COM/2020/66 final, https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1582551099377&uri=CELEX:52020DC0066

Task 1:

1.1. The question of access and **reuse of sensitive public sector data** which are currently not disclosed by public sector bodies and not covered by the Public Sector Information (PSI)/Open data directive[2] (e.g. health data, statistical microdata, company ownership data, microdata from public transport systems and others)[3].

1.2. The possibility of **establishing "data altruism" schemes** in Europe, defined as means of making data available (whether anonymised or non-anonymised) without expecting anything (not even services) in exchange.

1.3. The question of facilitating data sharing through the establishment of **metadata standards** across or within sectors and including both technical and legal standards.

1.4. The relevance of building a **certification framework** for European data intermediaries or data marketplaces which help data demand and supply to match through independent platforms.

Task 2:

2.1. Aspects related to **Business to Government Data Sharing (B2G) for the public interest** (i.e. for the development of better policies and delivery of better public services).

2.2. Possibilities for **empowering citizens** and putting them even more in control of their data, building on the General Data Protection Regulation[4] and establishing a human centric data economy.

2.3. The question of **rights over co-generated data** (i.e. in the context of connected and Internet of Things devices) for enabling further business to business (B2B) data sharing.

2.4. Aspects related to **conflict of laws at the international** level and possible obstacles for businesses subject to extra-territorial provisions and foreign jurisdictions.

For each of these important aspects, the study explores *the state of play in Europe and determines the impact of a number of possible policy options.*

Concerning the geographical scope, the study focuses on the **27 European Union Member States** but it also covers case studies, examples and literature coming from third countries when relevant (i.e. experiences of B2G data sharing). Furthermore, for specific subtasks (i.e. Subtask 1.2 on data altruism) the data collection and analysis activities focus on a sample of Member States.

From a stakeholder perspective, the study focuses on the relevant stakeholders in the data value chain for each of the topics in scope, meaning on data holders, data intermediaries and data re-users. The table below summarises the main categories of stakeholders involved in the subtasks' data collection and analysis activities conducted as part of Task 1.

---

[2] Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information, https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1561563110433&uri=CELEX:32019L1024

[3] In agreement with the Commission, this study focuses on the former two.

[4] 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), https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679

Table 1 - Stakeholder scope for Task 1 (data value chain mapping)

Subtask	Data holder	Data re-user	Intermediaries	Personal data?	Purpose
1.1	Public sector authorities (e.g. Health institutions, transport authorities, statistical offices)	Public sector authorities, researchers and businesses	Public sector authorities, research organisations, non-for-profit orgs.	Yes (and sensitive)	Research and innovation, public health, increased efficiency
1.2	Public sector authorities, businesses, NGOs and researchers	Public sector bodies, researchers and non-for profit orgs.	Public sector authorities, businesses, research orgs.	Yes (and sensitive)	Research, innovation, public health and other societal benefits
1.3	Businesses in traditional sector	Other businesses and researchers from various sectors	Public and private orgs. in charge of data spaces; standardisation initiatives	N/A	Innovation, competitiveness
1.4 *Generic Approach*	Businesses, Academia and research orgs., Governmental orgs., NGOs, Citizens	Businesses, Academia and research orgs., Governmental orgs., NGOs, Citizens	Certified data Intermediaries (i.e. Data marketplaces, data brokers, data repositories, PIMS/PDS, industrial data platforms, trusted third parties, data unions, data cooperatives, data collaboratives, data trusts)	Potentially	Business, R&I, Public Good
1.4 *B2B Approach*	Businesses	Businesses	Certified data intermediaries: Data marketplaces, industrial data platforms, trusted third parties, data collaboratives, data trusts	No	Business, R&I
1.4 *C2B Approach*	Citizens	Businesses	Certified data intermediaries: PIMS/PDS, data unions, data cooperatives, data collaboratives, data trusts	Yes	Business, R&I, Public Good

1.3   Methodology for the assignment

The evidence supporting this analysis from a number of different sources:

- Stakeholder mapping;
- Interviews;
- Legal analysis (Subtask 1.2);
- Workshops (Subtask 1.1 and 1.3);
- Case studies (Subtask 1.1 and 1.3);
- Market analysis (Subtask 1.4)

The first step of the assignment consisted of **mapping stakeholders** based on desk research to understand the current data economy stakeholders relevant for each of the topics under the four subtasks. This analysis

contributed to the development of the stakeholder map and potential interviewee list for further data collection.

The team carried out **interviews** with public sector entities and private sector organisations to collect primary data for this study. These interviews covered various industry sectors and Member States. The purpose of the interviews was to reach out to European companies and Member States to collect data on the data economy and on the related costs and effects.

Additional research methodologies where used for the various subtasks relevant for their respective topic. For subtask 1.1 and 1.3 **case studies** and **workshops** were conducted, a **legal analysis** for Subtask 1.2 and a **market analysis** for Subtask 1.4.

Workshops were conducted by Subtask 1.1 to measure facilitating secondary use of data the use of which is subject to the rights of others and by Subtask 1.3 for the purpose of reviewing the relevance of establishing a European structure for governance aspects of data sharing with stakeholders. The aim of these workshops was to:

- Discuss the study team's qualitative and quantitative assumptions and findings with regard to the baseline scenario and policy options, and discuss the policy options themselves (Subtask 1.1); and
- Discuss the policy options and validated the consolidated data on costs and benefits (Subtask 1.3.

Case studies were also conducted for in-depth and detailed investigation to understand better the state of play or baseline scenario for the topic at hand. The case studies were developed based on desk research and interviews with stakeholders from the public and private sector. Each case study built on the data coming from the stakeholder mapping and provided insights for the baseline scenarios and make hypothesis on the impact of the different policy options.

The **legal analysis** for Subtask 1.2 was based on desk research and interviews to understand the current legal status with regard to data altruism and what barriers this might cause for data altruism. The analysis supported the definition of the policy options and overall analysis.

In addition, Subtask 1.4 carried out at **market analysis** to understand the business environment and data based value chains as well as to identify the key players and key positions in the markets. The market analysis focused on data collection through desk research, triangulation of data and an analysis of the data to answer questions such as what are the main data intermediaries on the EU market. For the desk research, the study team relied on databases and data marketplaces such as Forrester Research, Gartner Research, IDC, Economist intelligence unit and EMIS Intelligence.

Limitations relating to the findings of this study

As part of this study, evidence was gathered from various sources, including desk research, interviews with businesses and other stakeholders, case studies, market analysis and legal analysis.

The data collection was hampered by the fact that the public and private sector are still relatively new to navigating the data economy and can only share insights into for example costs and benefits to a very limited extent.

This situation poses challenges on the findings of this study. While we were able to collect qualitative feedback from the public and private sector on the different policy interventions discussed for each subtask, it was more difficult to quantify their costs and benefits, e.g. because case numbers are still small or the data sharing practices are just emerging and stakeholders themselves do not yet know their scale and/or costs of making data available. In addition, the stakeholders consulted do not yet have a

---

final and consolidated perception on for example the potential benefits they could draw from increased data use and availabilities in their respective domain, besides speculative thoughts.

This report should be considered as a first attempt at examining this topic and gathering the existing data on these subjects. This analysis is therefore based on the limited data available and provides a preliminary (mainly qualitative) overview of the costs and benefits for the different topics under scrutiny. The conclusions reached are based on independent judgement and specific to this study.

2 Background and problem assessment

This chapter contains the problem assessment of issues related to the secondary use of data held by the public sector and the use of which is subject to the rights of others, data altruism, the governance of data sharing, and the role of data intermediaries.

Market developments and policy initiatives of the past decade have set the ground for a European data strategy that could enable the EU to become the world's most attractive, secure and dynamic data-agile economy, improving the lives of its citizens. Europe's technological and digital future depends on whether it seizes this opportunity. Thus, despite the action that the European Commission has taken so far, remaining issues need to be tackled for Europe to reap fully the data economy's benefits.

Data has started to disrupt European economies and markets. The European data market's value will reach EUR 77.8 billion in 2020 employing 8.25 million people, and the overall value of the data economy grew from EUR 247 billion in 2013 to almost EUR 477.3 billion in 2020, worth about 3.2 per cent of total EU GDP.[5] The European Monitoring Tool further predicts that by 2025, the value of the EU data market could reach EUR 93 to 141.6 billion. Likewise, the EU data economy is expected to increase to EUR 1,053 billion with an overall impact of 6.3% on the EU GDP under a high growth scenario. The data suppliers industry would increase from 255,000 companies in 2016 to 294,350 in 2020, and the number of data workers in Europe would increase up to 8.25 million by 2020.

This market development is in line with the perception of businesses. Economic growth and a higher level of competition and innovation in the EU were the key benefits and opportunities identified with regard to the European data economy and data mobility within the EU.[6] Many companies have recognised the potential of data-driven innovation and started to share and re-use data among them to enhance their business opportunities and improve internal efficiency. According to a 2018 study on Business to Business (B2B) data sharing and re-use[7], this trend would grow significantly in the following five years, while companies not critically investing in data may be missing business opportunities. In addition, a large proportion of SMEs perceive data sharing as important and actively acquire data from other companies.[8] Data also feeds into other new technologies with the potential to foster European economies, such as Artificial Intelligence and the Internet of Things.

Apart from its economic impact, the data transformation will affect European societies and daily lives. The volume of data produced in the world is growing rapidly, from 33 zettabytes in 2018 to an expected 175 zettabytes in 2025, and the amount of data is doubling every 18 months. Data-driven innovation will bring

---

[5] European Data Market Monitoring Tool, consulted on 7 July 2020, see: http://datalandscape.eu/european-data-market-monitoring-tool-2018
[6] European Commission (2017), Synopsis report: consultation on the 'building a European data economy' initiative, https://ec.europa.eu/digital-single-market/en/news/synopsis-report-public-consultation-building-european-data-economy
[7] European Commission (2018), Study on data sharing between companies in Europe, https://op.europa.eu/en/publication-detail/-/publication/8b8776ff-4834-11e8-be1d-01aa75ed71a1/language-en
[8] European Commission (2019), SME panel consultation on B2B data-sharing principles and guidance – Report on the results, https://ec.europa.eu/digital-single-market/en/news/sme-panel-consultation-b2b-data-sharing

substantial benefits for citizens, for instance through personalisation and enhancements in healthcare and well-being, transport, transparent governance, public services, energy consumption, product, material and food traceability, and could even contribute to the successful implementation of the European Green Deal. While identifying these benefits, the European strategy for data[9] emphasises that transformations could put European core values at risk. To ensure an open, fair, diverse, democratic and confident Europe, it needs to accompany the wide flow and use of data with high levels of privacy, security, safety and ethical standards.

2.1 Subtask 1.1: Measures facilitating secondary use of sensitive data held by the public sector

2.1.1 Background

2.1.1.1 Context

In recent years, there has been a growing trend towards open data – making public data available for reuse by the private sector, civil society and academics in order to enable research or help promote the development of new services. The EU Open Data Directive[10] is one manifestation of this global trend, backed by estimates that the opening up of EU public data could drive economic benefits of EUR 250 billion[11].

The Open Data Directive extended the scope of previous legislation, requiring Member States to make unprecedented amounts of data available for reuse. The new rules recognise, however, that some data held by the public sector is not suitable for entirely open access and reuse. Instead a number of exceptions are carved out, including for personal data, "documents for which third parties hold property rights", and other "sensitive data" protected by national legislation on the grounds including national security, statistical confidentiality, and commercial confidentiality.

Any of these types of data can be categorised as sensitive data in that additional steps are required before it is possible to share them publicly. Indeed, this task defined 'sensitive data held by the public sector' as "**data the use of which is subject to the conflicting rights of others**". For these reasons, this type of data is less likely to be made available by public administrations. However, in line with the high estimate of the value of public sector data, there is great potential to drive positive economic and social results through opening up some of this data. Health and social data, for example, would generally be classed as "sensitive" data under this categorisation as it is subject to the rights of patients having co-produced the data, and could be used for ends including to "develop personalised medicine or advance research to find cures for specific diseases".[12] Other data the use of which is subject to the rights of others could include genetic data, and statistical microdata.[13]

At European level, there is an ongoing discourse within the research community on how data the use of which is subject to the rights of others can best be shared and made available for research purposes.

---

[9] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, A European strategy for data, 2020, https://ec.europa.eu/info/sites/info/files/communication-european-strategy-data-19feb2020_en.pdf

[10] DIRECTIVE (EU) 2019/1024 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 20 June 2019 on open data and the re-use of public sector information. Available at https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1561563110433&uri=CELEX:32019L1024

[11] Deloitte, Open Evidence, Wik Consult, timelex, Spark, The Lisbon Council (2018), Study to support the review of Directive 2003/98/EC on the re-use of public sector information. Available at https://ec.europa.eu/digital-single-market/en/news/impact-assessment-support-study-revision-public-sector-information-directive

[12] European Commission (2020), Shaping Europe's digital future – Questions and Answers, Available at https://ec.europa.eu/commission/presscorner/detail/en/qanda_20_264

[13] EOSC-hub (2019), D2.8 First Data policy recommendations, Available at https://documents.egi.eu/public/RetrieveFile?docid=3419&filename=EOSC-hub%20D2.8%20v1%20Approved%20by%20EC%20Public.pdf&version=5

European projects including EUDAT CDI[14] and EOSC-hub[15] have engaged on this issue, producing recommendations[16] on how to enable the secure sharing of this data.

The types of measures considered go well beyond the anonymisation of personal data, for which there are concerns that it will nonetheless be possible to link the data back to the individual concerned.[17] Instead, they focus on points such as:

- Promoting free access to the metadata of data the use of which is subject to the rights of others;
- Providing a "safe haven" - a secure environment for research work on data the use of which is subject to the rights of others;
- The relative merits of central vs distributed storage of data the use of which is subject to the rights of others.

2.1.1.2 Ecosystem

This section outlines the types of stakeholders concerned by the sharing of data held by the public sector and the use of which is subject to the rights of others. The table below provides an overview.

Table 2 - Stakeholder scope (data value chain mapping)

Subtask	Data holder	Data re-user	Intermediaries	Personal data?	Purpose
1.1	Public sector authority (e.g. Health institutions, transport authorities, statistical offices)	Researchers/Public Sector Bodies/ Businesses	Public bodies, research organisations, not for profit organisations, partnerships	X (and sensitive)	Research & Innovation, Public health, increased efficiency

This study focuses on **data holder, data re-users and data intermediaries**. Data (co-) producers in this context are typically citizens and, depending on the case, employees at statistical offices performing surveys or general practitioners. They are unknown to the public and should remain so. There cannot be costs or benefits for (co-) producers as part of this study.

Data holders are defined by the OECD as "*a party who, according to domestic law, is competent to decide about the contents and use of (personal and non-personal) data regardless of whether or not such data are collected, stored, processed or disseminated by that party or by an agent on its behalf*".[18] In the context of this subtask, data holders are necessarily public sector entities, such as statistical offices, health institutions or transport authorities. They hold the data that re-users want to access, and constitute the **supply side** of the sensitive public data market.

Data re-users, which can be defined as "*generating the social and economic value of data sharing*" through their use of the data the use of which is subject to the rights of others,[19] may be public bodies, researchers,

---

[14] EUDAT Collaborative Data Infrastructure. See https://eudat.eu/eudat-cdi

[15] European Open Science Cloud hub – providing support services for the development of a European Open Science Cloud, and a single point of contact for researchers for resources for advanced data-driven research. See https://www.eosc-hub.eu/about-us

[16] EOSC-hub (2019), D2.8 First Data policy recommendations, Available at https://documents.egi.eu/public/RetrieveFile?docid=3419&filename=EOSC-hub%20D2.8%20v1%20Approved%20by%20EC%20Public.pdf&version=5; EUDAT (2018), EUDAT Conference Outputs and Recommendations, Available at https://b2share.eudat.eu/records/31b4347b771641e791991578b6731aa1

[17] European Commission (2018), Synopsis Report - Consultation: Transformation Health and Care in the Digital Single Market. Available at https://ec.europa.eu/health/sites/health/files/ehealth/docs/2018_consultation_dsm_en.pdf

[18] OECD, Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies. See: https://www.oecd.org/sti/enhancing-access-to-and-sharing-of-data-276aaca8-en.htm

[19] OECD, Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies. See: https://www.oecd.org/sti/enhancing-access-to-and-sharing-of-data-276aaca8-en.htm

or businesses, and accordingly may use data the use of which is subject to the rights of others for information, research or commercial purposes. They constitute the **demand side** of the sensitive public data market.

Lastly, **data intermediaries** are public sector entities which primarily "*enable data holders to share their data, so it can be re-used by potential data users*". These intermediaries facilitate the processes (such as data permit applications, or the process of searching for the data holder holding the desired dataset) required for re-users to obtain access to sensitive public data held by data holders.

Both **data holders and data intermediaries perform a public service function**, and therefore may not make pecuniary profits from these functions. However, they may charge for that service in order to cover the costs linked to its provision.

2.1.1.3 Ongoing initiatives

This section outlines a list of initiatives aiming at facilitating the reuse of data held by the public sector and the use of which is subject to the rights of others. It examines in more depth two of these – Finland's Findata and the German *Rat für Sozial- und Wirtschaftsdaten* (RatSWD).

Several data holders that already make data the use of which is subject to the rights of others available for access and re-use have been identified:

- The **United Kingdom's Department for Education** controls the **National pupil database** (NPD) containing over 21 million pupils (data stored in the NPD is never deleted).[20] Data is collected by state-funded education and higher education institutions in England only: other Devolved Administrations operate different systems.
- The UK National Health Service (NHS) **National Services Scotland (NSS) national safe haven service** – allows data from electronic records to be used to support research when it is not practicable to obtain individual patient consent, while protecting patient identity and privacy. It provides **secure** file transfer and submission services to data providers and additional services (e.g. analytics platforms) to researchers. It is currently operated by the Edinburgh Parallel Computing Centre (EPCC, University of Edinburgh), while the Farr Institute of Health Informatics Research provides the infrastructure.[21]
- The **Belgian Federal Public Service (FPS) Mobility and Transport** launched a **Smart Mobility call** to support digital mobility initiatives to improve mobility and to boost the use of open data for mobility.
- The **38 officially accredited *Forschungsdatenzentren* – Research Data centres** – in Germany, such as the Research Data Centre of the German Federal Employment Agency, or the German Microdata Lab.[22] These have as mission to make relevant data available for research on labour market, pensions, unemployment benefits, education, vehicles, migration, copyrighted content, and others, but currently do not cover health (although this will be setup in the near future). Of particular interest is the **German Forschungsdatenzentrum** (Research Data Centre) consisting of two Research Data Centres (one at federal level and one at Länder level) that enable access to official statistics microdata to researchers.[23] This enables data from different regional statistical offices to be centralised in one storage system, thus facilitating scientific analysis. It contains data related to health, income and living conditions, agriculture, education, energy, taxation and other statistics. An online application to access the data must be submitted, and data use is subject to a fee depending on *inter alia* the amount of requested data.

---

[20] Her Majesty's Government, National pupil database. See: https://www.gov.uk/government/collections/national-pupil-database

[21] EPCC, NHS National Services Scotland (NSS) national safe haven). See: https://www.epcc.ed.ac.uk/projects-portfolio/nhs-national-services-scotland-nss-national-safe-haven

[22] The full list may be found on the German Data Forum's website. See: https://www.ratswd.de/en/data-infrastructure/rdc

[23] Forschungsdatenzentrum. See: https://www.forschungsdatenzentrum.de/de

- Many statistical offices across the EU, such as for instance:
  - **Statistics Denmark** is Denmark's central statistical authority collecting, compiling and publishing statistics.[24] Data can be accessed for a fee covering the costs of development and operation. Research institutions may access Statistic Denmark's collection of register data and to anonymised microdata after having made a request to the Division of Research Services. In addition, Statistics Denmark provides customised services combining a range of statistical records, for a fee calculated based on the time spent on the custom request (with a defined hourly fee).
  - **Statbel** is Belgium's statistical office collecting, producing and disseminating figures on the Belgian economy, society and territory.[25] These figures are available at national, regional, provincial, municipal and even more detailed level, as well as within a European context. Apart from the readily available statistics that are often used by policy makers, consumers and businesses, and researchers, pseudonymised study microdata can be made available for public institutions or research institutes, through a standardised procedure in order to comply with the privacy law.
  - The **Czech Statistical Office** (CZSO) is Czechia's central statistical authority collecting, processing, and disseminating statistical information.[26] A wide range of data can be easily accessed free of charge. The use of confidential statistical data (including personal data) can only be used for scientific research purposes and must be officially requested.
  - The **National Institute of Statistics** is the main producer of official statistical data for **Romania**. They are responsible for the coordination of all activities at national level regarding the development and dissemination of European statistics.[27] Direct access to the anonymised microdata is provided by means of research contracts. Access is in principle restricted to universities, research institutes, national statistical institutes, central banks inside the EU and the EEA countries, and the ECB. As regards data intermediaries, the following initiatives were identified:
- The **eHealth platform Belgium** is a Belgian federal government service that offers an electronic platform where all stakeholders involved in public health (businesses, citizens, care providers, institutions…) can exchange information, including personal data, in a safe and efficient manner.[28] It offers a range of other services, such as **MyCareNet**,[29] a platform enabling data exchange between care providers and health insurance providers.
- The UK Office for National Statistic's (ONS) **Secure Research Service** is a facility that enables access to restricted data from surveys and other confidential datasets produced by the ONS to Accredited Researchers.[30]
- **Administrative Data Research UK** (ADR UK) is a partnership between ADR Scotland, ADR Wales, ADR Northern Ireland and the ONS, and which links together and anonymises data held by different public bodies and facilitates access thereto for approved researchers.[31]
- **The Health Data Research Hubs** are centres of excellence in the UK facilitating access to data held by the public sector and the use of which is subject to the rights of others. The Hubs bring together data from NHS hospitals and facilitate access for the public sector, academic and industry research.[32]
- The **French Centre d'accès sécurisé aux données (CASD)** is a partnership between several French universities, research centres and the national statistics institute (INSEE). CASD makes available data from INSEE and from different ministries (including Justice, Education, Agriculture and Economics), from

---

[24] Statistics Denmark. See: https://www.dst.dk/en

[25] Statbel. See: https://statbel.fgov.be/en/

[26] Czech Statistical Office. See: https://www.czso.cz/csu/czso/about-czso

[27] National Institute of Statistics. See: https://insse.ro/cms/en/content/about-nis

[28] Belgian Government, eHealth. See: https://www.belgium.be/fr/services_en_ligne/app_be_health

[29] CIN-NIC, Aperçu général de MyCareNet. See : http://fra.mycarenet.be/algemene-beschrijving

[30] Office for National Statistics, Secure microdata. See:
https://www.ons.gov.uk/census/2011census/2011censusdata/censusmicrodata/securemicrodata

[31] ADR UK. See: https://www.adruk.org/

[32] Health Data Research UK, The Hubs. See: https://www.hdruk.ac.uk/infrastructure/the-hubs/

hospitals, and even from some private companies.[33] It provides this data through SD-Box, a secure infrastructure (or 'bubble') accessible remotely and where confidential data is 'sanctuarised'.

- The **French Health Data Hub** is a database and service provider gathering health data from various databases with a view to facilitating their reuse by research institutions – in full respect of privacy.[34] It also aims at enabling increasingly personalised medicine and a more efficient public system. In addition, it is tasked with promoting standardisation norms regarding health data exchange and reuse, taking into account European and international standards.

- The French **AVIESAN** (Alliance nationale pour les sciences de la vie et de la santé) brought together representatives from the research and health sectors to develop a plan to advance precision medicine and encourage the emergence of a national and industrial sector for genomic medicine. This Genomic Medicine Plan 2025 takes into account technological progress in storage, analysis, and reporting of big data. A National Centre for Intensive Calculation will gather the large volumes of data generated by twelve sequencing services and provide services for health care practitioners. One of the targets is to create a national framework capable of driving scientific and technological innovation and economic growth in numerous fields including big data processing, semantic web and the Internet of Things, medical devices, and eHealth. The consequences of precision medicine on policy will most likely not be negligible.[35]

- The **Nordic eInfrastructure Collaboration** (NeIC) is hosted by NordForsk, which provides for and facilitates Nordic cooperation on research and research infrastructure across the Nordic region. Among their activities, **Tryggve** is a Nordic collaboration for data the use of which is subject to the rights of others funded by NeIC and by research institutes forming ELIXIR nodes of participating countries.[36] Tryggve develops and facilitates access to secure e-infrastructure for data the use of which is subject to the rights of others, suitable for hosting large-scale cross-border biomedical research studies. It develops state-of-the-art scalable infrastructure for safe, efficient, ethical, and legal storage, analysis and sharing of sensitive personal data for biomedical research between countries. The project supports open and transparent data access processes by engaging with the key stakeholders from each of the Nordic countries, and facilitates automated cross-border data exchange among Nordic countries with the ambition to scale this to Europe and beyond.

- The **Leuven Statistics Research centre** (LStat), created in 1988 as an interfaculty institute providing a coordinating link between all university research centres dealing with statistics. Currently, there are around 25 such centres (spread over 11 different faculties) involved in the activities of the LStat.[37]

- The **Edinburgh Parallel Computing Centre** (EPCC), which works closely with NSS and the Farr Institute to extend and enhance the new NHS Scotland safe haven service beyond its current basic computing capability in order to provide traditional High-performance Computing (HPC) services within the safe haven.[38]

In addition, two intermediaries are of particular interest, as they perform functions that are similar to those outlined in the proposed policy options (see section 3.2.1). Annex A provides additional details on these two initiatives.

---

[33] CASD. See: https://www.casd.eu/
[34] Health Data Hub. See: https://www.health-data-hub.fr/

[35] https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(16)32467-9/fulltext
[36] NEIC, Tryggve – Collaboration for Sensitive Data. See: https://neic.no/tryggve/ ; and ELIXIR, ELIXIR Nodes. See: https://elixir-europe.org/about-us/who-we-are/nodes
[37] Leuven Statistics Research Centre. See: https://lstat.kuleuven.be/
[38] EPCC. See: https://www.epcc.ed.ac.uk/

2.1.1.3.1 RatSWD

The German RatSWD is a **public advisory council to the German federal government** and was founded in 2004. The RatSWD aims at sustainably **improving the research data infrastructure** that underlies empirical research and at contributing to the international competitiveness of said research.

It is made up of an **independent body of researchers and representatives of data holders**, and acts as an institution of exchange and of mediation between the interests of science and data producers. As such, it is an important platform for **communication and coordination**.

RatSWD's core tasks are the following:

- To issue **recommendations on further improving the data infrastructure**, specifically:
  - Recommendations on how to secure and further improve data access, particularly by establishing and evaluating research data centres and data service centres according to a set of clear standards;
  - Recommendations on how to improve data use through the provision of scientific and statistical data (research data portal; metadata) and appropriate documentation;
  - Recommendations on research topics and research tasks pertaining to the conceptual development of data infrastructures on the national, European and international level;
  - Recommendations on how to optimise the production and provision of research-relevant data;
- To **advise science and policy**, specifically:
  - Advising the Federal Ministry for Research and the Länder governments on the development of the research-based data infrastructure;
  - Advising public and private data producers;
  - Advising data producers that are institutionally unaffiliated with independent scientific research on how to receive certification as a scientific research institution (certification);
- To **monitor legal and technological developments**, specifically:
  - Monitoring national and legal developments in data provision;
  - Monitoring technological developments, e.g. virtual research environments; and
- To **organise and host the Conference for Social and Economic Data** every three years.

Although RatSWD itself does not make data available to re-users, it is an **intermediary responsible for the accreditation of Germany's Research Data Centres** (RDCs) (see the discussion of the German Forschungsdatenzentrum above), which act as data holders and sometimes also as data re-users for research purposes. It coordinates these RDCs via a **Standing Committee Research Data Infrastructure (FDI Committee)** established in 2009. In addition, the RatSWD accredits RDCs when they meet a number of criteria, and monitors their continued compliance with these. The RatSWD also provides a search engine for datasets held by the 38 accredited RDCs in Germany (with a few exceptions).

2.1.1.3.2 Findata

Findata is Finland's recently established **one-stop-shop** responsible for streamlining and securing the secondary use of social and health data. It guarantees a flourishing ecosystem around the secondary use of social and health data streamlining the processes for the **issuing of research permits** and data collection and ensuring that data is being used in **secure environments**, thereby maintaining the trust that the general public have in authorities and the public sector.

Findata makes retrieving combined health and social data from different sources **easier, faster and possible with just one permit application**, removing the need to approach each authority and data source separately. These applications are not free for the re-user: for a Finnish or EU/EEA-based re-user, a

decision on a data request costs EUR 1,000, excluding an additional data processing fee of EUR 115/hour (for the combining, pre-processing, pseudonymisation and anonymisation of the data).

Findata is also responsible for ensuring the **ethically sustainable use of data**. It makes decisions on data permits concerning data held by other controllers, and is responsible for the collection, combination, pre-processing and disclosure of data for secondary use, in accordance with the Act on Secondary Use of Health and Social Data. Furthermore, the data permit authority maintains a data request management system to forward and process data requests and permit applications.

Findata also maintains a **secure hosting service** for receiving or disclosing personal data and a secure operating environment, in which the permit holder may process the personal data he/she has been disclosed on the basis of data permit. It also supervises compliance with the terms and conditions of the permit it has issued. The data permit may be revoked if the permit holder fails to comply with the law or the terms and conditions of the permit. Lastly, the data permit authority is responsible for the pseudonymisation and the anonymisation of personal data.

2.1.2 **The problem, its magnitude and the stakeholders affected**

Several issues can be identified as part of this subtask.

The first concerns **discovery of, and access to, data**. As stated in the European strategy for data, "*the value of data lies in its use and re-use*". The unavailability in some Member States of certain types of data the use of which is subject to the rights of others for re-use, results in a range of problems such as:

- The inability for re-users to access and use the data in order to conduct research and development that may have positive impacts for society, such as improved public services including public transport, healthcare or education;
- The inability for decision-makers to rely on re-users research outputs as input to their decision-making that could result in overall 'better' policies; and
- Barriers in the development of Artificial Intelligence (AI), which requires data to improve continuously.

The second issue is **data interoperability and quality**. Even when given datasets held by the public sector and the use of which is subject to the rights of others are made available for reuse, research may require combining datasets from different data holders. This is hampered by the insufficient data interoperability among datasets from different sources, which may structure their datasets differently – resulting in additional time spent by re-users on combining different datasets rather than on conducting research for the public good. This situation necessarily results in **fragmentation** as regards access to, and combination of, data of sufficient quality, and in an **imbalance** between large re-users (such as multinational pharmaceutical companies) that have the resources to perform this work, and smaller re-users (such as SMEs or some researchers) which do not always have that capacity. This imbalance is reinforced by the growing costs associated with re-use of the data, deterring smaller reusers.

An additional issue regards the **ability of citizens to exercise their rights**, notably under the General Data Protection Regulation, in the absence of transparent and streamlined processes to do so.

When examining the issue's **European dimension, the situation is even more fragmented**, as some Member States have taken steps to facilitate the re-use of data held by the public sector and the use of which is subject to the rights of others (such as establishing one-stop shops or single data permit authorities, as well as cross-border data re-use mechanisms) whereas others have not. This hampers research at a European scale – a scale that would enable both higher quality research, and increased AI development, due to larger datasets being available and interoperable.

While the benefits of reusing data the use of which is subject to the rights of others cannot be accurately estimated, data access and reuse of public sector data (including non-sensitive data) is estimated to bring social and economic benefits equivalent to between 0.1% and 1.5% of GDP, according to the OECD.[39] This therefore impacts society as a whole: data re-users, data holders, data intermediaries and citizens (both in their role as co-producers, and also as members of society more broadly).

Two broad categories of data holders can be differentiated: statistical offices and health- and social-related data holders. As regards statistical offices (and other public authorities responsible for the development, production and dissemination of statistics), the European Statistical System keeps an up-to-date list that currently contains 286 entities,[40] of which 27 are related to health (and therefore excluded from this count to avoid double-counting). As a result, the amount of **data holders in the EU27 when it comes to statistical microdata can be estimated to be around 260**.

As regards **health- and social-related data**, several broad types of data holders can be identified,[41] namely an estimated:

- 55 Ministries of Health (typically one per Member State, except in federal countries, namely Austria with 9 länder, Belgium with three federated governments, and Germany with 16 länder);
- 55 Ministries of Social Affairs (typically one per Member State, except in federal countries);
- 104 Medical insurance authorities (of which it can be assumed there is on average one per each 104 NUTS-1 region);
- 55 Pensions Authorities (typically one per Member State, except in federal countries);
- 104 Social Security Authorities (of which it can be assumed there is on average one per NUTS-1 region);
- 27 Medicine Authorities (typically one per Member State);
- 27 Population registers (typically one per Member State); and
- 104 Hospitals Authorities (of which it can be assumed there is on average one per NUTS-1 region)[42].

Thus, there are roughly **530 data holders in the health and social domains**.

In addition, there are a number of cross-sectoral data holders: public universities and research centres that one the one hand re-use data the use of which is subject to the rights of others (on which see below), but on the other also produce such data (e.g. when conducting surveys). To estimate the number of such public entities, the study team reused Eurostat's list of recognised research entities listing 666 recognised entities across the EU27.

In total therefore, there are around **1,500 data holders** in total impacted.

With regards to data re-users, an estimated number of re-users of statistical microdata can be derived from Eurostat's list of recognised research entities, which lists a total of **666 recognised research entities** in the EU27. The total number of data re-users for health and social data overlaps with the research entities recognised by Eurostat: 48 of these conduct research in *inter alia* social sciences, while 22 conduct research

---

[39] OECD, Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies. See: https://www.oecd.org/sti/enhancing-access-to-and-sharing-of-data-276aaca8-en.htm
[40] European Commission, List of National statistical institutes (NSI) and other national authorities. See: https://ec.europa.eu/eurostat/documents/747709/753176/20190607_List_other_national_statistical_authorities_IT.pdf/f3c3bddf-c378-4203-92a2-48d0dd789f3d
[41] This identification is an extrapolation based on the different types of data holders concerned by Findata, on the different public sector partners of the French Health Data Hub, and on the health and social affairs data holders listed in the Centre d'accès sécurisé aux données. It excludes statistical offices to avoid double-counting.
[42] Individual hospitals and doctors are considered to be data co-producers rather than data holders.

in a health-related domain. However, it also includes a large number of private companies – that number is estimated to be **147,000 companies**.[43]

Thus, there are roughly **150,000 data re-users** impacted in total.

Lastly, two public data intermediaries can be reasonably assumed to exist in each Member State (except those with a federal structure) – one for health and social data, and another for statistical microdata – this is a total of **110 public data intermediaries** in total.[44]

The table below summarises these figures:

Table 3 - Estimated number of stakeholders impacted in the EU-27

Stakeholder type	Health and social	Statistics	Total
Data holders (public sector)	55 Ministries of Health 55 Ministries of Social Affairs 104 Medical insurance authorities 55 Pensions Authorities 104 Social Security Authorities 27 Medicine Authorities 27 Population registers 104 Hospitals Authorities	260 Statistical offices	**Approx. 1,500**
	666 recognised research entities		
Data re-users (private sector and academia)	70 recognised research entities (overlapping with statistics) 147 000 private companies	666 recognised research entities	**Approx. 150,000**
Data intermediaries (public sector)	55	55	**Approx. 110**

These numbers are **non-exhaustive**, as there are many additional stakeholders in sectors other than health and statistics, such as mobility, business registers or financial reporting databases.

2.1.3 The causes of the problem

This situation is a result of several causes.

In many Member States, there is **uncertainty regarding applicable rules and legislation** – both to provide access to data the use of which is subject to the rights of others, and to access it. For instance, it remains unclear in some Member States whether national ministries have a legal obligation to make the data (the use of which is subject to the rights of others) they hold available for re-use.[45]

Second, data the use of which is subject to the rights of others cover a wide **range of different types of datasets** – ranging from health data to statistical microdata covering a variety of topics such as household

---

[43] This estimation was reached using a) the number of people employed in the healthcare industry (800,000 in 2012 in the EU, see https://ec.europa.eu/growth/sectors/healthcare_en#:~:text=A%20vibrant%20EU%20pharmaceutical%20sector,the%20EU's%20total%20manufacturing%20workforce.); b) the number of active businesses in the EU (27.5 million in 2017), and c) the number of employed persons in the EU (150 million persons in 2017), see https://ec.europa.eu/eurostat/statistics-explained/index.php/Business_demography_statistics. These figures were used to reach an average number of employees per active business (150,000,000/27 500 00=5.45); from which the number of healthcare businesses was derived (800,000/5.45=146,788.99) and rounded-up.

[44] Indeed, it is unlikely that a given Member State would have more than one public data intermediary for the same domain, since the reason behind their existence is to streamline procedures.

[45] As was revealed during a stakeholder interview.

composition. For example, the French CASD provides access to data categorised in a large number of themes, such as agriculture, businesses characteristics, unemployment, household consumption, housing, life expectancy, living standards, health, immigration, and others. [46]

Lastly, **different methods to enable the re-use of data the use of which is subject to the rights of others** co-exist, as detailed above.

2.1.4   The effects of the problem

The current situation results in a range of impacts. First, researchers must spend time and resources finding who holds the datasets they seek, discovering and understanding any specific procedure to request these datasets, filling in several separate applications to access the datasets, and potentially curating the data in order to enable their combination. This may **deter researchers with limited resources** from going through the process altogether, while researchers who decide to go through the process may see their application rejected – or one of their applications rejected (when they apply to more than one data holder). This makes it **more difficult** for them **to conduct their research**.

Similarly, businesses have to navigate through the same issues. While large companies may have the resources to do so, SMEs do not always have such resources, resulting in an **unequal access to data the use of which is subject to the rights of others and therefore reduced innovation and business opportunities** when such access is limited. This impact is **cumulative**, since in effect larger companies are in a better position than small ones to innovate and to develop new products and services.

Data holders that do not allow access to the data the use of which is subject to the rights of others they hold may have **fewer incentives to ensure the data's high quality**. In addition, they are harmed by the abovementioned hindrances to research, as this research could be used as input for **better-informed policy-making**. This has negative impacts across the board for society, health, and the environment.

2.2   Subtask 1.2: Establishing a certification/authorisation scheme for data altruism mechanisms

2.2.1   Background

2.2.1.1   Context

Building up "databases required for the development of artificial intelligence geared towards public service missions" will be a key development in the near future, as predicted in the recent French Parliamentary Mission Report on Artificial intelligence. [47] Through such databases, citizens or other stakeholders could choose to allow their data to be used for the public benefit.

As of today, data sharing is rapidly rising and the value of the EU27 data economy is already EUR 301 billion or 2.4% of the EU GDP. The predicted volume of data that will be shared will increase fivefold to 175 zettabytes in 2025 as compared to 2018[48]. This data includes personal data held by individuals, private sector and civil society organisations that can greatly benefit society. When focusing on the greater good of society, as opposed to the economic benefit of individual stakeholders, data altruism schemes are an interesting mechanism to consider for the public sector. Data altruism is, to date, a small part of data

---

[46] CASD, Les sources de données déjà disponibles au CASD. See: https://www.casd.eu/les-sources-de-donnees-disponibles-au-casd/
[47] Villani (2018), FOR A MEANINGFUL ARTIFICIAL INTELLIGENCE TOWARDS A FRENCH AND EUROPEAN STRATEGY. Available at https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf
[48] The European Data Strategy. Shaping Europe's Digital Future Factsheet. February 2020. Available at https://ec.europa.eu/commission/presscorner/detail/en/fs_20_283

sharing however one that gained increased attention in the previous months, during the COVID-19 pandemic, in the form of data altruism for public health reasons.

While there is an overall willingness to share for example personal data for the public good[49], to date, wide-spread-data altruism scheme does not exist, even though a shared approach on this topic could provide large gains for society. A data altruism scheme is understood as digital data sharing for public benefit. Where data altruism decisions are based purely on the consent of individuals 'donating' their own personal data, data altruism schemes must allow data subjects to revoke their consent for data processing according to the General Data Protection Regulation (GDPR). An example of this revocation right can be found in the Corona-Datenspende App[50]. It needs to be stressed that 'donating' does not mean that the data holder/owner loses rights to their data, instead access to this data is provided and should, under GDPR, always be able to be revoked.

A key challenge is determining exactly when a data sharing decision is driven by altruism, as opposed to being an economic decision. A potential indicator of the presence of an altruistic motivation is to examine the circumstances of the data sharing: if the data producer receives a direct benefit (such as a new or improved product or service) as a result of sharing data, their motivations are less likely to qualify as altruistic even if a social good also materialises. Note that data altruism schemes are generally understood to comprise data infrastructures or mechanisms that will benefit the greater society as a public good[51] and not for economic benefit of individual stakeholders. A data altruism scheme can take a number of different forms, depending on how it is set-up which again depends on various factors such as policy, legal, technology and organisational.[52,53]

Projects have begun in a number of European countries to explore different possibilities to enable data altruism. In Finland, for example, the MyData project aims to enable a 'paradigm shift in personal data management and processing that seeks to transform the current organization centric system to a human centric system'[54]. This conceptual approach goes well beyond the topic of data altruism, presenting a holistic approach to personal data management and developing a discussion framework that can host many implementations and models, including projects. However, one of the main ways that the infrastructure is described as creating value is 'as a common framework for different kinds of research data banks to easily acquire consumers' consent to collect their data"[55].

Similarly, the OwnYourData project has developed a Semantic Container for Data Mobility, supported by Horizon 2020 funding. The semantic container enables "secure and traceable data exchange between multiple parties"[56], with one of the use-cases named as "data donation" to support studies and research. Another example includes the Valencia.Data project in Spain which maintains a database of people who have chosen to make their personal data available through this project, together with a platform for the management of this data. The purpose of the project is to promote the reuse of data for research. This project was funded by the region of Valencia, public resources, and is still running various projects at the

---

[49] Halvorson, Permanente, and Novelli (2014), Data Altruism: Honoring Patients' Expectations for Continuous Learning. Available at https://nam.edu/wp-content/uploads/2015/06/dataaltruism.pdf

[50] Corona-Datenspende App FAQ. Robert Koch Institute, https://corona-datenspende.de/faq/

[51] Skatova A, Goulding J (2019) Psychology of personal data donation. PLoS ONE 14(11): e0224240. https://doi.org/10.1371/journal.pone.022424

[52] Kirkpatrick, R. A New Type of Philanthropy: Donating Data (2013). Harvard Business Review

[53] High-Level Expert Group on Business-to-Government Data Sharing (2020), Towards a European strategy on business-to-government data sharing for the public interest.

[54] Poikola, Kuikkaniemi, Honko (2014), MyData – A Nordic Model for human-centered personal data management and processing. Available at http://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/78439/MyData-nordic-model.pdf?sequence=1&isAllowed=y

[55] Poikola, Kuikkaniemi, Honko (2014), MyData – A Nordic Model for human-centered personal data management and processing. Available at http://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/78439/MyData-nordic-model.pdf?sequence=1&isAllowed=y

[56] Ownyourdata, SEMANTIC CONTAINER FOR DATA MOBILITY. See https://www.ownyourdata.eu/en/semcon/

Instituto de Biomecanica de Valencia. The institute runs several research and development projects focused on data altruism with a focus on the public good. These include projects to for example reduce the amount of textile waste in the textile industry by asking data holders, citizens, to donate physical data to improve clothing seizing. While it sounds commercial, this is a project to utilize data sharing to reduce environmental damages from the textile industry. To date, up to 4000 data holders have shared their data for various projects, however the institute must request consent from every data holder for every new project to reuse data. The data holder then has the opportunity to approve or decline consent for the various projects. All projects are funded by the Valencia government and the institute had built every project infrastructure for data sharing independently, for which no exact price estimation could be provided. While the institute finds that data holders are willing to share data, this strongly depends on the time they must make available to share data and what they will receive in return (research results). The less time is required to share data and the clearer they will know what they can except in return, the more wiling they are to share data and to do so repeatedly for future projects.

A current approach of a data altruism ecosystem is the Corona-Datenspende-App in Germany. For the public interest, specifically to safeguard health and to prevent new pandemic outbreaks, the German Robert Koch-Institute (RKI) is collecting via this app health data related to COVID-19 of Corona-Donation-App-users. With a scientific evaluation of the donated data that was made available, the RKI project team calculates a 'fever map' with the societal purpose to detect early possible hotspots.[57] The app and its use is voluntary and the data are pseudonymised. Data collection and data processing are subject to strict data protection guidelines.[58] Through a scientific blog (www.corona-datenspende.de) providing relevant and further information about the app and the processing of the data, RKI aims to enhance transparency and reliability.

SelfData Territorial is a project of Fing (Next Generation Internet Foundation), a French organization which aims that individuals take control of their personal data, rather than tacitly allowing it to be exploited by others, in order to recognize their right to use (not own) their data.[59] The concept of Self Data means that people are empowered to reuse their personal data themselves for their own purposes. This also includes their ability to access their personal data in a reusable format, and that they are equipped with tools and services to do so.[60] According to Fing, the hoster of MesInfo, SelfData complies with existing data protection regulations in France and complements it. Currently, three French regions, Nantes Métropole, La Rochelle, and Grand Lyon are participating in the SelfData Territorial project.

There are therefore multiple examples of data altruism schemes and infrastructures that could be applied to this purpose and a potential high level of demand for them from both individuals who want to make their data available under specific constraints, and research and other organisations, which want to re-use and analyse it.

2.2.1.2 Ecosystem

A data altruism ecosystem consists of various stakeholders: the data subjects, intermediary, data holder and data re-user. The four stakeholders are described, visualized and explained in a Table 1 below.

The ecosystem stakeholders:

---

[57] Robert Koch-Institut, Blog zur wissenschaftlichen Auswertung der Corona-Datenspende, 2020, https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Corona-Datenspende.html (13.05.2020).
[58] See previous footnote.
[59] FING, SelfData Territorial, 2020. See http://mesinfos.fing.org/self-data-territorial/.
[60] FING (The Next Generation Internet Foundation), Understanding Self Data, 2017. See http://mesinfos.fing.org/wp-content/uploads/2017/08/selfdata_FAQ_mydata2017.pdf.

Data subjects are individuals or organisations[61] that actively share their data for an altruistic purpose. The exact level of detail on the purpose of such data sharing can, but must not always be known once sharing the data. Such purposes of data sharing include research and development, public health, public interest, matching of and synergising cross-sectoral data, information about public administrations and regarding the society, economy or the environment, transparency, or improvement of the access to public services. Data sharing must always be compliant with the GDPR regulation and allow the data sharer to revoke the data sharing rights.

Intermediaries are organisations or institutions which act as a link between data subjects and data re-users in order to transmit the data to the latter. Intermittent stakeholders are e.g. patient associations and health insurance schemes which collect patient data, or research organisations which collect personal data for research and study purposes.

A **data holder** is an entity such as public sector bodies, companies or organisations that manages, hosts and provides the shared data. Depending on the specific data scheme, data holders have respective technical, organisational and/or governance functions in order to make the data accessible to interested re-users by the request of the data subjects.

A **data re-user** is an entity, usually a research body, that re-uses the shared data to create new knowledge insights that contribute to the benefits of the society. Data re-users can be researchers, public sector bodies or non-governmental organisations. The re-use of shared data should be solely for public benefit and not for private economic gains.

A data altruism ecosystem can be approached from the perspective of a public body or private body. Both approaches are visualized below and mapped in the ecosystem mapping in Table 1.

As an example, to illustrate that stakeholders can have multiple roles in the ecosystem, the Corona-Datenspende App ecosystem is also mapped in Table 1 as well as a Business-to-Government example from the maritime sector:

Table 4 – Data altruism scheme ecoystem

Example	Data subjects	Data holder	Intermediaries	Data re-user	Type of data	Purpose
Public sector data holder	-Citizens -Companies	-Public sector bodies -NGOs	-Public sector bodies -Organisations -Companies (hosting a platform/scheme for data altruism)	-Public sector bodies -Researchers (not for economic gains) -Organisations	-Personal sensitive data -Personal non-sensitive data -Other data	-R&D -Public benefit
Private sector data holder	-Citizens -Companies	-Companies	-Public sector bodies -Organisations -Companies (hosting a platform/scheme for data altruism)	-Public sector bodies -Researchers (not for economic gains)	-	-R&D -Public benefit

---

[61] Note that the data subject for the purposes of this note is therefore not necessarily a data subject in the sense of the GDPR, since a data subject under the GDPR must be an individual natural person, not an organisation or other form of legal entity.

Corona-Datenspende App	*-Citizens*	*-Public sector bodies (Robert-Koch-Institute)*	*-Robert-Koch-Institute*	*-Researchers (project team of the Robert-Koch-Institute)*	*-Personal sensitive data*	*-R&D -Public health (prevent new pandemic outbreaks) -Societal benefits*
Vessel Traffic Data PoC	*-Vessels*	*-Statistics Netherlands*	*-MarineTraffic*	*-Statistics Netherlands*	*-Other data*	*-R&D -Maritime (improve statistics for maritime related policymaking)[62]*

2.2.1.3 Ongoing initiatives/Market analysis

This section outlines a list of initiatives aiming at facilitating data altruism schemes and active data altruism schemes in the European Union. The focus in on data intermediaries and data reusers, in particular public research institutes and NGO's. Due to the relative recent development of data altruism schemes, there is a limited amount of examples, nevertheless it has to be noted that the current COVID-19 pandemic has led to an acceleration of national health related data altruism schemes. Germany and Italy have both developed data altruism schemes in the form of contact tracing applications for smartphone and will be presented in-depth.

Data altruism scheme- Data intermediary and data re-user

It has to be noted, as explained in the ecosystem, that the intermediary and data re-user can often be the same actor considering public bodies that manage the application (intermediary) are also reusing the made available data.

- The Instituto de Biomecanica de Valencia (IBV), part of the Universidad Politecnica de Valencia in Spain, is a technology research center that ran the **Valencia.Data** project from 2018-2019[63]. As a data reuser, the IBV established a data altruism scheme by creating an app that enables citizens to share personal data, such as anatomical-physiological data, for research purposes. The IBV is an intermediary and data re-user in this instance.
- The **MESINFOS Project** was a project, run by the City of Lyon, aiming to empower citizens to reuse and share their energy consumption data[64].

2.2.1.3.1 The German Corona-App approach

At the beginning of April, the German Robert Koch-Institut (RKI) released the **Corona Data Donation App** (in German: Corona-Datenspende-App) which aims to contribute information about the spreading of SARS-CoV-2 in Germany. With a scientific evaluation of data, the project team calculates a "fever map" in order to detect early possible hotspots.[65]

The app and its applications is voluntary and anonymized. Currently (05.05.2020), 509,532 persons have registered with a total amount of 15,259,595 data sets made available.[66] The data covers information about

---

[62] Towards a European strategy on business-to-government data sharing report. European Commission (2019)

[63] VLC.Data. Instituto de Biomecanica de Valencia. https://www.ibv.org/blog/proyecto/valenciadata-ecosistema-digital-centrado-en-las-personas/

[64] MESINFOS http://mesinfos.fing.org/english/

[65] Robert Koch-Institut, Blog zur wissenschaftlichen Auswertung der Corona-Datenspende, 2020, https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Corona-Datenspende.html (13.05.2020).

[66] Corona-Datenspende Blog, Robert Koch-Institut, Der Corona-Datenspende Blog, 2020, https://corona-datenspende.de/science/ (13.05.2020).

sex, age in a 5-year-interval, size in 5-cm-intervals, health status and activity data regarding sleeping behavior, heart frequency and temperature, and the regional code.

The algorithms in the app can recognize symptoms, which are correlated with an infection with the Coronavirus.[67] The purpose of the data is exclusively of scientific nature. After a careful analysis, the data contribute to the visualization of a map, which shows the regional distribution of infected persons up to a local level. The RKI has at no time access to personal information such as name or address of the app users. Data collection and data processing are subject to strict data protection guidelines.[68] These standards were verified before launching the app. Persons who choose to make their data available have the possibility to access, administer and delete all of their personal data. The app was edited by the RKI in collaboration with Thryve (mHealth Pioneers GmbH), a digital health company. Through a scientific blog (www.corona-datenspende.de) providing relevant and further information about the app and the processing of the data, RKI aims to enhance transparency and reliability.

The release of the Corona Donation App has been accompanied by an intense public discussion about where to store the data and who is ultimately controlling the data that was made available:[69] The two diverging approaches are either a decentralized or a centralized storage of data. Whereas within a decentralized model, a user in case of an infection only sends its own IDs to the server, in the central version the app sends additionally the codes of all contacted persons to the server. This information is sensitive and needs to be protected with special standards. The current version of the app follow the decentral model.

In addition, the Federal Government of Germany launched the **Corona-Warn-App** to "*help fight the coronavirus*"[70] on the 16th of June 2020. This contact tracing app aims to notify [the app] users when they have been in contact with an infected person which could result in a risk of catching the virus. Thereby, the government hopes to interrupt the chains of infection and reduce the spread of the virus in Germany. The application is a form of data altruism because it helps to break chains of infection and provides valuable data insights for the health ministry, and user of the app, who could not track everyone's interaction of the past 14 days. This is also an example that data altruism exists is many different variations that can help the greater public good, including limiting the further spread of a pandemic.

The app operates by relying on Bluetooth technology, which must be active on a user's smartphone, to measure the distance and duration of contact between people that have installed the app. The app is available for free to download on IOS and Android devices and serviced by the Federal Government. The technology is enabled to exchange temporary encrypted random IDs but does not allow connections to be made to a user's identity or location. When a user tests positive for the virus, they can voluntarily inform other users by notifying in the app that they have tested positive. The app then checks who has been in contact with the infected person and notifies them through a warning in the app. To conclude, the Federal Government is the data re-user, the application user the data holder and the application the intermediary[71].

Note, that data security and protection is of high importance in Germany and the Federal Government ensures users that the app meets European and German data protection requirements. To do so, the Federal Commissioner for Data Protection and Freedom of Information (BfDI) and the Federal Office for Information

---

[67] Robert Koch-Institut, Corona-Datenspende-App. Hände waschen, Abstand halten, Daten spenden – Ihr Beitrag gegen Corona, 2020, https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Corona-Datenspende-allgemein.html (13.05.2020).
[68] See previous footnote.
[69] Tagesschau, Corona-Tracing-App. Was heißt zentralisiert oder dezentral?, 2020, https://www.tagesschau.de/inland/coronavirus-app-101.html (13.05.2020).
[70] The Coronavius warning app. The Federal Government of Germany (2020): https://www.bundesregierung.de/breg-de/themen/corona-warn-app/corona-warn-app-englisch
[71] Ibid.

Security (BSI) were involved in the development of the Corona-Warn-App[72]. Users remain anonymous, they do not have to provide e.g. their personal data when they register to use the app, and all data is **encrypted** and **stored exclusively on your own smartphone**. This is significant because thereby the **data holder** provides access to data but does not share or donate the data itself with the data reuser.

To date, Friday the 26th of June 2020, the app has been downloaded by 13.3 Million users[73] and the Federal Government is investigating cross-border collaboration with Switzerland, the Netherlands and France to establish interoperability across national borders to fight the virus.

2.2.1.3.2   The Italian Immuni App

At the beginning of June 2020, "Immuni", the Corona App of Italy was released and has to date (26.06.2020) already been downloaded over 3 million times. As other Corona-apps, Immuni aims to contribute to the containment of SARS-CoV-2 and thereby to prevent potential outbreaks of the virus.[74] The functions are not yet activated nationwide, which is expected to happen in the upcoming weeks. As in Germany, the app is based on a source code which is openly available and based on contract tracking via Bluetooth Low Energy technology. The data is stored in a decentralized way; and personal data, such as name, date of birth, mobile phone number, identity of contact person or location is not asked for.[75] Anyone with close contact with a user who has tested positive for COVID-19 will receive a warning from the app regarding a potential risk of infection. In order to increase the number of potential users, altruism was used as a strategy to improve the trust in and acceptance of the app, according to Paolo de Rosa, Chief Technology Officer at the Ministry for Innovation Technology and Digital Transformation.

Data protection was key in developing the Immuni App. Data holder is the Ministry of Health in Italy and other public institutions and are stored locally at servers in Italy. Data and connection of the app to the server are protected.[76] The completely transparent approach was chosen by the government because it should create, according to de Rosa, trust in the app. Before the release of the app, these trust and reputation problems were addressed with communication campaigns and awareness raising. A specific fast track law, which is also GDPR compliant, was made for the Immuni app.[77]

The app was developed in a public-private partnership: the mixed stakeholder-combination consisted of one IT-company and, publicly, people from the academia and ministries. The cost of development, the provision of infrastructure, legal, implementation and other costs is estimated to 10 mio. EUR. However, everyone involved worked pro bono, so no real costs have arisen.[78]

In the meantime more Member States have, or are, developing and releasing Corona tracing applications.

2.2.2   **The problem, its magnitude and the stakeholders affected**

The European Union's aim to create a single market for data, to ensure Europe's global competitiveness and data sovereignty, and to create a data-driven society. This includes data sharing initiatives for the public good, such as in case of the healthcare sector as has been argued for especially during the current COVID-19 pandemic. To date, there is no European, or national, data sharing initiatives that enable data holders,

---

[72] Ibid.
[73] Anzahl der Downloads der Corona-Warn-App…im Juni 2020. Statista (26.06.2020)
https://de.statista.com/statistik/daten/studie/1125951/umfrage/downloads-der-corona-warn-app/
[74] Immuni ist ein weiteres Instrument im Kampf gegen die Pandemie. Presidenza del Consiglio die Ministri (2020):
https://www.immuni.italia.it/?gclid=EAIaIQobChMIso7XkMCf6gIVwZAYCh3c0QDpEAAYASAAEgJwLvD_BwE
[75] Ibid.
[76] Ibid.
[77] Paolo De Rosa, Chief Technology Officer, Ministry for innovation Technology and Digital Transformation, Italy in an interview on the 16th of June 2020.
[78] Ibd.

whether private or public, to share data for the public good. There are initiatives, especially for the healthcare sector, however, these are nationally bound.

Data altruism has been highlighted by the German Ethics Council since 2017, which addressed the importance and value of data altruism, however also a foundational problem - the definition of data altruism, how to handle and or grant consent for it and the subsequent reuse of shared data[79]. This is a legal debate to ensure adequate data protection, however the German Ministry of Health found that 79% of German would be willing to share their data for research purposes, highlighting a willingness to participate if such a data altruism mechanism would be in place. This data is from 2019, before the COVID-19 pandemic and the current public willingness to share data appears to be less when looking at the amount of participants that have downloaded the German COVID-19 mobile phone applications.

This highlights a problem within the problem. The first problem is the absence of data altruism mechanisms in the European Union, the second problem is the stakeholders that are essential to a functioning data altruism mechanism such as the data holders, often citizens or private companies, that need to trust the mechanisms to share data with the data re-users.

2.2.3 The causes of the problem

One of many reasons why data altruism schemes are created is to resolve, or at least to mitigate, a multitude of legal challenges which can be linked to data sharing. When done by individual persons on a voluntary basis, data altruism will almost inevitably involve the processing of personal data as defined under the GDPR, both through the initial transfer of personal data to the data re-users, and through any subsequent use of the data thereafter.

As a result, the requirements of European data protection law must be observed, including those included in the GDPR. This implies first and foremost that any re-use of the personal data made available must have a clear legal basis. In a typical data altruism scheme, a person will provide their consent to re-use their data, thereby providing a suitable legal basis.

However, reliance on consent is not as trivial as it seems: in order to be legally valid, consent must be freely given, specific, informed and unambiguous. Freely given implies that there may be no element of coercion in the consent, e.g. because consent for re-use must be provided in order to be able to gain access to a product or service. This also implies that the consent cannot be given when there is a relationship of authority between the data subject and the recipient, such as e.g. between an employee and an employer, or between a student and teacher.

More importantly, consent must be specific and informed, which means that the purpose of re-use must be described in a way that allows the data subject to understand at a sufficient level of detail what their data will be used for, and what the potential implications are. A generic description – e.g. making data available "for the public good" or "for scientific research" is in principle considered too generic to meet this requirement[80]. However, there is some flexibility on this point for scientific research, in cases where the purposes for data processing within a scientific research project cannot be specified at the outset. In those

---

[79] "Datenspende"- Bedarf fuer die Forschung, ethische Bewertung, rechtliche, informationstechnologische und organisatiorishce Ramenbedingungen. Bundeministerium fuer Gesundheit. March 2020; https://www.bundesgesundheitsministerium.de/fileadmin/Dateien/5_Publikationen/Ministerium/Berichte/Gutachten_Datenspende.pdf

[80] Recital 33 of the GDPR notes that "It is often not possible to fully identify the purpose of personal data processing for scientific research purposes at the time of data collection. Therefore, data subjects should be allowed to give their consent to certain areas of scientific research when in keeping with recognised ethical standards for scientific research. Data subjects should have the opportunity to give their consent only to certain areas of research or parts of research projects to the extent allowed by the intended purpose."

cases, recital 33 to the GDPR allows the purpose to be described at a more general level[81]. As a result, attempts have been undertaken – e.g. by the Medizininformatik Initiative[82] - to provide a more generic and broad consent template for scientific research, which have been approved by the Conference of Independent Data Protection Commissioners of the Federal Government and the German federal states. A scaled-up version of this initiative at the EU level could possibly provide greater legal certainty on the possibilities and constraints of such consent forms. Finally, consent must also be unambiguous, i.e. expressed through an affirmative action; consent cannot be deduced from circumstances, or induced by e.g. pre-ticking consent boxes and relying on the data subject's inaction.

Reliance on consent, while clearly a key way to ensure that re-use in data altruism is lawful, also has the limitation that it is not possible for children, that a person can only consent to the use of their own data (which precludes e.g. cases where the data describes interactions with other persons, since their personal data would then be made available as well), and that the consent must be revocable – implying that platforms relying on consent must allow persons engaging in altruism to essentially change their minds and put a stop to the use of their data. This has infrastructural implications as well, since consent management must be built into data altruism platforms, along with sufficient controls to allow future processing to cease after the revocation of consent.

Data altruism schemes can be a part of the answer to these problems, not only by offering a shared platform for consent management (essentially through a specialised form of Personal Information Management Systems (PIMS)[83], but also by streamlining the process of informing users appropriately about new re-use cases, and by building a governance mechanism on top of the altruism use cases that enables a degree of control over future re-use, and notably whether the re-use complies with the information provided when the consent was obtained. In this case, the intermediaries in data altruism can play the role of supervisors and enforcers of the scheme, at least to some extent, and notwithstanding the protections afforded by the GDPR.

The legal basis is not the only challenging factor in data altruism. The GDPR also is based on the purpose limitation principle, implying that the purpose of the data processing must be defined upfront, the data subject must be informed of the use that will be made of the data, and that the data may thereafter not be used in a manner which is incompatible with the communicated purpose. In this way, the GDPR ensures predictability for legal subjects, and avoids misuses which are based on overly broad, ambiguous or misleading phrasing. As a result, a data altruism decision requires that a reasonably precise description of the re-use is given, comparable to the informed consent requirement as described above, and that this description accurately describes the limitations to any future use of the data. This also implies that the data subject must be informed in a sufficiently systemic manner when a new purpose of re-use is identified.

The data subject also has rights that are unalienable and continue to apply even after the data has been made available for re-use. Beyond the right to revoke consent (where applicable), the data subject also has the right to restrict future processing of their data if they feel that a specific use is unlawful. Furthermore, the data subject has a data portability right to the data that they made available, implying that they may ask to receive it back in a structured, commonly used, machine-readable and interoperable format. Interestingly, this obligation might be perceived as a barrier for re-users (who must implement a way to support this right), but it is also an enabler for data altruism: the data portability right can only be respected in a relatively user friendly manner if a way is found to move it back to the user, or to third parties, thus

---

[81] The European Data Protection Board has affirmed this position. See the Article 29 Working Party Guidelines on consent under Regulation 2016/679, adopted on 28 November 2017, as last Revised and Adopted on 10 April 2018, WP257; https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=623051
[82] See https://www.medizininformatik-initiative.de/en/collaboration/consent-working-group
[83] Personal Information Management Systems (or PIMS) are systems that help give individuals more control over their personal data. PIMS allow individuals to manage their personal data in secure, local or online storage systems and share them when and with whom they choose. See EDPS Opinion 9/2016 on Personal Information Management Systems, https://edps.europa.eu/sites/edp/files/publication/16-10-20_pims_opinion_en.pdf

containing an implicit obligation to provide the interfaces and technical resources to facilitate data migration across a multitude of stakeholders.

The GDPR also contains safeguards against profiling, which can occur in data altruism cases where the personal data of the data subjects is used to firstly evaluate their situation, in order to then make inferences that affect third parties. In health care for instance, data from historic patients and their treatment might be analysed in order to create treatment profiles for future patients – resulting in profiling of future patients, on the basis of profiles created using the data made available by the original patients. This implies that care must be taken that the data made available is not used in a manner that automatically produces significant effects on the person concerned, e.g. by affecting the availability or cost of a service on the basis of the data made available.

Data altruism schemes can mitigate many of these legal problems by creating governance mechanisms that ensure that a homogeneous response can be given to all of these questions, and that the relevant safeguards (consent, revocability, purpose restriction and so forth can be verified and policed to some extent.

2.2.4   The effects of the problem

The current state of development on data altruism schemes is very fragmented in the European Union. While the number of examples of such schemes seems to be increasing, the examined schemes seem to struggle to scale up to a European level, or generally to grow beyond a strictly defined and geographically bounded context. Based on currently available data and conducted interviews, to a large extent the legal uncertainty and lack of consensus on how to address the legal challenges with regards to data privacy, permissible reuse, governance and enforcement, act as a barrier. As a result, researchers and project initiators spend much time addressing legal questions to set-up their own schemes in accordance with local rules and sensitivities. This **legal fragmentation**, **lack of awareness** and **lack of consensus on best practices deters** researchers and initiators from establishing data altruism schemes or at least **significantly increases legal costs** to establish such a scheme. The ultimate effect is that **data altruism initiatives are harder and more costly to organize in the EU, resulting in both internal market fragmentation and a competitive disadvantage towards other regions of the world when it comes to using data for the public benefit.**

2.3   Subtask 1.3: Establishing a European structure for governance aspects of data sharing

2.3.1   Background

Data sharing and data re-use are essential to data innovation. OECD assess that data sharing can "generate social and economic benefits worth between 0.1% and 1.5% of GDP in the case of public-sector data, and between 1% and 2.5% of GDP (in a few studies up to 4% of GDP) when also including private-sector" (OECD, 2019).

Yet a set of conditions needs to be in place so that businesses (and in particular across sectors) can reap the benefits from data-sharing. One of such conditions includes *the agreement and implementation of data standards, metadata standards, data schemes and interoperability principles*.

2.3.1.1 Context

Data sharing and reuse requires data holders' agreements on data standards widely adopted across industries. Those standards are not only difficult to negotiate but also, and most importantly, to implement, being often fundamental to unleash innovation.

Data standards refer to "reusable agreements that make it easier for people and organisations to publish, access, share and use better quality data"84. A data standard is considered *open* when it is accessible to everyone for use or share. The main functionalities of data standards are to[85]:

• Agree on a consistent vocabulary and common attributes for data, which are defined in registers, taxonomies, vocabularies or ontologies.
• Facilitate the exchange of data within and across organizations by employing common data formats and shared rules, which are defined in specifications, schemas or templates.
• Offer guidance for sharing better quality of data and understanding information flows, which are defined in models, protocols or guides.

2.3.1.2 Ecosystem

The data ecosystem related to subtask 1.3 refers to:

The different stakeholders in the data ecosystem considered under subtask 1.3 include (see Table 2):

• **Data holder**, which refers to companies in traditional sectors that collects, maintains and publishes data, making it available for others to use.
• **Data reusers,** which refers to any companies who use data and extracts benefits from information and insight – including both business in traditional sectors and data companies
• **Intermediaries,** which refers to any organization that facilitates data flows between data holders and data users. Data intermediary can take the role of partnerships, consortium, platform, non-governmental bodies, data standardization associations and any organizational form that facilitates data sharing across organizations.

Table 5 - Stakeholder scope (data value chain mapping)

Subtask	Data holder	Data reusers	Intermediaries
1.3	Business in traditional sector	Other business (competitors)	Public and private organisations in charge of data spaces and standardization initiatives
		Other business in the same sector (downstream/upstream)	
		Business and researchers from different sectors, esp. tech	

In order to obtain an understanding of the current problems in data standardization activities in EU, its causes and effects, we applied purposive sampling to select key informants that can provide us with a comprehensive overview. We use our judgement to choose the key informants based on i) geographical, ii) cross-sector and iii) inclusiveness of data stakeholder type, to inform our sampling, rather than aiming to construct a representative sample. Table 3 provides an overview of the on-going interviews, which are consistent with the stakeholder mapping provided for subtask 1.3 in the previous deliverable.

Table 6 - On-going interviews consistent with stakeholder mapping for subtask 1.3

Function	Category	Name of organization	Domain (if relevant)	Status
Mariane Van Teen, Director	Data intermediary	INNOPAY/iShare	Logistics	Interview done
Christoph Mertens, Head of adoption	Data intermediary	International Data Spaces	Cross domain	Pending

---

84 Open Data Institute, 'Open Standards for Data Handbook', Retrieved from https://standards.theodi.org/.
85 Ibid.

Henk Jan Vink – Managing Director	Research organisation	TNO	Cross domain	Interview done
Simon Hodson, Executive director	Research	CODATA	Research	Interview done
Eric Pol	Data intermediary	eNewGovernance	Cross domain- Personal data	Interviews done (2)
Lindsay Frost	Data intermediary	ETSI/ BDVA	Cross domain- Data standards development	Interview done
Ray Walshe	Data intermediary	BDVA	Cross-domain- Data standards development	Interview done
Tirza Van Daalen	Data intermediary	TNO – Director Geological Data Netherlands	Oil and Gas	Interview done
Eric Monteiro	Expert data standards	NTU- About OSDU	Oil and Gas	Interview done
Jochen Friedrich	Data holder and user	IBM	Cross-domain	Scheduled July
Andre Radon	Data holder and user	Volkswagen AG	Automotive	Pending
Jaana Sinipuro, Director	Data intermediary	Sitra IHAN	Cross-sector	Pending
Lars Nagel	Data intermediary	International Data Spaces	Cross-sector	Pending (in coordination with 1.4.)

2.3.1.3 Ongoing initiatives/Market analysis

Top-down Standard Development Organizations (SDOs) coordinate the development of compatibility of data standards that ensure technological progress. They are legally mandated processes where data holders come together in a participatory process of consensus building that seeks to enable the development and diffusion of data standards that are democratically agreed and aligned with broader policy goals. Yet, formal standardization processes often suffer from major drawbacks, due to the lengthy and challenging process of consensus forming and sometimes lack of market orientation.

In response to such problems, ad-hoc and industry-lead SDOs emerged with heterogeneous origins, goals and institutional logics. Industry-lead standards development processes are self-organized and not mandated by law. They include industry consortia, loosely coordinated temporary working groups and task forces, but also not-for-profit organizations that help stakeholders to organize data standard making processes in a more permanent fashion. Such de facto data standards result from market-based standard setting processes, in which data standards are generated through competitive forces.

While top-down SDOs emphasize consensus and social welfare implications, de-facto data standards stress speed, agility and the needs of the industrial constituents. Both standardization efforts co-exist to develop compatibility standards; which implies that the process of standardization evolves within and across multiple SDOs.

While different formal or informal SDOs have emerged to foster data standardization needs, a set of **intermediaries** are facilitating the agreement between data holders and users on data standards. As part of the research effort in the present study, a set of data intermediaries have been identified and interviewed (see table 3). We provide a brief description of them below and some estimates about the benefits and costs

that they incur according to the sources gathered in the desk research and the evidence provided during the interviews (see detailed analysis in section assessing the four policy options):

2.3.1.3.1 International Data Spaces Association[86]

IDSA consists of a trustworthy architecture where more than 101 companies and institutions across industries from more than 20 countries. The goal of IDSA is to guarantee data sovereignty by **reference architecture for peer-to-peer network** providing usage control of data from all domains. Trust and security are the core pillars structuring its work. IDSA aims at providing the architecture that supports sharing data between different endpoints while ensuring data sovereignty. Main components in IDS architecture are the so-called 'IDS Connectors', which are the gateways that ensure control over data sharing at any source and point of use.

One key element of these connectors is the automated enforcement mechanisms of the relevant data policy – such as restrictions, limited persistence, disallowing transfer to third parties and so on, based on the standard policy language XACML (eXtensible Access Control Markup Language) and Information Model's Usage Control module, which provides machine-readable specifications of usage control policies. In other words IDS has enforcement mechanisms built in the architecture to control which data are used, how and by whom.

The governance and control in IDS ecosystem is expressed in the *certification* criteria for tiered security levels, specifically appointed evaluation facilities and certification bodies also at global scale. It enables the secure exchange of data and easy integration and aggregation of data in business ecosystem. By employing certified core components and certified technical and organizational security measures, IDS guarantees to its member that the architecture operates under the principles of trust. The certification of participants and components takes place in two phases:

- IDS_ready Review, which is implemented by members of the certification working group and the Head Office in charge of issuing the IDS_ready statements.
- IDS Certification, which consists in an evaluation implemented by evaluation facilities and approved by the IDS Certification Body, which is responsible for issuing the IDS certificates.

Organizations engagement in IDSA requires around 20% dedication of one person and is estimated to generate in average around 15% efficiency savings for the companies. The current 22 use cases of IDSA that span from logistics, defence and manufacturing sectors reflect how a common reference architecture (i.e. technical, procedural, organizational and legal) leads to companies efficiency gains.

2.3.1.3.2 CODATA[87]

CODATA is the Committee on Data of the International Science Council (ISC). CODATA exists to promote global collaboration to improve the availability and usability of data for all areas of research. CODATA Works towards fostering scientific data sharing. The principle of CODATA is that research data should be as open as possible and as closed as necessary. CODATA works also to advance the interoperability and the usability of such data: research data should be FAIR (Findable, Accessible, Interoperable and Reusable). Recent studies have estimated that the annual financial cost of not sharing FAIR data to be at least EUR 10.2bn for the European economy; an additional estimate of the impact of FAIR on potential economic growth is EUR 16bn annually (European Commission 2019b).

---

86 International Data Spaces Association: https://www.internationaldataspaces.org/

87 CODATA: https://codata.org/

In consequence, one of the current tasks of CODATA is to support scientific data sharing across research domains. CODATA facilitates the dialogue across disciplines to agree on minimum common denominator across research domains about metadata and data structure that facilitates data interoperability across research domains. To achieve such goal, CODATA has set up a number of standing committees and strategic executive led initiatives, and through its Task Groups and Working Groups. It also collaborates on major data conferences like SciDataCon and International Data Week.

The benefits of research data sharing include: to improve reproducibility; to accelerate scientific processes and research velocity; increased scientific quality; to prevent scientific fraud; and to increase scientific productivity by reducing redundancy and innovation gains (Borgman 2015; Edwards et al. 2011; European Commission 2019a; OECD 2015; Tenopir et al. 2015).

Yet the average estimated costs of introducing the metadata and contextual information required for scientists to re-use the data are around 5% of the total research budget. Other sources estimate that such production of metadata and the contextual descriptions of datasets could span an estimated two to three weeks from an average of a two-year research grant application (OpenAire 2019). In a dedicated study to examine high-energy physics practices, the vast majority of respondents (94.3%) thought that "the additional effort needed for preparing data for preservation in a re-usable form is substantial (more than 1% of the overall effort invested in producing and analysing the data) whereas 43.0% think that the supplementary effort is more than 10%"(Holzner, Igo-Kemenes, and Mele 2009).

2.3.1.3.3   iSHARE[88]

iSHARE consists in a stable set of agreements that makes possible that organizations give access to their data to a pool of unknown organizations without requiring bilateral or ad hoc agreements. With the scope in the logistics sector, iSHARE was set up through a collaborative project that started in 2017 lead by Innopay, a consultancy firm in the Netherlands, and soon the benefits of the scheme were acknowledge and lead to the creation of iSHARE foundation.

By giving all organizations the same identification, authentication and authorization methods, companies joining iSHARE scheme gain efficiency and do not incur in costs every time they want to share data with a specific organizations. Overall, organizations can avoid time-consuming integrations when they want to share data. ISHARE set of agreements allow the data owner to remain in full control over their own data at all times. Additionally, data owners decide the terms under which their data will be shared, with whom and for how long.

As such, iSHARE combines functional, technical, operational and legal agreements that organizations adhere. These set of agreements support both Machine to Machine (M2M) or Human to Machine (H2M) interaction. It also supports portable identity(s) for parties and humans; flexible authorizations, applicable in heterogeneous context; facilitates data exchange based on delegations; control over own data through management of consent; and provides a trust framework.

The operational description of what iSHARE offers to the data holders and users is the following: participants sign one contract with the Scheme Owner, which implies having a contract with all participants of iSHARE automatically. While participants remain free to develop additional contracts that do not conflict with iSHARE scheme, by signing the contract with the Scheme Owner, participants are able to share their data amongst them. As part of iSHARE scheme, an important aspect is the trust framework that the scheme designs, where licenses define the conditions under which data can be shared or the services that can be offered or

---

[88] ISHARE: https://www.ishareworks.org/en/node/6

consumed. The trust framework relies on technological solutions that allow organizations to authenticate with the other in a reliable way. Within the iSHARE scheme there is an API (Application Programming Interface) architecture for identification, authentication and authorization, which is based on a modified version of OAuth and OpenID Connect standards. While the setup of iSHARE scheme cost around few million Euros, the maintenance of the scheme is considered to require less than one million operational cost, which can be supported through members fee and transitioning public grants.

iShare and IDSA are working on similar issues and have formalized their collaboration in December 2019. iShare is at a more advanced stage of deployment but it is narrower in scope as it focusses only on the first two stages of trust and security, as illustrated by the chart below. The iShare solution is now aligned with the IDSA reference architecture and can therefore be consider as one solution to implement the IDSA framework.

Figure 1: Relation between iShare and IDSA
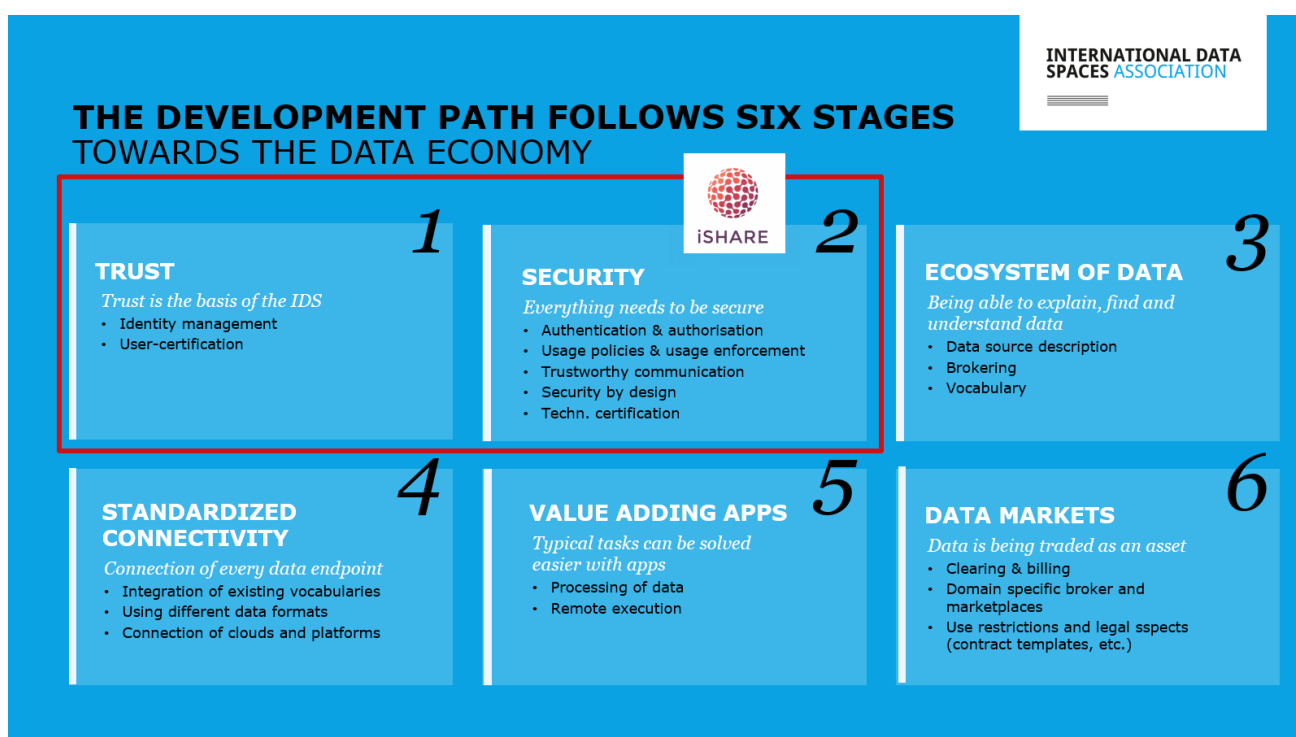
2.3.1.3.4 ABOUT ML

Data sharing for machine learning purposes has specific challenges, notably that thee way datasets have been created can lead to bias. Hence the need for full documentation of the origin and purpose of machine learning datasets. Different companies have adopted their own solution for metadata of these datasets, such as Google dataset fiche or Microsoft datasheet for datasets, but the lack of a standard limits the scalability of reuse. [89]

This is the rationale behind ABOUT ML (Annotation and Benchmarking on Understanding and Transparency of Machine Learning Lifestyles) is an initiative created by the leading AI industrial consortium representing all the main players, with the intention of establishing, encouraging and promoting novel standards for

---

[89] See Margaret Mitchell and others, 'Model Cards for Model Reporting', Proceedings of the Conference on Fairness, Accountability, and Transparency - FAT* '19, 2019, 220–29 <https://doi.org/10.1145/3287560.3287596>; Timnit Gebru and others, 'Datasheets for Datasets', ArXiv:1803.09010 [Cs], 2020 <http://arxiv.org/abs/1803.09010> [accessed 29 June 2020].

transparency within machine learning systems by way of documentation. This aims to be done by studying best practices from inception to deployment.

The project is broken down into 8 phases: Understand latest research; understand current practice; combine research theory and results of current practice into testable pilots; run pilot test with PAI partners/ organizations (not individually specified); collect data from pilot transparency practices; iterate on pilots with the latest research and practice; when enough evidence has been collected, elevate it to a best practice; and promote effective practices to establish new industry norms for transparency. The partner organizations include: Facebook, Xbook, IBM, Leverhume Centre, Accenture, Quantumblack, Mckinsey & Co.,Future of Humanity Institute, EFF, Future of Privacy Forum, Deepmind, Berkman Klein Center, Tech Policy Lab (UoW), Google, Policy Link, AI Now, Berggruen Institute, Data & Society, Center for Internet and Society, Sony, BBC, UCL, Microsoft, Intel, Vision and Imaging Processing. The main drivers for industry engagement in AboutML are to avoid misuse and harm that arise from ML systems by creating guidelines for transparency documentation which, if implemented early on, can prevent future harm. Additionally the partnership provides major corporations to gain legitimacy through being associated with one of the first great push for transparency initiatives. It also seeks to insulate major corporations from future challenge by regulators or public opinion by highlighting the influence and guidance of civil society and social justice organizations in the creation of these standards. The partnership annual revenue has been the following: USD 7.25M (2017); USD 10.53M (2018); USD 8.14M (2019) where USD 3.91M spent on all programs including About ML.

2.3.1.3.5 eNewGovernance[90]

aNewGovernance is a Public Private Partnership launched in 2020 designed to support free flow of data in a human centric approach. The partnership seeks to support start-ups, SMEs, corporates, local authorities and governments alike to develop new services based on data re-use with no trust or liability issues. By supporting themselves in already on-going initiatives such as MyData, Fiware or Gaïa-X, aNewGovernance seeks to govern the data landscape to empower users by allowing their interaction with their personal data via technological tools that enable them to enact their right to data portability as claimed in GDPR.

ANewGovernance has the goal to develop an infrastructure that fosters data sharing and enables interoperable data ecosystem. The partnership seeks to ensure that organizations storing the data are not managing the permissions over the data use. Some of the expected outputs by aNewGovernance are: agreed data models, liability model for data sharing, personal data sharing APIs and a common consent or permission layer.

2.3.1.3.6 BDVA- Standardization working group

The Big Data Value Association (BDVA) is an industry-driven international not–for-profit organisation with more than 200 members across Europe, which contributes to the implementation of the Big Data Value PPP program. As part of their activities, the organization fosters a wide range of activities to facilitate data sharing across industries. Under these activities, the organization has a task force devoted to foster data standardization (under task force 6). Data standardization activities are estimated a dedication of 2 to 3 hours per week of a person; around 3 to 5 meetings per year with an average of 3 to 6 days of meeting and the correspondent (and usually continental) travel and accommodation costs of such 3-5 meetings. Organizations dedication can go from 1 to 7 people dedicated in the participation in the standardization process. As a result, data standardization activities face incentive problems for companies (in particular SMEs) who need to have a clear business case before engaging in such high-effort-intensive tasks.

---

90 ANewGovernance: https://www.anewgovernance.org/

2.3.2 **The problem, its magnitude and the stakeholders affected**

The overall problem analysis rests on a set of causal relationship summarized in the table below.

In this specific context, the problem is represented by the suboptimal adoption of data sharing by companies, which leads to lower innovation and productivity in traditional sector as well as in data business. Two of the main barriers to data sharing lie in the limited standardisation of data and metadata, lack of interoperability and trust. While fear and perceived risks of sharing data, reduces the likelihood in data sharing, other factors such as limited standardisation of data and metadata come into place when an organization negotiates access to data. Standardization is in such context a cost reduction strategy. The ongoing initiatives are therefore not designed at stimulating data sharing *per se*, but the setting up of governance mechanisms to support and accelerate standardisation, within sectors and across them. In particular, the goal is to facilitate the speeding up and scaling up of the standardisation activities which fall fully under the European strategy for ICT standardisation.[91]

Table 7 - overview of the problem analysis

Ongoing initiatives	Causes	Problem	Effects
Standardisation and coordination initiatives	Lack of data and metadata standards, data schemes within sectors Lack of technical interoperability across sector	Lack of data sharing within/across sector	Lower productivity and innovation

Data sharing among business is increasing but remains below optimal leading to missed economic opportunities. In a 2018 report by Everis, 60% of companies do not engage in b2b data sharing. Deloitte estimates that the vast majority of the benefits expected from IoT data in different sectors by 2027 stems from data sharing, but that data sharing has reached only a minor part of its potential: 32% for horizontal (between competitors) data sharing, 47% for vertical (business in the same value chain) and 31% for data sharing across sectors.[92] In particular, the opportunities stemming from data sharing across sector are remarkable but clearly smaller than data sharing within sector – from one third in the case of manufacturing to about 60% in the case of automotive.

More data are available on scientific data sharing to illustrate the problem. Only 14% of researchers deposit their data on trusted scientific repositories which gather less than 20% of overall scientific data.[93]

This is particularly important for advanced, data intensive machine learning applications. In fact, access to data is the second most frequently mentioned barrier for artificial intelligence in Europe.[94]

The stakeholders affected by the problem are of three types:

- The data holders are any data generating company in Europe. There are 22 million companies in the EU. However, of these 22 million, only around 700,000 are considered genuine "data users" by the EU data market study because of their intensive use of data, increasing to 844,000 by 2025 in the most favourable scenario.[95]
- The data reusers are any other company in Europe, since by nature data spaces allow for peer to peer data sharing between companies. In addition, specific benefits will be drawn by a subgroup of reusers,

---

[91] COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS. Digitising European Industry. Reaping the full benefits of a Digital Single Market. COM/2016/0180
[92] Deloitte, Realizing the economic potential of machine-generated, non- personal data in the EU.
[93] The Lisbon Council and others, 2019. STUDY ON OPEN SCIENCE: MONITORING TRENDS AND DRIVERS. European Commission
[94] Claire Beatty. The global AI agenda: Europe. MIT Technology Review Insight, 2020.
[95] These data come from the EU data market study. See www.datalandscape.eu

the technology companies. According to the same study, there are an estimated 280,000 data companies in the EU.

- Intermediaries are composed by dedicated intervention (market or government led) to facilitate data sharing and data standardization. They are difficult to quantify but can be considered in the order of 10 to 100 if we limit ourselves at intervention with visible footprint at EU level.

Table 8 - Estimated number of stakeholders impacted in the EU-27

Stakeholder type	Business in traditional sectors	Data companies	Intermediaries	Total
Data holders	22 million companies of which, 700,000 use data extensively			**Approx. 700,000**
Data re-users	The same 22 million companies of which, 700,000 use data extensive	280,000 data companies		**Approx. 1,000,000**
Intermediaries (data sharing and standardisation)			10-100	**Approx. 10-100**

2.3.3 The causes of the problem

Barriers to data sharing include the use of varying and non-compatible data standards that make difficult to integrate, aggregate and combine different data from diverse data holders in the ecosystem. These varying data standards are a bottleneck for data reuse. The same study points out that the "lack of common sharing protocols and standards" is one of the main barrier in manufacturing and implies the loss of about 40% of valuable data sharing – mainly when it comes to vertical data sharing among players in the same value chain. On the same tone, a second barrier is that "the cost of normalising data to be shared is high". On the other hand, when standards are implemented in cases such as the OpenActive standards for physical activities, this can result in a visible increase in data sharing – 200,000 new activities were shared after the introduction of the standard, resulting in 150 to 500,000 new activities carried out by consumers per month. The case of OpenActive also shows that the development of standards was only part of the problem solved, but the OpenActive also coordinated work across the sector to communicate the benefits of standard adoption. The case reflects that standardisation involved more than just technical work, but also engagement towards adoption in order to unlock such benefits.

Hence, interoperability sits at the core of data sharing goals and it implies engaging in data standardisation processes, whether bottom-up (industry-led) or top-down via *de jure* Standards Development Organizations (SDOs) mandates.

The challenges of making industries to agree and widely adopt data standards to achieve the desired interoperability for data re-use are substantial. While data standards can potentially ensure that industry sectors are more competitive and support a vibrant ecosystem of innovative new business, the lengthiness, time-consuming efforts by businesses, lack of incentives of for-profit to engage in such standardization process when no clear concrete business case on data re-use is clear, and the complexity to achieve practical and wide consensus, make data standards a real barrier for data sharing. Hence, standardisation needs to be driven by either regulatory intervention that addresses market failures, or to address specific goals across a business ecosystem. If standards are not connected to the goals of the sector then they are less likely to be adopted. In other words: generic standards for data sharing and metadata will be more challenging to develop and adopt than those that support specific use cases. Yet at the same time, there is a clear economic case for greater findability and interoperability of data across sector – this is where most of the benefits will

come in sectors such as automotive.[96] It can be argued then that the more generic the data standard, the less likely it is to be entirely market driven. It is worth noting that many of the initiatives for generic, scalable data sharing are backed up by government agencies (as it is the case for iShare and SITRA IHAN as well as FAIR for research data).

A second major bottleneck is trust. Companies are reluctant to share data because of the loss of control it implies. Data reuse can harm them from a competitive perspective, by letting other companies understand valuable information about the processes. And it can expose them to legal problems, for instance in relation to GDPR or commercial secrets. In the same study, Deloitte points out that for manufacturing the remaining top barriers are: "Exposing machines to attack and/or inadvertently disclosing commercial secrets" and "Legal procedures will need to replicated for every data-sharing partner, which is time consuming". In the 2018 report by Everis, the 60% of companies which do not engage in b2b data sharing attribute this to privacy concerns, trade secrets, and fear of misappropriation. In another report by Accenture, data breach lead to almost 10 percent-decline in revenue for up to six months after the breach compared to companies who did not suffer a breach. [97]

Another trust related issue refers to the potential risks of reusing datasets for machine learning without a full understanding of their limitation, leading to possible mistakes in decisions as well as discrimination. Datasets useful for some purposes are not for others, because of their inherent limitations. Companies have started to work on "datasheets" similar to those of the electronic devices. Just as a transistor provides information on the range of temperature for its use, the datasheet would provide information on the key limitations of the sample of the datasets.[98] However, we are still at an early stage where each company is starting to provide this information in different ways.

The further away the data travel from the original data holder, the more the need to provide valid documentation to support reuse – mainly through metadata describing the data and additional aspects including the method of collection, the purpose of reuse, the consent of the various data holders (including personal data) but also additional documentation is required beyond metadata, as the cases such as Datasheets show. The more investment in good metadata and documentation, the more reused the data will be, as reflected in the current European work on high value datasets. Yet there is an issue of effort optimization: While datasets that are less likely to be reused, e.g. from small studies or experiments, may not need the same level of documentation, standardisation, data that is expected to be widely re-used needs to appropriate investment on metadata and documentation. A clear example of the relevance of metadata and documentation more broadly comes from scientific data, where the FAIR principles require documentation and formatting that enables the widest reuse by different communities. Based on the experience of the interviewee, it is widely recognized that providing high quality metadata and documentation for scientific datasets requires 5 to 10% of the total project budget – a very substantial expenditure.

In other words, both interoperability and trust issues are solvable. There is a long history of standardization, metadata, interoperability and definition of sharing agreements. But while those activities require time, consensus and coordination efforts across stakeholders, in the meantime present needs for data sharing are

---

[96] Deloitte study for Vodafone group, Realising the economic potential of machine-generated, nonpersonal data in the EU, see: https://www.vodafone.com/content/dam/vodcom/files/public-policy/Realising_the_potential_of_IoT_data_report_for_Vodafone.pdf
[97] See Everis, 2018, Study on data sharing between companies in Europe, European Commission and Accenture, 2019. Maximize collaboration through secure data sharing
[98] Timnit Gebru and others, 2020, 'Datasheets for Datasets', ArXiv:1803.09010 [Cs]<http://arxiv.org/abs/1803.09010> [accessed 29 June 2020].

mostly solved with bilateral contractual arrangements which do not scale adequately and entail excessive cost and most importantly an opportunity loss for most organizations.

There is a need of a set of high-level guiding principles to support data standardization efforts that can cut across any data governance attempt. As FAIR principles are currently guiding the data standardization efforts in research and the activities providing the appropriate metadata, data structure and descriptions, some complementary guiding principles could support the spread and fragmented efforts towards data standardization across the different industries. While there is no such a one-size-fits-all approach for data standardization, grounding efforts in general underlying principles can provide a source of clarity across the diverse verticals and application areas.

While there are diverse data intermediaries that have emerged to support or even coordinate some data standardization efforts, yet there co-exist many conflicting and non-interoperable data standards. The cause of the problem is that standards are usually developed locally and vertically to facilitate data sharing. For instance, where Pistoia Alliance or Global Alliance for Genomics and Health set up data standards to support scientific data sharing in biomedical research, when organizations need to aggregate health patient data we move towards what is considered another sector (health) were data standards are agreed by parallel standardization efforts in different SDOs. If we need to move further and for epidemiological purpose aggregate data about the environment then the data standards will be again different and conform to the consensus of 'another' community. In sum, while data standards agreed locally or in a sectorial basis work, they lack the ability to travel beyond and far from its origins, in the absence of some minimal and generic prescriptions of how the structure and data descriptions should look like. The absence of such standardization effort that allows data to travel across sectors impedes innovation and supposes high-cost.

In sum, there is a need to invest in standards development around specific use cases e.g. by sector or broader challenges but also to bridge across environments and sectors. While data standards agreed locally or in a sectorial basis work, we need also standards that make possible for data to travel across sectors far from its origins. The adoption of common principles for standards development, data access and data governance could help make that achievable. Overall, the absence of such standardization effort within but also across sectors impedes innovation and supposes high-cost.

2.3.4 **The effects of the problem**

The effects of the fragmented efforts towards data standardization and the lack of incentives of for-profit to actively engage in such processes, in particular cross-sector, to facilitate data sharing across industries has an effect not only in restricting data-driven innovation but also in incapacity to reap the benefits of data in terms of efficiency and productivity.

In manufacturing alone, data sharing of IoT data is expected to generate - if fully implemented – 1.4 trillion euros in increased productivity by 2027. Another study by the World Economic Forum estimates at 100bn the current opportunity for data sharing in manufacturing.[99]

In the broader context, the historical contributions to Germany economic growth rate attributed to standardization is 0.9%, and for the period from 2002 to 2006 the total economic benefit of standardization averages about 16.77 billion Euros per year.[100]

Additionally, widely adopted data standards contribute to more transparent competition: Not only individuals can benefit by having better products and services at lower prices, but also organizations can benefit by

---

[99] WEF, 2020. Share to Gain: Unlocking Data Value in Manufacturing.
[100] Manchester Institute of Innovation Research, 2013. The Impact of Standardization and Standards on Innovation, NESTA.

driving up their profits. The collaboration of governments and for-profits is required in such standardization activities to ensure a minimum stack of standards for authentication, consent interoperability, sector-specific API available. Scalable portability and interoperable consent management is important to enable data sharing and some type of standardization or agreement on data formats, descriptions and workflows is needed. Ultimately, by providing general guidelines that inform data standards across sectors may expect to increase the gross value added to the economy through increased productivity.

2.4 Subtask 1.4: Establishing a certification framework for data intermediaries

2.4.1 Background

2.4.1.1 Context

Data intermediaries play an increasingly pivotal role in the thriving data sharing and selling market. The OECD report 'Enhancing Access to and Sharing of Data' defines **'data intermediaries'** as organizations that 'enable data holders to share their data' which 'may also provide additional added-value services such as data processing services, payment and clearing services and legal services, including the provision of standard-license schemes".[101] Additionally, the EC Communication "Building a European Data Economy" of 2017 and its accompanying SWD present data market places as organizations that facilitate data use and exchange and identifies three types of data intermediaries (i.e data marketplaces, industrial data platforms, personal information management services),[102] while the above-mentioned OECD report identifies five types of data intermediaries (i.e data repositories, data brokers, data marketplaces, Personal information management systems/personal data stores, trusted third parties). Other types of data intermediaries include data unions, data cooperatives, data collaboratives and data trusts. The Open Data Institute (ODI) has published a visualisation of the different concepts in use with a proposed clustering.[103] This study, aims to examine the intermediaries that are neutral in the sense that they are at least functionally/organisationally separate from both data holders and data users.

This overarching and very broad definition of data intermediaries encompasses many different types of organisations which can have very different characteristics:

- **They can be focused on personal or non personal data (or both)**: the Staff Working Document[104] accompanying the Communication on Building a EU Data economy[105] for instance distinguishes between 1) non-personal data Industrial Data Platforms (which can be vertical an sectorial like AutoSar[106] or community led and horizontal like FiWare[107]) and 2) Personal Information management services (such as generic solution providers like Mydex[108], digi.me[109], Meeco.me,[110] Polypoly[111] or sector specific solution providers like MiData Cooperative[112] etc.). However, the line between personal and non-personal data intermediaries is liked to become more blurred in the future and for certain specific sectors (i.e. health and automotive).

---

[101] OECD Report (2019) 'Enhancing Access to and Sharing of Data', chapter 2
[102] COM(2017)9 and SWD(2017)2
[103] https://theodi.org/project/the-data-access-map/
[104] COMMISSION STAFF WORKING DOCUMENT on the free flow of data and emerging issues of the European data economy, Accompanying the document Communication, Building a European data economy, 2017, https://ec.europa.eu/digital-single-market/en/news/staff-working-document-free-flow-data-and-emerging-issues-european-data-economy
[105] Communication on Building a European Data Economy, 2017, https://ec.europa.eu/digital-single-market/en/news/communication-building-european-data-economy
[106] https://www.autosar.org/
[107] https://www.fiware.org/about-us/
[108] https://mydex.org/
[109] https://digi.me/
[110] https://meeco.me/
[111] https://polypoly.eu/en/home
[112] https://www.midata.coop/en/home/

- **They can be completely independent from data holders or stem from data holders' initiatives**: in some cases, data intermediaries are established by data holders in order to enhance the access to their data, control how their data is being used and gain access to third parties' data. These types of organisations can be found within many different value chains. For instance, in the railway sector, a specific "data space" has recently been established by railway infrastructure service providers, original equipment manufacturers, railway operators and other stakeholders to pool together everybody's data and securing data exchange while maintaining data sovereignty.[113] Other examples of industrial data platforms developed by big industrial players include among others Mindsphere (Siemens), Skywise (Airbus), RIO (Traton Group), Predix (GE Digital), FieldView, Xarvio, as well as the Data Intelligence Hub of Deutsche Telekom and Radianz of BT Group.

- **They can provide only access to data or services on top of the data**: as the OECD suggests, some data intermediaries provide services on top of the data and they specialise in offering data storage or access management features to their clients. This is the case for instance of data intermediaries like Nallian[114] which provides standard license schemes for sharing the data uploaded on the platform as well as the possibility of plugging in applications for smart billing and smart auditing.[115]

- **They can be well-established players with a long history of providing data or start-ups and newly established businesses:** some industries and especially the financial industry are used to the existence of big data brokers such as Bloomberg, Thomson Reuters, etc. and some of these players date back of several decades. For other industries, such as the aerospace or automotive industries to name two, these players are new and respond to the changes brought by the data and internet of things economy.

- **They can be profit driven or not**: the OECD mentions that, on top of business driven data intermediaries, public data repositories such as those set up by public libraries or scientific communities can also be considered as data intermediaries[116].

These and other differences make data intermediaries a very heterogeneous category of players. However, their common characteristic lies in their role of **matchmakers between demand and supply of data**. For this reason, they are also sometimes called "data marketplaces" or even "data brokers" although these terms are also used to identify more specifically certain types of data intermediaries. According to the Summary report of the open public consultation on the European strategy for data, almost 60% of the 772 respondents to this section considered that emerging novel intermediaries, such as 'data marketplaces' or 'data brokers', are useful enablers to the data economy, while almost 22% don't know or remain neutral to the question.[117]

2.4.1.2 Ecosystem
The data-sharing ecosystem includes various types of stakeholders involved in the value chain of the data intermediaries, including in particular data holders, data re-users and (certified) data intermediaries.

The **data holders** in this value chain are the data providers sharing their data with the data users through the certified intermediaries. The **data intermediaries** in this value chain are the enablers of data sharing between data holders and data re-users. The **data re-users** will be the clients of the data intermediaries.

---

[113] https://www.internationaldataspaces.org/knorr-bremse-establishing-data-sovereignty-and-data-ecosystems-in-the-rail-industry/

[114] https://www.nallian.com/solution/how

[116] OECD, Enhancing Access to and Sharing of Data, Reconciling Risks and Benefits for Data Re-use across Societies, 2019, http://www.oecd.org/going-digital/enhancing-access-to-and-sharing-of-data-276aaca8-en.htm

[117] https://ec.europa.eu/digital-single-market/en/news/summary-report-public-consultation-european-strategy-data

The table below provides three different approaches of the main stakeholders identified for the subtask 1.4, as potentially affected from the establishment of a certification framework for data intermediaries: a) a generic approach, b) the B2B data-sharing scenario and c) the C2B data-sharing scenario.

Table 9 - Stakeholder scope (data value chain mapping)

Subtask	Approach	Data holder	Data (re)user (whole dataset)	Intermediaries	Personal data involvement	Purpose
1.4	Generic Approach	Data Providers: Businesses/ Academia and research organisations /Governmental Organisations/ NGOs/ Citizens	Intermediaries' Clients: Businesses/ Academia and research organisations /Governmental Organisations/ NGOs/ Citizens	Certified data Intermediaries (i.e. Data marketplaces, data brokers, data repositories, PIMS/PDS, industrial data platforms, trusted third parties, data unions, data cooperatives, data collaboratives, data trusts)	Potentially	Business, R&I, Public Good
1.4	B2B Approach	Data providers: Businesses	Intermediaries' Clients: Businesses	Certified data intermediaries: Data marketplaces, industrial data platforms, trusted third parties, data collaboratives, data trusts	No	Business, R&I
1.4	C2B Approach	Data providers: Citizens	Intermediaries' Clients: Businesses	Certified data intermediaries: PIMS/PDS, data unions, data cooperatives, data collaboratives, data trusts	Yes	Business, R&I, Public Good

2.4.1.3 Data holders

For the subtask 1.4, data holders will be the data providers who will be enabled to share their data through the data intermediaries. In a generic approach, the types of data providers might vary depending on data intermediary category and could include businesses, academia and research organisations, NGOs or citizens. **This study focuses on two different data sharing scenarios: a) Business to Business data sharing (B2B) and Consumer to Business data sharing (C2B).**

In the B2B scenario, businesses is the most common source of data for certified data intermediaries like data marketplaces, industrial data platforms, trusted third parties, data trusts and data collaboratives. In the C2B scenario, individuals are the data providers for certified data intermediaries like personal information management service (PIMS) or personal data stores (PDS), data unions, data cooperatives, data collaboratives and data trusts.

2.4.1.4 Data Intermediaries

- Depending on the type of data sharing **Data marketplaces**: There is no uniform definition of what a data marketplace is. The Commission in its 2017 Staff Working Document *(SWD(2017)2)* followed the definition of Stahl et al. describing data marketplaces as electronic marketplaces where data is traded as a commodity, an electronic marketplace being "the concrete agency or infrastructure that allows

participants to meet and perform the market transactions, translated into an electronic medium",[118] while the OECD report 'Enhancing Access to and Sharing of Data' defines them as online platforms that host data from various publishers and offer the (possibly enriched) data to interested parties.[119] Finally, a Forrester research report entitled "The Insights Professional's Guide to External Data Sourcing, Beginner Level: Data Practices For Insights-Driven Businesses" defines data marketplaces are data exchanges that enable sellers to offer data products and services and enable buyers to find and acquire data, often as a self-service, transactional model.[120]

- **Industrial Data platforms** are defined in SWD(2017)2 as virtual environments facilitating the exchange and connection of data among different companies and organisations through a shared reference architecture, common governance rules and within a secure business ecosystem.[121]

- **Trusted third parties / Data intermediary acting as a third-party certification authority:** According to the OECD report 'Enhancing Access to and Sharing of Data' in some cases, data intermediaries can act as a certification authority as in the case of the Industrial Data Space (IDS). The certification authorities of the IDS certifies all participants based on standards defined by the IDS regarding, for example, security, privacy, and terms of use. Data owners define terms of use and the fees of data use, which data brokers use to match with other data owners and users.[122] Participants and core components shall provide a sufficiently high degree of security regarding the integrity, confidentiality and availability of information exchanged in the Industrial Data Space. Therefore, an evaluation and certification of the core components as well as of the technical and organizational security measures is mandatory for participating in the Industrial Data Space. This applies to both organizations that develop software components intended to be deployed within the Industrial Data Space (i.e., prospective software providers) and to organizations that intend to become participants in the Industrial Data Space. During the certification process, the primary focus of the evaluation will be either on the product or on the organization itself.[123]

- **Data collaboratives** are defined as a new form of collaboration, beyond the public-private partnership model, in which participants from different sectors—including in particular private companies, research institutions and government agencies - exchange their data to solve public problems and create public value.[124]

- **Personal information management services (PIMS) and personal data spaces**[125] : The OECD report 'Enhancing Access to and Sharing of Data' defines PIMS/PDS as platforms to give data subjects (consumers) more control over their personal data and thus to restore user agency, including in the context of the Internet of Things.[126] The SWD(2017)2 further defines (PIMS) as "a developing set of technical means, currently in its infancy, for individuals to manage control over their personal data.

[118] SWD(2017) 2, p. 17
F. Stahl, F. Schomm, G. Vossen, & L Vomfell, A Classification Framework for Data Marketplaces, Vietnam J Comput Sci, 2016, p. 137.
[119] OECD Report (2019) 'Enhancing Access to and Sharing of Data', chapter 2 (p.36)
Dumbill, E. (2012), Microsoft's plan for Hadoop and big data, http://radar.oreilly.com/2012/01/microsoft-big-data.html
[120] Forrester research, The Insights Professional's Guide to External Data Sourcing, Beginner Level: Data Practices For Insights-Driven Businesses, May 2019
[121] SWD(2017) 2, p. 18
IDC and Open Evidence, European Data Market Study, 2016, publication forthcoming, https://docs.google.com/a/open-evidence.com/viewer?a=v&pid=sites&srcid=b3Blbi1ldmlkZW5jZS5jb218ZG93bmxvYWR8Z3q6NjJiZTQ1NTYyZjdlOGNhNg
[122] OECD Report (2019) 'Enhancing Access to and Sharing of Data', chapter 2
[123] https://www.internationaldataspaces.org/publications/whitepaper-certification/
[124] https://ec.europa.eu/knowledge4policy/online-resource/data-collaboratives_en & https://oecd-opsi.org/toolkits/data-collaboratives-canvas/ & http://thegovlab.org/the-emergence-of-data-collaboratives-in-numbers/
[125] Also defined by other terms, including among others personal data stores (PDS)/vaults/wallets/clouds or infomediaries, vendor relationship management tools, life management platforms, personal information management systems; information fiduciaries, mediators of individual data - MID, information banks. [source Understanding MyData Operators White paper, https://mydata.org/wp-content/uploads/sites/5/2020/04/Understanding-Mydata-Operators-pages.pdf]
[126] OECD Report (2019) 'Enhancing Access to and Sharing of Data', chapter 2
Urquhart, L., N. Sailaja and D. Mcauley (2017), "Realising the right to data portability for the domestic Internet of things", Personal and Ubiquitous Computing, http://dx.doi.org/10.1007/s00779-017-1069-2.

While considerable conceptual differences exist, PIMS can be summarised as technical means which individuals can use in order to exercise their right to data portability under article 20 GDPR. PIMS in this respect can serve as a means to receive back personal data from data controllers (within the limits of the right under article 20 GDPR). PIMS would then also give individuals the means to provide personal data through a web or mobile application for processing by others on the basis of one of the legal bases of the GDPR (e.g. consent, performance of a contract)".[127]

- **Data unions:** A Data Union is a framework, currently being built on the Streamr Marketplace that allows people to easily bundle and sell their real-time data and earn revenue. On its own, our data does not hold much value, but when combined in a Data Union, it aggregates into an attractive product for buyers to extract insights. This is crowdselling, and has the potential to generate unique data sets by incentivising trade directly from the data producers.[128]

- **Data cooperatives:** Similarly to the above mentioned data unions, data cooperatives can be defined as structures that enable the creation of open data and personal data stores for mutual benefit; they could rebalance what many perceive as asymmetric relationship between data subjects (people with personal data) and data users (people who use data to develop services and products)[129]

- **Data trusts:** The ODI defines data trusts as legal structures that provide independent, fiduciary stewardship of data. Data trusts are an approach to looking after and making decisions about data in a similar way that trusts have been used to look after and make decisions about other forms of asset in the past, such as land trusts that steward land on behalf of local communities. They involve one party authorising another to make decisions about data on their behalf, for the benefit of a wider group of stakeholders. With data trusts, the independent person, group or entity stewarding the data takes on a fiduciary duty. In law, a fiduciary duty is considered the highest level of obligation that one party can owe to another – a fiduciary duty in this context involves stewarding data with impartiality, prudence, transparency and undivided loyalty.[130]

2.4.1.5   Data re-user

For the subtask 1.4 the data (re-)users will be the data intermediaries' clients. In a generic approach of the value chain, the client base of a data intermediary could entail various categories of organizations including businesses (e.g. buyers and suppliers), academia and research organisations, NGOs, public sector organisations and citizens. The category of data re-users vary according to the type of data intermediary and the services provided. In the B2B and C2B scenarios in the context of this study, the client base of the intermediaries will be mainly comprised of businesses. In particular, for industrial data platforms (B2B data platforms) the client base will most likely be comprised of businesses such as Original Equipment Manufacturers (OEMs) first and second tier buyers and suppliers.

2.4.1.6   Ongoing initiatives/Market analysis

Non-exhaustive listing, providing examples of B2B and C2B data intermediaries, active in the European market is presented in the tables below.

Table 10 - Data Intermediaries - B2B Data Sharing European Market Overview

Data Marketplaces	Industrial Data Platforms	Trusted Third Parties	Data Collaboratives (B2B Data Sharing)	Data Trusts (B2B Data Sharing)	Other B2B Data Sharing Operators
Dawex	Mindsphere (Siemens)	International Data Spaces Association	Industrial Data Space Project (German	OpenCorporates	Ocean Protocol

---

[127] SWD(2017) 2, p. 19
[128] https://medium.com/streamrblog/what-are-data-unions-how-do-they-work-which-ones-can-i-use-887e67fb7716
[129] http://opendatamanchester.org.uk/2015/04/14/open-data-cooperation-building-a-data-cooperative/ &
https://medium.com/@opendatamcr/open-data-cooperation-building-a-data-cooperative-264eef373b63
[130] https://theodi.org/article/what-is-a-data-trust/

			Federal Ministry of Education and Research-BMBF) - International Data Spaces Association		
DataPace	Skywise (Airbus)	Smart Connected Supplier Network (SCSN)	Amsterdam Data Exchange (AMDEX)	Truata	Refinitiv
Streamr	RIO (Traton Group)		DeepMind & NHS Machine Learning for Health		Meeco.me
OpenDataSoft	Predix (GE Digital)		Data and Analytics Facility for National Infrastructure (DAFNI)		
Databroker DAO	FieldView		Smart Connected Supplier Network (SCSN)		
Rocketgraph	Xarvio		SmartFactoryKL		
Smart Jobs S.L	Data Intelligence Hub (Deutsche Telekom)				
Spaziodati	Radianz (BT Group)				
WhoApi	Nallian				
City Context Open Data API	AutoSar				
Datalayer	FiWare				
DataScouts	Far-edge				
dmi.io	Arrowhead				
GLOBMOD					
Helix Nebula Science Cloud					
Open Corporates					
gDatum					
Advaneo					
Caruso					
The IOTA Foundation					
Kasabi					
Datafairplay					

Table 11 - Data Intermediaries - C2B Data Sharing European Market Overview

PIMS/PDS	Data Unions	Data Cooperatives	Data Trusts (C2B Data Sharing)	Data Collaboratives (C2B Data Sharing)	Other Personal Data Operators
Digi.me	Streamr	MiData	UK Biobank	SalusCoop	Meeco.me
Mydex	The Data Union	SalusCoop		Copenhagen's City Data Exchange	Vastuu Group
CitizenMe	Swash	Holland Health Data Cooperative		Grampian Data Safe Haven (DaSH)	Peercraft
Datawallet	Tracey Project - TX/ WWF Philippines/ UnionBank/ Streamr Partnership	The Good Data Cooperative		Consumer Data Research Centre	Criteo
Schluss		Polypoly		Decode	Worker Info Exchange
Qiy Foundation		Healthbank Cooperative		Next Generation Internet - Engineroom	Digita
Polypoly					Datavillage
Solid Inrupt					Happy-Dev
BitsaboutMe					Ontola
Coelition					1001 Lakes
Comuny GmbH					Business Finland
Cozy Cloud					Caelum Labs
Datafund					City Of Oulu
DataYogi					Conseils Oy
esatus AG					de Volksbank
Ockto B.V.					Fair & Smart
OwnYourData					MyLife Digital
iGrant.io					Diabetes Services ApS
					Demos Helsinki
					Enfuce
					Electronic Frontier Finland

2.4.2 **The problem, its magnitude and the stakeholders affected**

The lack of a certification framework for data intermediaries, or more generally of mechanisms to differentiate neutral data intermediaries from the others leads to two main clusters of problem that are coming to prominence. The first cluster of problems involve misuse and overuse of data; the second set of problems involves underuse of data.[131] In both cases, this further leads to a generic lack of trust between

---

[131] https://medium.com/@vincejstraub/the-new-ecosystem-of-data-trusts-36901fc59010

the actors involved in the data intermediaries' ecosystem. As a final consequence, a fair and well-functioning market level playing field at European level is not ensured.

Additionally, according to the summary report of the open public consultation on the European strategy for data, almost 80% of the 512 respondents to the question have encountered difficulties in using data from other companies. These difficulties relate to technical aspects (data interoperability and transfer mechanisms), denied data access, and prohibitive prices or other conditions considered unfair or prohibitive. A very large share of respondents (87.7%) supported the idea that the EU should make major investments in technologies and infrastructures that enhance data access and use, while giving individuals as well as public and private organisations full control over the data they generate. Around the same proportion of respondents considered that the development of common European data spaces should be supported by the EU in strategic industrial sectors and domains of public interest. [132]

2.4.2.1 Estimation of Stakeholders affected

The wide definition of data intermediaries used for this study, and their several different categories constitute difficult the calculation of the total number of stakeholders affected. An estimation of the total number of data intermediaries active in the European market could include an average number of 150 organisations, while the number of data users or data holders affected could entail any European company or individual wishing to buy or sell data through the intermediaries.

The companies present big differences in the scale of client base. In particular, Siemens' Mindsphere counted more than 6.100 customers in March 2020; the client base of the late-stage Dawex includes approximately 10,000 organisations; the example of the data trust UK Biobank holds data from about 0.5m people and it includes the number of 946 researchers using its data in its annual accounts of 2018. This would therefore give a ratio of roughly 50000:1:1000 (data holders : data intermediary : data re-users). At the same time, there are several data intermediaries of early or growth stage in the European market with a client base of less than 100 clients.

2.4.3 The causes of the problem

Currently, different rules and legislation might apply to data intermediaries in Europe, depending on their category, country of establishment, sector of activity, functionalities offered and use cases and type of data handled. This might often create legal uncertainties and generate burdens to the cross-border activities of data intermediaries. Furthermore, given that the appearance of the majority of data intermediaries has recently happened within the last decade, such companies, being still in early or growth stage, might lack incentives to align on best practices. Finally, there is also a lack of mechanisms for data intermediaries to assess the quality and neutrality of intermediaries' operations, creating a further lack of trust within the market. These barriers might create difficulties in the establishment a common certification framework of data intermediaries, covering all types and market needs.

2.4.4 The effects of the problem

The lack of a certification framework for data intermediaries and its interwoven lack of trust between the actors involved in this ecosystem presents various effects and impacts for the stakeholders affected. The intermediaries respecting already certain neutrality requirements present no competitive advantage in the market compared to the others due to the lack of mechanisms for their clients to assess their neutrality. Furthermore, there are currently no mechanisms that could support such data intermediaries to scale up, as many of them are in early or grow stage. As a broader impact, the economic and societal value of data is not maximised in the European market.

---

[132] https://ec.europa.eu/digital-single-market/en/news/summary-report-public-consultation-european-strategy-data

An overview of the above-mentioned intervention logic is presented in the following table.

Table 12 - Intervention logic for Task 1.4

Measure	Barriers	Problem	Broader Impact
Establishing a certification framework for data intermediaries	Different rules applying to data intermediaries depending on sectors and types of data handled (creating legal uncertainty)	A fair and well-functioning market level playing field is not ensured, due to the lack of trust between the actors involved in the ecosystem, which does not allow data intermediaries to scale up.	Economic and societal value of data is not maximised
	Lack of mechanisms for data intermediaries clients to assess the quality of intermediaries' operations (i.e. in terms of respect of GDPR or other legislation)		
	Lack of incentives for data intermediaries to align on best practices		
	Different categories, business models, functionalities offered and use cases of data intermediaries active in the European market, creating difficulties in the establishment a common framework covering all types.		

3 Policy objectives and policy options

This chapter contains a description of the policy objectives, which could be pursued in relation to the barriers, problems and effects identified above. It also presents a list of relevant policy actions.

3.1 Policy objectives

The general objective of this initiative is to **set the foundations of a Single Market for Data**. This will contribute to **maximising the potential of data for the EU economy and society**, in particular through the empowerment of the individuals and businesses with respect to the use data they generate and create value for society. This vision will be implemented through the creation of **common European data spaces** in strategic sectors and domains of public interest, and will contribute to a more **rapid recovery** from the current economic crisis.

Across spaces, companies, public sector bodies, researchers and individuals themselves should be able to use data, personal as well as non-personal data, irrespective of the sector, domain or Member State, **in line with EU rules and fundamental values**, in particular personal data protection, consumer protection legislation and competition law.

On the global stage, the Single Market for Data will **increase Europe's sovereignty** on data and on all the key enabling technologies and infrastructures that are essential for the data economy. It will underpin a new **European approach** to data as an alternative to the platform model.

This Single Market for Data is an important element that will **complete the EU internal market**, increasing growth and jobs, modernising public services, empowering citizens to exercise their rights, and **accelerating innovation** as data is more widely used for the **common good**.

To reach this general objective, this initiative has **three specific objectives**:

- Creating **trust** in common European data spaces;
- Building **common data spaces**, making more data usable where data holders could agree to it through technical, legal and organisational support; and
- Ensuring data **interoperability** across sectors.

Figure 2 – Objective tree
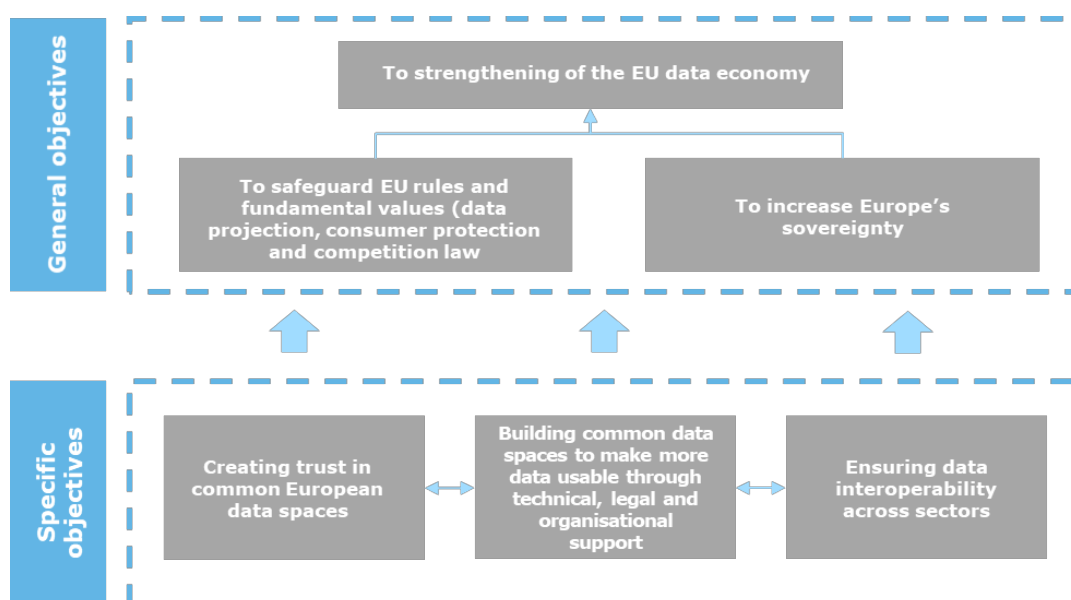
In line with these objectives, there are three levels of data governance addressed by the first phase of this study:

Table 13 – Levels of data governance

Levels	Subtask 1.1	Subtask 1.2	Subtask 1.3	Subtask 1.4
Trust in common data spaces	X	XX	X	XX
Reusable data (technical/legal)	XX	X	XX	X (certification of neutrality of intermediaries)
Cross data space interoperability	X	X	XX	X

These layers can be considered subsequent levels of enhanced abstraction and scalability of providing an enabling environment for data use in the data economy.

Trust in common data spaces deals with ensuring data is available for reuse. This mainly deals with ensuring that the appropriate management of the rights of different stakeholders. Whether data is held by public sector, citizens or businesses, this layer deals with the appropriate rights of data processing. For personal data, including sensitive data, it concerns making sure that consent and other forms of legitimate data access and reuse are ensured.

The second level concerns scalable, reusable tools for data sharing. That means, in addition to the availability of data, ensuring that there are common rules and methods for accessing and reusing. This covers standards for data schemes, for metadata, for architectures, for consent sharing, for the certification of data intermediaries. There are many such examples within different sectors.

The third level is more general and abstract. It refers to ensuring interoperability rules and standards for reuse among a wide variety of actors and use cases, across different sectors.

3.2   Policy options

Policy options have been developed in close cooperation between the consortium and DG CNECT. The policy options are listed per subtask below:

3.2.1   Policy options Subtask 1.1: Measures facilitating secondary use of sensitive data held by the public sector

The table below lists the policy options developed for Subtask 1.1.

Table 14 - Policy options Subtask 1.1

Option	Description
Option 0	Baseline scenario - No horizontal action at EU level
Option 1	Coordination at EU level and soft regulatory measures only: Guidelines
Option 2	Regulatory intervention with low intensity: One-stop shop
Option 3	Regulatory intervention with high intensity: Single data authorisation body

3.2.1.1   Policy Option 0: Baseline scenario- No horizontal action at EU level

In the baseline scenario, **no horizontal action is taken at EU level** on data governance and interoperability of common European data spaces. However, action may be taken at sectoral or Member State level as announced in the European Strategy for Data. This will lead to further interoperability issues and regulatory fragmentation in the internal market. Only certain cross sector data sharing will happen in limited cases between those common European data spaces that have compatible sectoral legislation, standards and infrastructures. Ultimately, there will be less data available for reuse across sectors. This will prevent the EU from reaping the full benefits of horizontal data sharing which account for 20% of all the benefits of data sharing in general. In addition, this situation could result in unnecessary duplication of efforts (and costs) among, for instance, data holders in different Member States in the setting up of data sharing infrastructures.

3.2.1.2   Policy Option 1: Guidelines

The first policy option would consist of issuing **non-binding Recommendations or guidelines** encouraging Member States to set up structures enhancing the re-use of publicly held data subject to the rights of others such as rights under GDPR, but also intellectual property rights, and legitimate interests to keep commercially sensitive information private. These Recommendations or guidelines would identify best practices (for instance, FinData, the French Health Data Hub or the German Forschungsdatenzentren) and promote their emulation by Member States. In addition, a network of data sharing experts would be set up as an informal Commission Expert Group issuing technical guidance for cross-border and cross-sectoral data sharing – for instance on interoperability issues, generic standards or metadata descriptions.

This option would contribute to some of the action's objectives: setting up structures enhancing the reuse of data held by the public sector and the use of which is subject to the rights of others may facilitate (and therefore likely increase) the re-use of data by companies, and researchers in line with applicable legislation. In addition, an Expert Group working on cross-border and cross-sectoral data sharing would be a first step towards interoperability across sectors.

3.2.1.3 Policy Option 2: One-stop shop

The second policy option would consist of a Directive or a Regulation requiring Member States to set up and/or maintain **capacity and services to facilitate the re-use of publicly held data** subject to the rights of others. These services would notably include a **one-stop shop** that would:

- Orientate re-users to the relevant data holders (i.e. provide information and guidance to re-users on whom to speak to); and
- Provide technical and legal advice to data holders on the permissible uses of such data and on de-identification of data.

Member States would be required to set up **secure processing environments** for the reuse of data the use of which is subject to the rights of others. Member States would have the possibility to either set up a single data processing environment, or to mandate each data holder to maintain its own.

This policy option would **not entail a right for re-users to access publicly held data** subject to the rights of others. However, there would be a best effort obligation to support innovative uses of sensitive public sector data. The use of these services and of that data would be limited to **entities established in the EU,** and potentially to entities located in third countries offering comparable mechanisms (whether this is the case in a third country would be determined by the Commission). Lastly, exclusive arrangements for data not covered by the Open Data Directive would be prohibited to ensure a **level playing field** among re-users.

3.2.1.4 Policy Option 3: Single data authorisation body

The third policy option would consist of a Directive or a Regulation requiring Member States to set up a **single data authorisation body** that would:

- Assess and grant (or reject) data re-use requests on behalf of data holders (although this may legally require the centralisation of different registers);
- Provide a secure data processing environment and data analytics tools for the re-use of publicly held data subject to the rights of others;
- Put re-users in contact with data holders (i.e. provide information and guidance to re-users on whom to speak to); and
- Provide advice to data re-users on the procedures to request a data re-use permit and on the likelihood of success of such requests.

These data authorisation bodies should not diverge excessively across the EU, for instance in terms of operational procedures and basic rules on accessibility of data. In addition, the re-use of data would not be limited in terms of purpose, and so **commercial purposes would be allowed** also. Lastly, exclusive arrangements for data not covered by the Open Data Directive would be prohibited to ensure a **level playing field** among re-users.

This policy option would neither, however, require Member States to reorganise competences internally among different data holders, nor make legislative changes to rules on secondary use of data. Member States would remain free to organise their registers, responsibilities and the grounds on which data can be reused as they see fit.

This policy option would contribute to the policy objectives of this action: as for PO2, it would facilitate (and therefore likely increase) the re-use of data by companies, and researchers in line with applicable legislation, and may contribute – through the support and advice it would provide – to creating trust between re-users and holders. In addition, by providing a secure data processing environment and data analytics tools, it

would ensure interoperability of data across sectors, thus making data more usable and contributing to building common data spaces.

3.2.2 Policy options Subtask 1.2: Establishing a certification/authorisation scheme for data altruism mechanisms

The table below lists the policy options developed for Subtask 1.2.

Table 15 - Policy options Subtask 1.2

Option	Description
Option 0	Baseline scenario- no horizontal action at EU level
Option 1	Coordination at EU level and soft regulatory measures only
Option 2	Regulatory intervention with low intensity
Option 3	Regulatory intervention with high intensity

3.2.2.1 Policy Option 0: Baseline scenario- No horizontal action at EU level

The baseline scenario assumes costs and benefits of the future if the situation was to remain the same as it is today. In the current situation, there is no European data altruism scheme. This implies that each Member State may explore different possibilities to enable data altruism individually or in cooperation with other Member States. This includes different infrastructures and approaches including legal and governance aspects in Member States. The discussion around the future role of data donors in the management of their data to support data exchange between different parties for a variety of purposes, among them research, public policy usage, public access to official documents or generally to increase efficiency and save transaction costs in one or more specified contexts, is fragmented as well. The European Data Strategy issued in February 2020 by the European Commission highlights the importance of a unified and coherent approach towards a shared data economy. There are several problems resulting of the status quo. First, the lack of European alignment on data altruism scheme leading to multiple and independently developed schemes that could face interoperability issues in the future. Second, a fragmented approach will lead to regulatory fragmentation in the internal market, where data sharing for altruistic motives will be limited to a multitude of separated silos, each acting as isolated data spaces with compatible legislation, standards and infrastructure, but without any realistic option for data to break out of that silo. Third, the rapid growth of data production and sharing which – without a common approach – cannot benefit the public sector. Fourth, the lack of a data donations for research purposes that could hamper innovation, including the development of for example AI, and impact other sectors in the European Union with, consequently, , a negative impact on the EU competitive advantage. Overall, this could prevent the EU from reaping the full benefits of horizontal data sharing which account for 20% of all the benefits of data sharing in general[41].

3.2.2.2 Policy Option 1: Coordination at EU level and soft regulatory measures only

The European Commission and Member States could explore mechanisms that encourage Member States to collaborate on efforts to sharing of personal data[42]. To facilitate this, the European Commission could adopt a Recommendation or guidelines, with no binding power, to address coordination and cooperation issues with regards to data altruism schemes and ethical guidelines on data use, considering (where applicable) the *Ethics guideline for trustworthy AI,* the "Ethics of information and communication technologies" opinion of the European Group on Ethics in Science and New Technologies, or the European Code of Conduct for Research Integrity, and the GDPR (among other authoritative sources). The recommendations could address MS to establish structures to increase reuse of publicly held data, to support data altruism, and to create processes to lower the transaction cost of data sharing. As a supporting initiative, the European Commission can also set up an informal data sharing expert group, coordinated by the European Commission. This group would be tasked with issuing more detailed guidance on multiple topics (e.g. metadata, semantics,

pseudonymisation techniques, equal and non-discriminatory access, the role and rights of the individual, compatible further processing, etc.); all of these guidelines could increase cross-border data sharing on a voluntary basis. Participation in this expert group would be voluntary and open for Member States and subject experts. The expert group could also assess if Member States are interested in trainings and funding, and provide proposals for such further support mechanisms.

Voluntary coordination could be organised at a general and horizontal level – i.e. focusing on the definition of universal principles for data altruism that would be valid independent of the sector or nature of the data – or could be integrated at a vertical sector, thus taking into account the specificities and sensitivities that may be present in individual situations.

3.2.2.3 Policy Option 2: Regulatory intervention with low intensity

The European Commission is actively engaging with the European community to advance the data economy and define a path towards a data market. To achieve this the European Commission could pursue regulatory intervention with low intensity such as mandating Member States to establish legislation and or administrative processes that allow data altruism within the Member State, without constraining them too much with respect to the practical approach to be followed. Furthermore, the Commission could (i) oblige Member States to set up certification schemes for data altruism mechanisms and/or organisations offering such mechanisms, (ii) such certification could be issued by private certification bodies under a specific Member State mandate and (iii) encourage voluntary certification of data altruism schemes. These measures would provide for a stronger and more homogeneous governance layer over the related data altruism schemes, thereby providing a more forceful and consistent response to some of the challenges described above. The responsibility to oversee this certification process would be of the Member States.

3.2.2.4 Policy Option 3: Regulatory intervention with high intensity

In 2016, the European Commission implemented the General Data Protection Regulation (GDPR) to protect, amongst others, citizens' data against unlawful (re)use. It contains safeguards against re-use of personal data for different purposes than those which were originally communicated to the data subject (i.e. the citizen); this principle can act as a complicating factor for data altruism, as has already been explained above.

The European Commission has increased the availability and re-usability of public and publicly funded data into the scope of the Open Data Directive[43] and while this Directive is likely conducive to supporting data altruism, it leaves the safeguards and constraints of the GDPR intact. A directive or regulation regarding data sharing (including but not necessarily limited to personal data) could facilitate data altruism, in several ways. One approach could be to introduce a tightly limited mandatory European authorisation mechanism for altruism schemes in relation to certain types of data (e.g. data generated or collected using government funding) or for certain purposes (e.g. donation of certain medical records to academic research institutions under specific constraints). Such authorisation would be issued under the auspices of a dedicated national authority, with mutual recognition mechanism between Member States. In some circumstances, it could be made compulsory to certify data altruism mechanisms and/or organisations operating such mechanisms. The responsibility to oversee this certification process would be on the Member States.

Alternatively, a more open approach could be considered, e.g. by establishing a governance structure at the national or EU level that would generically permit data altruism schemes to be established provided that certain safeguards are met. These safeguards could include the establishment of independent supervisory bodies and/or monitoring bodies to oversee compliance with the schemes and their use in practice; certification of schemes and/or technologies, platforms or infrastructures which would be used for data donations; codes of conduct that beneficiaries of such schemes (i.e. the recipients of donated data) or intermediaries in charge of a scheme or a technology would need to sign up to in order to become eligible

for data altruism; and/or the establishment of auditing and verification mechanisms accompanied by credible sanctions in case of violations of the framework. Such a framework could facilitate data altruism by providing greater clarity and legal certainty on the conditions for lawful data altruism, including the role and rights of the donors and other stakeholders. Scoping is however critical in legislative interventions: without suitable constraints, mandatory data donation would likely lead to citizen objection. The Commission should therefore consider how to protect citizens' personal data in altruism schemes by defining data ethics requirements.

3.2.3 Policy options Subtask 1.3: Establishing a European structure for certain governance aspects of data sharing

The table below lists the policy options developed for Subtask 1.3.

Table 16 - Policy options Subtask 1.3

Option	Description
Option 0	Baseline Scenario.
Option 1	Coordination at EU level and soft regulatory measures only – Informal Expert Group
Option 2	Regulatory intervention with low intensity: Formal Expert Group - European Data Innovation Board
Option 3	Regulatory intervention with high intensity: Independent European body - European Data Innovation Board

The main barrier to business data sharing lies in the lack of interoperability and scalable trust mechanisms. Simply put, companies are reluctant to share data because of the risks and the lack of control that it entails. And even when they are willing to do it, they often lack interoperable schemes and protocols to exchange data with other companies.

Solving the barriers of interoperability and trust at scale is the objective of a wide range of measures with different degrees of maturity, as illustrated below.

Option 0 includes no action and maintaining the baseline scenario. Option 1, 2 and 3 refer to the creation of an entity (informal, formal or with legal personality) to support data standardisation. The difference lies in the formal requirements and level of engagement, but the type of initiatives covered are similar in the different options:

1.3.0. Sector based data standards, such as those developed within different sectors. This is a mature area and falls outside of the scope of the policy intervention, but is added here to clearly distinguish it from the following three points

1.3.1. Metadata standards on findability and data quality for machine learning, such as those bring developed by AboutML

1.3.2. Data sharing schema such as iShare, IDSA and IHAN

1.3.3. Prioritisation of standardization for data use across sectors/data spaces (interoperability), which at the moment is fairly limited but could be similar to the FAIR principles applied beyond science and potentially include data standards, metadata and data sharing schemes.

In particular, **sectoral data standards** address how data should be formatted and made available to third parties in order to be fully reusable. This is the most basic issue, related to technological interoperability. It is designed to assure that when data are shared, the reuser can immediately make use of them without additional effort and, in the case of open standards, independently from the hardware and software used. Data standards have been in place for decades in different sectors, they have well defined process for

standard setting and as such they fall beyond the scope of this initiative, with the exception of cross-sector standards which is part of the last point.

Metadata standards refers to metadata that allow data reuse, both in terms of standardised schema for metadata and standardised conventions on how to describe individual metadata elements (field values), and to the proper reuse, namely in order to understand the limitations and the origin of the data, as well as their purpose of use. This is particularly important when it comes to machine learning, where the data used as an input directly affect the judgements performed by the algorithm. There are many ways to provide this information, from Microsoft datasheet for dataset to Google descriptive fiche which includes many criterion precisely designed to make the dataset bias transparent and manageable. The AboutML initiative is precisely designed "to develop, test, and implement machine learning system documentation practices at scale".

Data sharing schemes are more complex arrangements around data sharing. If data and metadata standards are designed to allow data sharing to happen when companies decide to do it, data sharing schemes aim to increase their propensity to share. They do so by ensuring data sovereignty of the business, reducing the mistrust and reassuring companies about the risks of data sharing – as well as the consent of the individuals. Concretely, this takes the shape of a series of **technical arrangements** and **legal protocols** on top of the data that describe "how to use the data", in terms of provenance, management of consent of the different parties, purpose and limitations of use, as well as tracking of who accessed the datasets for what purposes. Sector-based initiatives such as IHAN, iShare and IDS reference architecture aim to provide a frictionless and scalable way to create multilateral agreements among companies to reuse the data. They basically set up standards on protocols to document, manage and track consent (IHAN), on legal agreements about what data are owned, accessed and shared by whom (iShare), on technical architectures that ensure data sovereignty at every stage of the data value chain and at every data sharing point, for instance through the IDS connector.

Interoperability across sectors aims to define principles and frameworks for interoperability that allow data sharing across sector, including all of the three points above. In this case, the technological and legal challenges are far greater, hence the need for more abstract principles that can ensure the interoperability between the standardization initiatives taking place within the sector, so that they do not constitute de facto barriers to cross sectoral data sharing.

All these initiatives are addressing long standing and well known problems. Data sharing is one of the oldest issues in computer sciences. To develop a technical format, a metadata scheme or to define legal agreements between two companies for sharing data is time consuming but ultimately a matter of costs. But the major difference is that to grasp present opportunities **data sharing has to happen at a scale and speed never seen before**. It is entirely another matter when the companies involved are tens, hundreds or thousands. These metadata standards, data sharing schemes and interoperability are designed to enable the scaling up of data sharing beyond bilateral relations, allowing for data reuse for the widest set of purposes and for serendipitous innovation. Schemes such as IHAN, iShare and IDSA reference architecture are designed to make these legal and technical agreement as much "plug and play" as possible to facilitate deployment at scale and reduce transaction costs.

In other words, such standards and schemes allow for "many-to-many" network effects in data reuse. IDSA refers explicitly in their white paper to the analogy of peer-to-peer communication. Following the analogy, the beneficial effects of the wide adoption of such protocols could be compared to the benefits of TCP/IP or HTTP.

With regard to the stated activities, the entity should aim:

- To work with data users to capture, understand and address current and emerging standards requirements, and share best practices
- To facilitate an effective method of forming and running collaborative special interest groups and new standards initiatives.
- To work with data holders and intermediaries to develop consensus and facilitate interoperability, to evolve and integrate data specifications
- To offer a set of guiding principles and guidelines to enhance operational efficiency towards data interoperability
- To raise awareness about successful data sharing schemes that can eventually scale-up and widely facilitate data sharing

3.2.3.1 Policy Option 0: Baseline scenario - No horizontal action at EU level

In the baseline scenario, no horizontal action is taken at European level on data governance and interoperability of common European data spaces and data standardization. Yet, actions may be undertaken at national and sectorial level as announced in the European Strategy for Data.[133]

This policy option would rely on industry led initiatives such as iShare, AboutML and IDSA, on national or sectoral initiatives without any guidance or orchestration at European level. As stated before, the traction of these initiatives is only emerging and the level of data reuse today remains far below optimal.

3.2.3.2 Policy Option 1: Coordination at EU level and soft regulatory measures only

The first policy option would consist of EU coordination and soft measures, which have been used in the area of data sharing over the past decade. Until present, it is estimated that the impact of coordination and soft policy measures is limited. Under this first policy scenario, the European Commission would adopt a recommendation or guidelines with no binding power to address the different problems identified in section 2.3.

The recommendation would suggest to the Member States to set structures in place to support processes that can help lower the transaction costs of data sharing. This scenario would also create a network of data sharing experts as an *informal Expert Group* of the European Commission. This group would be tasked with issuing technical guidance for cross-border and cross-sectoral data sharing such as on interoperability issues, generic standards or metadata descriptions.

3.2.3.3 Policy Option 2: Regulatory intervention with low intensity

The second policy option would consist of creating a European Data Innovation Board. The board would be a coordination mechanism at European level that would take the form of a *formal Expert Group or a scientific committee* set by legislation, hosted by the European Commission. The functions of the Expert Group would be limited to the general technical guidance on issues related to data standards, data specifications, metadata, ontologies or findability.

3.2.3.4 Policy Option 3: Regulatory intervention with high intensity

The third policy option would consist of a European Data Innovation Board. However, under this option the board would be an *independent European body* with legal personality and supported by a secretariat. This body would be inspired by the structure and operational characteristics of European Data Protection Board (EDPB). The functions of such board would be of low intensity and specific mandate for the accreditation of certification schemes for data intermediaries.

---

[133] COM Strategy data

3.2.4 Policy options Subtask 1.4: Establishing a certification framework for data intermediaries

The table below lists the policy options developed for Subtask 1.4.

Table 17 - Policy options Subtask 1.4

Option	Description
Option 0	Baseline scenario- No horizontal action at EU level
Option 1	Coordination at EU level (industry driven self-regulatory certification framework)
Option 2	Regulatory intervention with low intensity (voluntary certification framework)
Option 3	Regulatory intervention with high intensity (compulsory certification framework)

3.2.4.1 Policy Option 0: Baseline scenario- No horizontal action at EU level

In the baseline scenario, **no horizontal action is taken at EU level** on regulation of European data sharing platforms and interoperability of common European data spaces. In the current situation, there are no specific regulatory and non-regulatory actions taken at the EU level targeting data intermediaries or data sharing platforms, and therefore no certification framework established for data intermediaries in the European market. However, action may be taken at sectoral or Member State level. This might lead to further interoperability issues and regulatory fragmentation in the internal market. Cross-sector data sharing will happen only in limited cases between those common European data spaces that have compatible sectoral legislation, standards and infrastructures. This will prevent the EU from reaping the full benefits of cross-sector data sharing, which account for 20% of the benefits of data sharing in general[134].

3.2.4.2 Policy Option 1: Coordination at EU level (industry driven self-regulatory certification framework)

As a policy option that aims to promote coordination at EU level, the European Commission could ask private operators, such as representatives of data intermediaries (providers of data sharing services) active in the European market, industry associations and certification bodies to create a network of data sharing experts as an informal expert group or stakeholder forum of the European Commission. This would enable the stakeholders involved to coordinate, exchange and present their views and experience on the topic, aiming to align on best practices and the way forward. An outcome of these discussions could be the creation of an **industry-driven, self-regulatory code of conduct** by the stakeholders.[135] This would not be a compulsory regulatory measure but it would remain at the choice of the data intermediaries to decide whether they would like to sign it or not. The code of conduct could further lead to the development of **self-regulatory certification scheme-** by the data intermediaries. The certification scheme would include parameters agreed by the industry that help bring trust to data intermediaries offering data sharing services in B2B contexts and/or personal data spaces, by ensuring that the certified intermediaries function as "neutral intermediaries". These could include, among others, rules regarding specified sources of data, the nature of the intermediary, its business model and the service offered, compliance with legislation, cybersecurity measures, transparency and non-discrimination in data sharing (non-discrimination might not apply for certain type of data, i.e. in cases of criminal activities or poor data quality, but it will be important to be transparent on that). This industry-driven approach to establish a self-regulatory certification framework could entail the finance of a private certification agency, while the government role would be

---

[134] Realising the economic potential of machine-generated, non-personal data in the EU, Deloitte Report for Vodafone Group, July 2018
[135] Similar to the self-regulatory code of practice on Disinformation, agreed by online data platforms in 2018

limited by participating as an observer or providing guidance. Finance from the public sector actors for the self-regulatory certification might also be available in the case that it is deemed needed.

Similar efforts for the establishment of a self-regulatory certification framework are already in place at the European and international level, within the MyData Community[136] and the NYU GovLab. In particular, a "self-description" process has been initiated within the MyData Community, targeting organisations that have signed the operator interoperability MoU. This is a voluntary, self-description process to allow operators how show their services meet the MyData human-centric criteria as described in the Understanding MyData Operators white paper[137]:

- to demonstrate alignment with the MyData principles. In the future, the development seems to be towards governed ecosystems and thus more neutral operators;
- to describe the systems for personal data management with respect to the MyData operator reference model;
- to show that they follow the two criteria of transparency and the person as the primary beneficiary.

As a result, in July 2019 16 organisations from 12 countries, who are working for human-centric approaches to personal data, were awarded the inaugural status of MyData Operator 2020.[138] Furthermore, the NYU GovLab has also developed a list of "Trusted Intermediaries" for data collaboration, for third-party actors support collaboration between private-sector data providers and data users from the public sector, civil society, or academia.[139] Finally, a certified data pool list is available in the frame of Global Data Synchronization Network (GDSN) for computer systems exchanging information through data pools, enabling collaborators to operate based on standards that support live data sharing and trading updates.[140]

3.2.4.3 Policy Option 2: Regulatory intervention with low intensity (voluntary labelling framework)

A policy option for the low intensity regulatory scenario could entail the adoption of a legislative/regulatory measure establishing a **voluntary labelling framework** for novel data intermediaries which would allow them to function as organisers/orchestrators of data sharing or pooling within such spaces and to obtain a label/kitemark/seal. This could be implemented by the means of a legislative act adoption (regulation or directive) and further developed by a delegated act, defining in detail the core criteria and certification requirements, that should be met by all labelled intermediaries in order to demonstrate their neutrality and absence of conflict of interest, in particular the absence of competition with data users (providers of services seeking to use data shared by data holders). The adoption of the legislative act would not alter –in substantive law- the rights and obligations of persons and organisations on data, but would establish a structural enabler encouraging data sharing through data intermediaries. The aim would be to lower data sharing transaction costs, bring trust among stakeholders in the data sharing market or pooling within the common European data spaces, in light of the current distrust in platform business models and the limited brand recognition of the novel services providers that are emerging. The certification criteria might entail softer "neutrality" requirement for B2B data intermediaries providing data-sharing services addressing business users and handling exclusively industrial data (e.g industrial data platforms, data marketplaces, trusted third parties, data collaboratives, data trusts, data trusts), compared to C2B data intermediaries or "personal data spaces", addressing individuals. The certification criteria would be stricter for data intermediaries dealing with consumer's personal data (e.g PIMS/PDS, data unions, data trusts, data cooperatives, data collaboratives), as neutral operators of personal data spaces should limit themselves to

[136] https://mydata.org/about/
[137] https://docs.google.com/document/d/1e3hvYSgsNas8ZWW3HXvg5V9a_H6r0AlpvaCC9HWvvQw/edit
[138] https://mydata.org/2020/07/29/press-release-mydata-operator-2020-status-awarded-to-16-organisations-from-around-the-world/
[139] https://datacollaboratives.org/explorer.html?#trusted-intermediary
[140] https://www.gs1.org/services/gdsn/certified-data-pools-list

data sharing services only and consent management (not added value services based on the data) and have fiduciary duties towards the individuals using them. The ambitions for this policy option would entail: a) a quick applicability process after the adoption of the legal instrument setting the criteria, b) verification by the data permit authorities and c) strict deadlines for receiving the results of the verification process. The potential role of Member States's governments to set up the process will have to be examined, in line also with the policy options developed under the other sub-tasks of this study, particularly under 1.1, as the handling of the application process and the awarding of the labels/certification would be done by the one-stop shop mechanisms set up by Member States which would also handle requests regarding the reuse of public sector data. The certification criteria could include among others:

- Strict notion of function structural separation: Structural separation of data intermediation services from both data holders and potential data users: Data intermediation services may not propose any service building on the data transacted [alternatives possible: merely legal or functional separation];
- Questions of data dominance and ownership as well as fair and non-discriminatory access to the data intermediation service for both data holders and data users.
- Data intermediaries' establishment in Europe: Providers of data sharing services offering services to business users shall have their principle place of business within the European Union.

In terms of international data flows, it is proposed that providers of data sharing services need to take adequate organisational and legal measures to prevent that jurisdictional decisions of third countries that would require access to data relating to European companies and individuals would take effect without making recourse to mutual legal assistance request that would ensure European jurisdictional control over these decisions. In practice this may mean that global players need to create legal entities in Europe that are entirely separated from the corporate structure in the third country, including at the level of ownership (cf. previous collaboration between Microsoft and Deutsche Telekom). In addition to the above, for C2B data intermediaries it is proposed not to add elements to the existing adequacy decision regime of the GDPR.

3.2.4.4 Policy Option 3: Regulatory intervention with high intensity (compulsory certification framework)

A policy option for the high-intensity regulatory scenario could entail the establishment of a European **mandatory certification framework** for all types of data intermediaries. Similarly to the previous policy option, this could be implemented by means of a legislative act adoption (regulation or directive) defining hard neutrality requirements and criteria to be respected by both B2B and C2B data sharing platforms. In this case, the certification would be compulsory for all the data sharing platforms in order to ensure the compliance of their activities with the specific provisions defined by the regulatory measure. Neutral operators of personal data spaces should only offer data sharing services and consent management (and no added-value services based on the data), while B2B data intermediaries may offer additional data sharing services, but subject to conditions of structurally separating data intermediation services from other services. Certification would be awarded by private conformity assessment bodies, based on criteria developed at the European level. Such bodies would be accredited by the European Data Innovation Board. This policy option might also require, without being necessary, a level of Member States' governmental involvement and responsibilities (e.g MS to financially support the private conformity assessment bodies, provide guidance and overseeing of the certification process).

Similar efforts to regulate data sharing platforms have been conducted also at the international level with particular examples in the US with the adoption of the Data Broker List Act of 2019,[141] in Japan with the Release of the Guidelines of Certification Schemes Concerning Functions of Information Trust ver. 1.0,[142] only for C2B data intermediaries in both cases, as well as in India. At European level, other types of certification frameworks, (including i.a. GDPR certification and cybersecurity certification) have been used in the past to ensure trust in certain markets and provide an added-value to the companies.

---

[141] https://www.govtrack.us/congress/bills/116/s2342
[142] https://www.meti.go.jp/english/press/2018/0626_002.html

4 Assessment of the policy options

This chapter presents the assessment of the policy options per subtasks identified in the previous chapter with regard to their effectiveness, efficiency and coherence and who will be affected.

4.1 Introduction

This chapter presents our draft assessment of the impacts of all the options, including the baseline scenario.

The following assessment criteria were agreed on for the assessment of the impacts of the options:

- **Effectiveness** in achieving the policy objectives:
  - Achievement of specific objectives;
  - Achievement of general objectives;
- **Efficiency**:
  - Costs of the option;
  - Benefits of the option, including reductions in some of the costs as well as other positive effects on (some of) the stakeholders;
- **Coherence** of the option.

Proportionality and legal/political feasibility criteria will be also considered when comparing the policy options.

To the extent possible, the assessment is built on **quantitative and qualitative information, including costs and benefits**. For this purpose, we took various data sources into account for the assessment of the impacts, including:

- Desk research, including a legal analysis;
- Interviews;
- Workshop (Subtask 1.1 and 1.3)[143];

The study team aimed at collecting as comprehensive quantitative data as possible. However, consulted stakeholders and pre-existing studies only provided data for some types of costs and benefits. In this D1 report, illustrative examples of quantitative and qualitative feedback from stakeholders with regards to costs and benefits of each policy options per subtasks have been included.

4.2 Subtask 1.1: Measures facilitating secondary use of sensitive data held by the public sector

This section assesses the baseline and three policy options for Subtask 1.1.

---

[143] Results to be included in D2

4.2.1 Stakeholders affected

The following table provides an overview of the key stakeholders affected by the possible policy options and how:

Table 18 – Overview of stakeholders affected by subtask 1.1 policy options

Who?	How?
Data holders	Data holders would in essence have a reduced range of tasks to perform when it comes to sensitive data reuse. Indeed, a number of functions currently performed by (most) data holders would be centralised under both policy options 2 and 3 (the latter entailing a larger number of such tasks that would be centralised). As a result, data holders should see a significant reduction in their running costs. Indirectly, data holders which also serve as decision-making bodies (such as national ministries) or which offer public services to citizens (such as in the public healthcare sector) would benefit from new insights generated by research reusing sensitive data. This could lead to more effective and/or efficient policy-making, and concrete benefits in health, such as lower costs, higher efficiency, better treatments, and lives saved.
Data intermediary	In all likelihood, public sector data intermediaries would be the actors taking on the role of one-stop shop under policy option 2 or data authorisation body under policy option 3. This will result in increased costs linked to these additional tasks – although ultimately, these costs would be borne by data re-users and/or by the taxpayer.
Data (re)users	Data re-users would in essence see their activities facilitated under policy options 2 and 3, since many tasks currently performed by a range of different actors would be centralised. As a result, transaction costs associated with having to deal with a range of actors would be greatly reduced, resulting in time savings. In parallel, data re-users would in all likelihood pay for the service performed by either the one-stop shop under policy option 2 or the data authorisation body under policy option 3. Whether these costs will be lower than costs currently incurred will depend on the specific case.
Society	Overall, society would benefit from greater access and re-use of sensitive data. Indeed, new insights generated from research would in theory lead to more effective and/or efficient decision-making in a range of domains, including health, social affairs, transport, and the environment. In addition, individual citizens will have greater control over the re-use of their data through increased transparency.

4.2.2 Policy option 0: Baseline

4.2.2.1 Effectiveness in achieving the policy objectives

This subsection examines the effectiveness of a baseline scenario in achieving the policy objectives.

4.2.2.1.1 Achievement of specific objectives

In the absence of EU action, **Member States would remain free to take their own approach** with regards to the re-use of data held by public bodies and the use of which is subject to the rights of others. Uncertainty with regards to applicable rules and legislation would likely continue in some Member States, and only some Member States would likely take steps towards interoperability of data cross sectors. Yet, over 75% of respondents to the Open Public Consultation on the European strategy for data believe that public authorities should do more to make a broader range of sensitive data available for research.[144]

As a result, it is **uncertain whether data held by the public sector and the use of which is subject to the rights of others would be generally more available for reuse**. Reusers would therefore be unable to increase their use of such data for research and development or new business opportunities, while policy-makers would not benefit from improved input to guide their decisions. The development of Artificial

---

[144] European Commission, Summary report of the public consultation on the European strategy for data. See: https://ec.europa.eu/digital-single-market/en/news/summary-report-public-consultation-european-strategy-data

Intelligence (AI) would not benefit from improved access to data the use of which is subject to the rights of others, and would therefore be impeded.

Likewise, **interoperability issues** across sectors and Member States **would likely persist**, causing reusers to continue spending unnecessary time pre-processing (i.e. pseudonymising and anonymising) the data in order to combine it. In the absence of further reuse of their data, data holders would have **no incentive to ensure their data is of the highest possible quality** and accuracy. Fragmentation as regards access to, and combination of data of sufficient quality would continue. Thus, imbalances would persist between reusers with the resources to overcome these issues and reusers without such resources.

Citizens wishing to **exercise their rights under the GDPR** – for instance, retracting their consent for their data to be reused – would continue facing **opaque and/or cumbersome** procedures for doing so in some Member States, with potentially **negative consequences for fundamental rights and for trust** in reuse of data the use of which is subject to the rights of others (and thus in common European data spaces).

This all results in a **limited positive economic impact** overall, particularly in terms of:

- Time and resources spent by data re-users as a result of these issues;
- Duplication of time and resources spent by some data holders to provide data the use of which is subject to the rights of others ; and
- Absence of gains due to increased re-use of data, increased innovation and new business opportunities (and therefore growth and competitiveness), and potentially better economic policies.

Furthermore, the absence of better policies resulting from better information would **limit the potential for positive social and environmental impacts**. This would be compounded by the duplication of re-use mechanisms for data the use of which is subject to the rights of others existing within a single Member State – resulting in a duplication of energy-intensive IT infrastructures enabling such reuse.

4.2.2.1.2 Achievement of general objectives
Absence of EU action would not contribute to setting the foundations of a Single Market for Data, since Member States would set their own policies. While the data economy of some Member States would likely be strengthened as a result of national policy, this would not be the case across the EU. This fragmentation would not contribute to increasing the EU's sovereignty on data and on the key enabling technologies and infrastructures, and neither would it contribute to completing the internal market.

4.2.2.2 Efficiency: Costs and benefits of the option
This subsection presents the costs and benefits associated with a baseline scenario.

4.2.2.2.1 Costs of the option
Data holders and intermediaries would continue to bear a number of costs in some Member States, namely:

- The costs of establishing and maintaining specific reuse mechanisms for data the use of which is subject to the rights of others when applicable, including the associated IT infrastructure;
- The costs of pseudonymising and anonymising data prior to making it available;
- The costs of examining applications for data access;
- The costs linked to training sufficient staff to perform these tasks;
- The opportunity costs, for data holders, linked to not accessing more research made available by data reuse; and
- Potential costs linked to data breaches.

For instance, the statistical office of a mid-sized EU Member State has approximately two FTEs working on pseudonymising and anonymising data, representing a cost of at least EUR 140,000 per year (for 70-100 requests a year); while the statistical office of a large Member State has approximately 10.6 FTEs working on tasks related to making data the use of which is subject to the rights of others available for re-users (including handling applications and pre-processing data, excluding IT costs that represent 20% of staffing costs).

Data reusers would continue to bear a number of costs in some Member States, namely:

- Time and resources spent on identifying the data holder holding the desired data;
- Time and resources, including of a pecuniary nature, spent on producing and submitting different (and not always successful) applications to access data from different holders;
- Time and resources spent combining data which is not necessarily interoperable.

For instance, a private sector data re-user in a large Member State estimates that the time spent on one data request application is equivalent to roughly five to 15 days of effective work depending on the complexity of the application.

4.2.2.2.2 Benefits of the option, including reductions in some of the costs as well as other positive effects on (some of) the stakeholders

In the absence of EU action, no particular benefits were identified for stakeholders across the EU.[145]

4.2.2.3 Coherence of the option

This policy option does not entail any piece of legislation which might be incoherent with other policy options. However, no action would be incoherent with the policy priorities identified.

4.2.3 Policy option 1: Guidelines

This section assesses the first policy option for Subtask 1.

4.2.3.1 Effectiveness in achieving the policy objectives

This subsection examines the effectiveness of policy option 1 in achieving the policy objectives.

4.2.3.1.1 Achievement of specific objectives

The **extent to which this policy option contributes to the specific objectives of the action is contingent on the degree to which Member States decide to follow the Commission's Recommendations or guidelines**. At any rate, this policy option is highly unlikely to be detrimental to the objectives. Indeed, recommendations to set up structures enhancing the re-use of publicly held data, if followed, would likely result in an increase in access and re-use of data the use of which is subject to the rights of others – contributing to **building common data spaces**.

These recommendations could have **positive economic, social and environmental impacts** (due to increased research and improved decision-making, on which more below), and **enhance trust** as a result of increased transparency, if:

- They recommend best practices such as a one-stop shop (policy option 2) or a single data authorisation body (policy option 3), and
- They are followed by Member States.

---

[145] There would however be benefits for stakeholders in the Member States that have domestically implemented something similar to policy options 2 or 3.

In addition, setting up an informal Commission Expert Group issuing technical guidance for cross-border and cross-sectoral data sharing – for instance on interoperability issues, generic standards or metadata descriptions – **may on the long term lead to increased interoperability** across sectors and across Member States. This would depend on the extent to which this guidance is implemented.

4.2.3.1.2 Achievement of general objectives

The extent to which this policy option contributes to the general objective of the action is contingent on the degree to which Member States decide to follow the Commission's Recommendations or guidelines. In the event that they are, the increased data access and re-use, trust and interoperability would contribute to setting the foundations of a Single Market for Data through the creation of common data spaces.

4.2.3.2 Efficiency: Costs and benefits of the option

The non-binding recommendations/guidelines encouraging Member States to set up structures enhancing the re-use of publicly held data subject to the rights of others were discarded from the CBA and subsequent macroeconomic analysis for several reasons.

- First, several stakeholders expressed **doubts as to their overall effectiveness**, noting that recommendations and guidelines on data sharing and reuse abound but are **not always followed**. At the validation workshop organized on 8 July, stakeholders present indicated they **expect a third to half of Member States to implement such guidelines**.
- Second, stakeholders interviewed and present at the workshop indicated that the **level of ambition of such guidelines or recommendations would likely be inversely proportional to the number of Member States adopting them**. This is because the more ambitious the guidelines, the more effort and resources would be needed for their implementation.

As a result, the study team estimated that policy option 1 would have a limited effectiveness, while any measure of its efficiency would be over reliant on assumptions linked to the content and uptake of such recommendations or guidelines.

4.2.3.3 Coherence of the option

The study team did not find any incoherence of this option with existing legislation.

4.2.4 **Policy option 2: One-stop shop**

This section assesses the second policy option for Subtask 1.

4.2.4.1 Effectiveness in achieving the policy objectives

This subsection examines the effectiveness of policy option 1 in achieving the policy objectives.

4.2.4.1.1 Achievement of specific objectives

The establishment of a once-stop shop would contribute to achieving the specific objectives. Indeed, providing information and guidance to data holders may incentivise them to make more data the use of which is subject to the rights of others available for access and re-use, and may increase demand for data the use of which is subject to the rights of others due to increased transparency, particularly for smaller re-users. This potential increase in available data the use of which is subject to the rights of others would **contribute to building common data spaces**, while increased fairness resulting from equal access to information and guidance would be a **positive social impact**.

This potentially increased demand for, and re-use of, data the use of which is subject to the rights of others would translate into improved and increased research, and therefore into better policy-making resulting in positive **economic, social and environmental impacts**. This would be in line with the results of the OPC,

in which 91,5% of respondents agreed that "more data should be available for the common good, for example for improving mobility, delivering personalised medicine, reducing energy consumption and making our society greener."[146]

Furthermore, this option would foster **trust through transparency** between data re-users and data holders, as well as trust among the general public – particularly if the one-stop shop provides legal guidance to citizens on how to exercise their rights under data protection laws. This is consistent with the results of the OPC: 84,6% of respondents believe that it should be made easier for individuals to give access to existing data held on them, in line with the GDPR.[147]

4.2.4.1.2   Achievement of general objectives
An increase in the re-use of data held by the public sector and the use of which is subject to the rights of others, contributing to building national common data spaces, complemented by greater trust in how sensitive (and particularly personal) data will lead to **more integrated national markets for data**. This is a significant **first step setting the foundations of a Single Market for Data**.

4.2.4.2   Efficiency: Costs and benefits of the option
This subsection presents the costs and benefits associated with policy option 2.

4.2.4.2.1   Costs of the option
The establishment of a one-stop shop **would entail costs for data re-users**, in the form of potential fees to fund the one-stop shop's operations – resulting in benefits neutering the costs for the one-stop shop.[148] These costs for data reusers are expected to be neutered – or potentially outweighed – by the benefits incurred by access to new data.

It would entail **costs for data holders**, in terms of additional staff required to cope with potential additional demand for data the use of which is subject to the rights of others (as a result of it being more easily accessible). In addition, coordinating and liaising with the one-stop shop on a regular basis would cost time and resources to data holders. For instance, the statistical institute of a large Member States estimates that each data holder in that country spends about two weeks of effective work per year coordinating with the country's one-stop shop. At the same time however, data holders would gain time from the amount of work (in, for instance, answering queries from data reusers) that would performed by the one-stop shop instead of them.

Establishing a one-stop shop **may also entail costs for society** – specifically, for taxpayers – if fees charged to data re-user do not cover all the costs associated with the one-stop shop. For instance, the one-stop shop of a large Member State employs 8 FTEs.

The figures in the table below originate from interviews with several stakeholders, including the German Data Forum (RatSWD) that acts as a one-stop shop in Germany, one of Germany's accredited Research Data Centres (acting as a data holder), and Statistics Denmark (for the cost linked to maintaining a secure data processing environment). Specifically, as regards costs associated with the one-stop shop, RatSWD's running costs are currently EUR 900,000 per year,[149] while a secure data processing environment costs on average roughly EUR 610,000 per year – thus amounting to an overall annual cost of approximately EUR 1,510,000. On the other hand, the German Forschungsdatenzentrum estimates that each data holder spends roughly 2

---

[146] European Commission, Summary report of the public consultation on the European strategy for data. See: https://ec.europa.eu/digital-single-market/en/news/summary-report-public-consultation-european-strategy-data
[147] *Ibid.*
[148] In addition, Member States could establish mechanisms for distributing these benefits among data holders. This however would be at the discretion of Member States.
[149] While some of these costs may be covered by user fees, this is currently not the case in Germany with RatSWD and may not be the case in all Member States.

weeks of work annually preparing and attending meetings at RatSWD. One hour of work is assumed to cost EUR 45 on average, while two weeks of work are assumed to correspond to 80 hours, which corresponds to a total of EUR 3,600 per annum.

These figures are conservative estimates: participants to the workshop held on 8 July 2020 indicated that the running costs of a one-stop shop are likely to be higher than EUR 900,000, and that data holders would likely spend a little more than 2 weeks per year coordinating with the body.

In addition to these, the study team assumed costs for reusers (in the form of fees payable to the one-stop shops) of EUR 500 per application. This assumption is based on the current fees charged by Findata (Finland's single data authorisation body). Since the one-stop shops would perform fewer tasks than a data authorisation body – in particular, they would not handle data access applications – these fees are assumed to approximate half of the fees charged by Findata. The actual amount however would vary depending on the Member State.

The extrapolation of these figures to the EU level can be found in the table below, while the full overview of costs (and benefits, including the cost benefit analysis) is in Annex C.

Table 19 – Overview of costs for subtask 1.1 | PO 2

Overview of costs (EUR) – PO 2

		Data holders		Once-stop shop (data intermediary)		Data re(users)	
		One-off	*Recurrent*	*One-off*	*Recurrent*	*One-off*	*Recurrent*
Measures facilitating secondary use of sensitive data held by the public sector	*Direct costs*	-	7.6 million p.a.	286.4 million	16.5 million p.a.	-	41.8 million p.a.
	Indirect costs	-	-	-	-	-	-

4.2.4.2.2 Benefits of the option, including reductions in some of the costs as well as other positive effects on (some of) the stakeholders

The establishment of a one-stop shop in each Member State would result in a variety of **benefits for data reusers**:

- Time and resources saved when identifying the data holder holding the desired data (these are assumed to be equal to 20 hours of work per application, or EUR 590);
- Increased fairness in access of data held by the public sector and the use of which is subject to the rights of others, i.e. all re-users would have equal access to valuable information on how to access that data;
- This improved access to data the use of which is subject to the rights of others would likely result in an increase in re-use of the data, since potential reusers currently not reusing data the use of which is subject to the rights of others may start doing so once it is easier;
- Access to legal guidance potentially resulting in saving time and resources related to legal training; and
- Time and resources saved by accessing already interoperable data across sectors.

It would also entail potential **benefits for data holders**, in terms of:

- Access to legal guidance potentially resulting in saving time and resources related to legal training;
- Access to technical guidance on how to allow data reuse, resulting in a decreased risk of data breach and the associated costs;
- Time and resources saved by not providing, and maintaining, a secure data processing environment;

- As noted above, such an environment costs on average EUR 610,000 p.a.
- Access to an increased amount of research resulting from an increased demand for data the use of which is subject to the rights of others – leading to better policy-making;
- Increased visibility; and
- Decreased difficulty in cross-border access to data.

This policy option would entail **benefits for society** more broadly:

- A one-stop shop would increase transparency regarding re-use of data the use of which is subject to the rights of others, which may contribute to increased trust;
- This increased transparency – along with additional research resulting from additional data re-use – would likely improve decision-making, with economic, social and environmental benefits for society.

Lastly, the one-stop shop is assumed to charge EUR 500 per application to cover (part of) its running costs.

These figures are extrapolated to the EU level in the table below, while the full overview of benefits (and costs, including a cost-benefit analysis) is in Annex C.

Table 20 – Overview of benefits – Subtask 1.1 | PO 2

Type of action	Description	Amount (EUR)	Stakeholders
Measures facilitating secondary use of sensitive data held by the public sector	**Direct benefits**		
	Resources saved as a result of not maintaining secure processing environment and analytics tools[150]	*684 million p.a.*	**Data holders**
	Direct revenues (fees)[151]	*41.8 million p.a.*	**Data intermediary**
	Time/resources saved as a result of easier data discovery[152]	*49.2 million p.a.*	**Data re-users**
	Impact on policymaking and decision-making	*Not quantifiable due to lack of data*	**Data holders and (re)users**
	Potential new scientific insights with positive outcomes on research and innovation	*Not quantifiable due to lack of data*	**Data (re)users**
	New economic base	*Not quantifiable due to lack of data*	**Data intermediaries**
	Cost savings and revenue generation from results created by data use. Possibility to enter new business sectors, research fields, generation of new correlation of data, which enables new insights.	*Not quantifiable due to lack of data*	**Data (re)users**
	Indirect benefits		

---

[150] This is based on the assumption that 20% of data holders would relinquish their data processing environment to use the environment established as part of the policy option, and that 30% of the data pre-processing and provision work would no longer be done by data holders but by the one-stop shop.

[151] This is based on the assumption that the one-stop shops would charge on average EUR 500 per application. This would cover part of the one-stop shops' running costs.

[152] This is based on the assumption that each data re-user would save about 20 hours of work per application.

Effect on Gross Domestic Product (GDP) Innovation and competitive advancement	*Not quantifiable due to lack of data*	**Data holders**
New insights	*Not quantifiable due to lack of data*	**Data intermediaries**

4.2.4.2.3 Findings of the Cost-Benefit Analysis

This policy option, although it carries a large one-off cost for the establishment of the one-stop shops, brings equally large – and recurrent – benefits that greatly outweigh the recurrent costs associated with the one-stop shop's operation. The Cost-Benefit Analysis conducted as part of this study (and which can be found in Annex C) indeed finds a benefit-cost ratio (BCR) of 9.2. In other words, benefits incurred by this policy option are expected to be over nine times larger than costs.

4.2.4.3 Coherence of the option

A one-stop shop performing discovery and advisory services is coherent with existing legislation as well as with ongoing efforts by some Member States to set up such one-stop shops.

4.2.5   **Policy option 3: Single data authorisation body**

This section assesses the third policy option for Subtask 1.

4.2.5.1 Effectiveness in achieving the policy objectives

This subsection examines the effectiveness of policy option 1 in achieving the policy objectives.

4.2.5.1.1 Achievement of specific objectives

The establishment of a single data authorisation body would contribute to achieving the specific objectives. Indeed, guaranteeing a single, streamlined application process for accessing data the use of which is subject to the rights of others, and providing a secure data processing environment to reusers, would consist of the technical, legal and organisational support needed for data holders to make more data usable – contributing to **building common data spaces**. This would be reinforced by the likely increase in demand for data the use of which is subject to the rights of others resulting from the process of accessing it being easier. This would also ensure more fairness in access to data the use of which is subject to the rights of others, since all re-users, small and large, would have equal access to information on how to access data and to data analytics tools – a net **positive social impact**.

This increased demand for, and re-use of, data the use of which is subject to the rights of others would translate into improved and increased research, and therefore into better policy-making resulting in positive economic, social and **environmental impacts**. In addition, **positive economic impacts** would be reinforced by the non-duplication of efforts by public sector data holders to make data available for re-use.

The single data processing environment and the upstream work conducted by the single data authorisation body would **ensure interoperability** of the data across sectors within a given Member State, and could be a stepping stone towards interoperability across Member States.

This option would foster **trust through transparency** between data re-users and data holders, as well as trust among the general public that their data is re-used following a single, streamlined procedure managed by a publicly **accountable** body. However, this impact may be counterbalanced by reduced trust from the public as a result of sensitive (and particularly personal) data being available for reuse for strictly commercial purposes.

66

The single data permit authority of a mid-sized Member State expressed doubts as to the feasibility of this option in all Member States, pointing to different levels of governance centralisation and to the centrality of trust among citizens that their data will be re-used in a secure way and for the purposes to which they have agreed. Trust related issues from citizens have already prevented the establishment of a single data permit authority in one Member State.

4.2.5.1.2 Achievement of general objectives

An increase in the re-use of data held by the public sector and the use of which is subject to the rights of others, contributing to building national common data spaces, complemented by greater interoperability within Member States and greater trust in how sensitive (and particularly personal) data will lead to **more integrated national markets for data**. This is a significant **first step setting the foundations of a Single Market for Data**.

4.2.5.2 Efficiency: Costs and benefits of the option

This subsection presents the costs and benefits associated with policy option 3.

4.2.5.2.1 Costs of the option

The establishment of a single data authorisation body **would entail costs for data re-users**, in the form of potential fees to fund the single data authorisation body's operations.

It would entail **costs for data holders**, in terms of staff required to coordinate and liaise with the single data authorisation body on a regular basis.

Establishing a single data authorisation body **may also entail costs for society** – specifically, for taxpayers – if the fees it charges do not cover the body's running costs.[153] For instance, the single data authority of a mid-sized EU Member States has an overall budget of EUR 5.2 million this year (expected to decrease), following one-time R&D costs of approximately EUR 10 million (corresponding to a pilot project). It will employ an estimated 25 FTEs once fully running, which each FTE costing approximately EUR 75,000 and requiring one to two weeks of training. This single data authority has had an approximate 7 FTEs specifically working on 38 data access applications since early May 2020.

In addition, enabling the re-use of data the use of which is subject to the rights of others, and especially of personal data, for **strictly commercial purposes would raise ethical questions and may undermine citizens' trust** – which might lead to a reduced amount of personal data available to reuse (due to potentially fewer citizens agreeing to the reuse of their data), in turn negating the positive impacts of increased reuse of data the use of which is subject to the rights of others. In addition, it may contravene national laws of several Member States where for instance statistical microdata may only be reused for research purposes.

The table below provides an overview of annual costs associated with this policy option, based on Finland's single data authorisation body for health and social data, Findata. Prior to Findata's establishment, the Isaacus pilot project was launched with a EUR 14 million budget, of which approximately 75% (i.e. EUR 10.5 million) were directly linked to Findata's establishment. The running costs of Findata, once it is fully up and running (i.e. in a couple of years), are estimated to range between EUR 4 and 5 million per annum. Currently, Findata charges reusers a fixed fee of EUR 1,000 per application for a new permit, and an additional EUR 115 per hour worked on pre-processing and combining datasets from different holders. Findata expects to handle an average of 600 applications per year, while one Finland-based re-user interviewed indicated they

---

[153] The legislative design would suggest that the fees charged to reusers could cover entirely the cost of operating the single data authorization body. However, the distribution of costs linked to the operation of the single data authorisation body, between re-users (in the form of fees) and the taxpayer (in the form of taxes), would be left at the discretion of Member States.

submit applications regarding around 30 projects per year (i.e. 30 applications via Findata, assuming they all concern data from multiple holders). It is currently expected that on the medium term, Findata will receive approximately **EUR 1 million each year from public funds**, with the remaining costs being covered by user fees. To avoid double-counting the running costs of Findata that will be covered by user fees, the EUR 1 million figure is used as an estimate of running costs while the remaining costs are counted as costs for reusers (in the form of the user fees).

These estimates may be conservative, as participants to the workshop organised on 8 July estimate that both the one-off R&D costs and the recurrent running costs of a single data authorisation body may be somewhat higher. However, Findata's figures themselves may be higher than they will be in the future, as Findata was established very recently and has only begun its operations. Indeed, Findata representatives indicated that the figures are based only on a few months of operation, and that more accurate numbers should be available by the end of 2020 at the earliest. Nevertheless, the figures are extrapolated in the table below, while a full overview of costs (and benefits, including a cost benefit analysis) is in Annex C.

Table 21 – Overview of costs for subtask 1.1 | PO 3

Overview of costs (in EUR) – PO 3		Data holders		Single data authorisation body (data intermediary)		Data re(users)	
		One-off	Recurrent	One-off	Recurrent	One-off	Recurrent
Measures facilitating secondary use of sensitive data held by the public sector	*Direct costs*	-	-	572.7 million	329.7 million p.a.	-	212.7 million p.a.[154]
	Indirect costs	-	-	-	-	-	-

4.2.5.2.2 Benefits of the option, including reductions in some of the costs as well as other positive effects on (some of) the stakeholders

The establishment of a single data authorisation body in each Member State would result in a variety of **benefits for data reusers**:

- Increased fairness in access of data held by the public sector and the use of which is subject to the rights of others, i.e. all re-users would have equal access to valuable information on how to access that data, as well as equal access to data processing tools;
- This improved access to data the use of which is subject to the rights of others would likely result in an increase in re-use of the data, since potential reusers currently not reusing data the use of which is subject to the rights of others may start doing so once it is easier;
- Access to legal guidance potentially resulting in saving time and resources related to legal training;
- Time and resources saved when identifying the data holder holding the desired data and by handing in one data access request application for access to data from more than one holder, instead of having to apply several times for one project;
- Time and resources saved by virtue of the application process being accelerated – since centralising this process would result in economies of scale; and
- Time and resources saved by accessing already interoperable data across sectors

---

[154] This corresponds to the fees payable to the single data processing authority. However, this fee replaces pre-existing fees payable to the different data holders currently processing requests. As a result, according to stakeholders, the costs for data re-users would be lower under PO3 than in the baseline – but current fees could not be quantified by the study team.

For instance, three data reusers estimate that having to deal with a single body as opposed to multiple data holders would save about half the time spent on a typical application (down from between 50 to 90 hours to 25 to 45 hours, with each hour assumed to be worth EUR 45, i.e. cost savings in the order of EUR 1.125 to EUR 2.025 for each application). During the 8 July workshop however, participants estimated that these savings may be even higher.

One stakeholder estimates that not having to pre-process data from different holders (since the single data permit authority will have done so) would save them several days of work each time[155] – although this varies and could not be confirmed with other stakeholders (one of which prefers continuing to perform this task to avoid missing potential insights from the raw data).

It would also entail potential **benefits for data holders** resulting from the centralisation of services in the data authorisation body and the above-mentioned economies of scale. Specifically, these would be:

- Time and resources gained as a result of:
  - Not processing data access applications;
    - This cost is estimated to approximate EUR 400,000 per annum for Statistics Denmark, although participants to the 8 July workshop believe this figure would be lower;
  - Not pre-processing the data and providing, and maintaining, a secure data processing environment and data analytics tools;
    - These costs are estimated to approximate EUR 1,200,000 (pre-processing and combining data) annually for Statistics Denmark, although participants to the 8 July workshop estimate that the figure would in reality be lower. The costs of maintaining a secure processing environment are estimated to approximate EUR 610,000.
- Access to an increased amount of research resulting from an increased demand for data the use of which is subject to the rights of others – leading to better policy-making.

Lastly, this policy option would entail **benefits for society** more broadly:

- A data authorisation body would increase transparency regarding re-use of data the use of which is subject to the rights of others, which may contribute to increased trust;
- This increased transparency – along with additional research resulting from additional data re-use – would likely improve decision-making, with economic, social and environmental benefits for society; and
- The benefits for data holders would translate into public savings, and therefore either reduced tax or increased spending in other areas.

These figures are summarised in the table below, while a full overview of benefits (and costs, including a cost-benefit analysis) is in Annex C.

Table 22 – Overview of benefits – Subtask 1.1 | PO 3

Type of action	Description	Amount (EUR)	Stakeholders
Measures facilitating secondary use of sensitive data held by the public sector	**Direct benefits**		
	Resources saved as a result of not processing data access applications	*569.4 million* p.a.	**Data holders**
	Resources saved as a result of not pre-processing and combining datasets[156]	*512.5 million* p.a.	**Data holders**

---

[155] The stakeholder was unable to give a precise number.
[156] This is based on the assumption that 30% of the data pre-processing and combination work would no longer be done by data holders but by the data authorisation body.

Resources saved as a result of not maintaining secure processing environment and analytics tools[157]	*171.5 million p.a.*	**Data holders**
Resources saved as a result of not having to submit separate applications for one research project	*167.1 million p.a.*	**Data re-users**
Direct revenues (fees)	*212.7 million p.a.*	**Data intermediary**
Impact on policymaking and decision-making	*Not quantifiable due to lack of data*	**Data holders and (re)users**
Potential new scientific insights with positive outcomes on research and innovation	*Not quantifiable due to lack of data*	**Data (re)users**
New economic base	*Not quantifiable due to lack of data*	**Data intermediaries**
Cost savings and revenue generation from results created by data use. Possibility to enter new business sectors, research fields, generation of new correlation of data, which enables new insights.	*Not quantifiable due to lack of data*	**Data (re)users**
Indirect benefits		
Effect on Gross Domestic Product (GDP) Innovation and competitive advancement	*Not quantifiable due to lack of data*	**Data holders**
New insights	*Not quantifiable due to lack of data*	**Data intermediaries**

4.2.5.2.3 Findings of the Cost-Benefit Analysis

This policy option carries benefits that are over twice larger than those associated with Policy Option 2 – in line with its greater ambition and with its higher potential to realise the policy objectives of this intervention. At the same time, it incurs much higher costs – over eight times higher – than Policy Option 2. This is due to the much higher recurrent costs associated with operating a single data authorisation body compared to a one-stop shop, and also reflects the higher level of ambition of this policy option. On balance however, the BCR for this policy option is 2.4 (see Annex C for the full CBA). While positive, this ratio is significantly lower than for Policy Option 2.

4.2.5.3 Coherence of the option

Setting up a single data authorisation body is coherent with existing EU policies, and complements Directive (EU) 2019/1024 on open data and the re-use of public sector information (PSI Directive) as it would cover sensitive data currently excluded from the PSI Directive. It is coherent with the GDPR and indeed facilitates the exercise of one's rights under it. However, opening up the reuse of data the use of which is subject to

---

[157] This is based on the assumption that 20% of data holders would relinquish their data processing environment to use the environment established as part of the policy option.

the rights of others to private companies for exclusively commercial purposes would contradict national law in Member States when only researchers are allowed to access and re-use for instance statistical microdata. In addition, a single data authorisation body similar to Findata, which would give access to data the use of which is subject to the rights of others on behalf of data holders, may be incoherent which national laws in at least two Member States stating that only the data holder may provide access to the data it holds for re-use.[158]

4.2.6 Summary of the impacts

The following table summarises the possible impacts of the policy options:

Table 23 – Summary of impacts for subtask 1.1

Economic impacts	- Costs for public authorities (one-stop shops/data authorisation bodies) - Direct (R&D, staff, premises, equipment etc.) - Indirect (coordination with data holders) - Benefits for public authorities (data holders) - Direct (time and resources saved) - Indirect (new insights from research leading to better decision-making) - Costs for data re-users - Direct (fee for the use of one-stop shop/data authorisation body) - Benefits for data re-users - Direct (time and resources saved, more equal access) - Indirect (increased re-use, better quality data) - Benefits for society - Public savings resulting in better spent taxpayer money, or reduced tax
Social impacts	- Fairer access to, and re-use of, data held by the public sector and the re-use of which is subject to the rights of others - Increased trust among society about the re-use of their data - Better policy-making due to new insights
Environmental impacts	Better policy-making due to new insights potentially leading to more effective environmental policies
Fundamental rights impacts	Increased transparency as regards reuse of personal data (e.g. related to health) and how to exercise one's rights

4.3 Subtask 1.2: Establishing a certification/authorisation scheme for data altruism mechanisms

4.3.1 Stakeholders affected

The following table provides an overview of the key stakeholders affected by the possible policy options and how:

Table 24 – Overview of stakeholders affected by subtask 1.2 policy options

Who?	How?
Data holders	Data subjects would have more transparency on how their data is being processed and for e.g. companies or organisations that hold data and plan to make it available through data altruism schemes, a certification scheme would be an added benefit. Overall, data altruism schemes, including certifications, would decrease privacy concerns and increase transparency as well as have an added authentication value. It could also enable an equal playing field for also small or medium-sized companies that were previously excluded from data altruism due to high legal or ICT costs.
Data intermediary	Data intermediaries are public sector bodies, organisations and companies. In specific cases this could be research institutes who are hosting data altruism mechanisms for specific research purposes directly, however often a third party - intermediary - will be involved.

---

[158] Furthermore, one stakeholder insisted that moving away supporting structures from the underlying data will come with a loss in the capacity to provide targeted expertise on how the specific data assets in question can be used.

	Data intermediaries would benefit from reduced costs to adapt to various data altruism schemes.
Data (re)users	Data (re)users would greatly benefit from a data altruism scheme. The more data is received, the better data (re)users can make a positive impact on the public good through for example enhanced policy making. The benefits are difficult to quantify but could be tremendous.

4.3.2 Policy option 0: Baseline scenario - No horizontal action at EU level

4.3.2.1 Effectiveness in achieving the policy objectives

This subsection examines the baseline policy option for Task 1.2

4.3.2.1.1 Achievement of specific objectives

In the absence of EU action, Member States and private entities would continue to take their own approach to data altruism. If relevant, they would potentially collaborate bilaterally or initially seek a national approach in federal Member States. Uncertainty with regards to legislation and data handling, think of interoperability and data guidelines, could lead to further fragmentation in the European Union. This would **increase costs** for Member States, companies and organisations such as NGO's because they would have to negotiate data altruism schemes with each Member State individually.

Legislative issues and **fragmentation** is costly and time consuming. Large corporations that want to engage in B2G data altruism schemes have the financial and manpower capabilities to address these issues, however SMEs and possible also NGOS and research organisations could consider this a roadblock to participate in data altruism schemes **due to uncertainty of data security** issues.

4.3.2.1.2 Achievement of general objectives

The general objective of creating a Single Market for Data would likely not be achieved very efficiently or effectively considering that every Member State would continue developing data altruism schemes according to its respective political agenda, which greatly varies across the EU27.

4.3.2.2 Efficiency: Costs and benefits of the option

4.3.2.2.1 Costs of the option

Costs for data re-users:

- Varying degrees of (macroeconomic) competitive advantages between Member States depending on the advancement of data altruism schemes and legal frameworks
- Loss of additional GDP for the European Union due to lack of coherent data altruism schemes to benefit from
- Unwillingness to share personal data by citizens and private entities companies due to lack of transparency and security issues resulting in increased awareness raising campaigns
- Political ambition, or lack thereof, to prioritize data altruism schemes could limit the creation of a EU27 data economy
- Technical infrastructure costs and interoperability issues for public bodies due to lack of guidelines for data altruism schemes
- Costs related to incorrect data management due to legislative fragmentation
- Data collection at regional level and local levels could be disrupted and additionally complicated when it is a Federal state (such as Germany or Italy) due to different laws within a single Member State which could increase costs for the public body re-using data

Costs for data holders:

- Increased time and resources spent on legislative questions, possible excluding certain Member States and SMEs from data altruism because of high costs, due to legislative fragmentation
- Increase awareness raising and transparency costs to gain consumer trust for the permission to participate in data altruism schemes with the public sector

4.3.2.2.2 Benefits of the option, including reductions in some of the costs as well as other positive effects on (some of) the stakeholders

Benefits for data-reusers

- Member States would have more time to define their data altruism scheme needs according to their national standards
- Member States would have more time to prepare citizens for data altruism schemes and increase digital skills to share data
- Lack of a common data altruism regulation and strategy could provide short-term innovation benefits to certain Member States

4.3.2.3 Coherence of the option

This policy option does not entail any piece of legislation, which might be incoherent with other policy options. However, no action would be incoherent with the policy priorities identified. Considering that data portability and privacy rights already have national legislative differences which is an issue and could become bigger through no-action.

4.3.3 Policy option 1: Coordination at EU level and soft regulatory measures only

This section assesses the first policy option for Task 1.2

4.3.3.1 Effectiveness in achieving the policy objectives

This subsection examines the effectiveness of policy option 1 in achieving the policy objectives.

4.3.3.1.1 Achievement of specific objectives

Coordination at EU level and introduction soft regulatory measures could contribute to the specific objectives of creating trust, building a common data space and ensuring data interoperability.

However, this would be to a limited extent since the policy option relies on coordination and soft regulatory measures that could be voluntary or for example an Expert Group. An Expert Group would have to be established, which takes time, and then have 18 months to work on and submit a report on a European data altruism scheme. Until these recommendation are then implemented, it could take many more months. Therefore this would be a very timely exercise. A similar length process would apply to coordination meetings to define soft regulatory measures such as guidelines of metadata or cross-border data sharing.

Most importantly, a lack of incentive or perception of necessity and benefits could deter many data re-users and holders to participate and later implement *soft measures.*

4.3.3.1.2 Achievement of general objectives

This policy option would be a step towards establishing the foundation of a Single Market for Data and establishing a common European data space. Nevertheless, due to its voluntary or soft measures it would not immediately accelerate the creation of a Single Market for Data since not all Member States would be bound to these guidelines or it would simply take time to implement them.

4.3.3.2 Efficiency: Costs and benefits of the option

Policy Option 1 was not further included in the CBA and subsequent macroeconomic analysis for the following reasons:

- Coordination at EU level would **require willingness from Member States to participate**. Some interviewed Member States stated that they are already involved in bilateral talks with other Member States, and while coordination at EU level would possibly reduce their workload, discussions around table instead of several, this would not necessarily accelerate the discussions. In addition, only the Member States that are already actively pursuing data altruism mechanisms would likely participate according to the interviewed Member States.
- The private sector raised further concerns that coordination at EU level **could take very long and yet not result in concrete action**. Considering that data altruism, and the data economy, are considered a priority, it should be addressed as such with concrete actions.
- Furthermore, adoption of the measures would be voluntary and **could widen the data altruism gap between different Member States and companies** by widening the fragmentation between those that are actively pursuing this and those that are not.

To conclude, this policy option was considered to have very limited effects, possible even **discourage certain stakeholders to participate** and coordinate due to fear of inefficiency. Indicative categories of costs and benefits are nonetheless provided below.

4.3.3.2.1 Costs of the option

Costs for data re-users:

- Costs are difficult to predict due to a lack of clarity on *soft measures*
- Could still create legislative fragmentation is certain Member States move faster or beyond the *soft guidelines*
- Willingness of Member States to participate in EU level coordination efforts
- Organisational, legal and technical costs to implement any soft measures
- Interoperability issues could persist limiting cross-border data sharing
- Transparency and trust could still be low among data holder, limiting the amount of data data re-users can reuse
- New, even soft, regulation could create additional costs if it is not build on existing regulations

Costs for data holders:

- Willingness of data holders to participate in EU level coordination efforts
- Organisational, legal and technical costs to implement any soft measures
- Costs by still facing legal and technical fragmentation in Member States
- Transparency and trust issues could persist among customers and individuals holding data

Table 25 – Overview of costs for subtasks 1.2 | PO 1

Overview of costs – PO 1

		Data holders		Data intermediaries		Data re(users)	
		One-off	*Recurrent*	*One-off*	*Recurrent*	*One-off*	*Recurrent*
A certification scheme for data altruism mechanisms	*Direct costs*		Time spent on understanding various data altruism schemes to	-	-	-	-Set-up scheme -Raise awareness of scheme

	make data available[159]					-Gain trust of data holders[160]
Indirect costs	-	-	-	-	-	-

4.3.3.2.2 Benefits of the option, including reductions in some of the costs as well as other positive effects on (some of) the stakeholders

Benefits for data re-users:

- Coordination among Member States would ensure a multi-lateral discussion about data altruism schemes at 'one' table instead of various bilateral discussion, savings time and costs
- Coordination to align on e.g. open source code and a legal basis could increase transparency and trust in data altruism. For example, Germany agreed on a coordinated agreement to request 'consent' from data holders to share their data

Table 26 – Overview of benefits – Subtask 1.2 PO 1

Type of action	Description	Amount	Stakeholders
A certification scheme for data altruism mechanisms	**Direct benefits**		
	Effect on Gross Domestic Product (GDP)	*Not quantifiable due to lack of data*	**Data holders**
	New business relationships with other stakeholders (e.g. data holders)	*Not quantifiable due to lack of data*	**Data intermediaries**
	Easy and transparent way to access data of various fields, contributing to research and development as well as improved decision-making	*Not quantifiable due to lack of data*	**Data (re)users**
	Indirect benefits		
	R&I and competition advancement Impact on policy- and decision-making	*Not quantifiable due to lack of data*	**Data holders, Data producers**
	Contribution to a societal goal If donated to research, development of new scientific insights (including medical research, leading to lives saved)	*Not quantifiable due to lack of data*	**Data donors**
	Value of personal data, assumption: the higher the amount of data, the greater the benefit Possibility to enter new business sectors, research fields, generation of new correlation of data -> new insights	*Not quantifiable due to lack of data*	**Data (re)users**

4.3.3.3 Coherence of the option

The study team is unaware of any incoherence with any existing EU or national legislation, largely due to the absence of any data altruism legislation.

---

[159] Further data collection will be performed to provide an estimate.
[160] Further data collection will be performed to provide an estimate.

4.3.4 Policy option 2: Regulatory intervention with low intensity

This section assess the second policy option of Task 1.2

4.3.4.1 Effectiveness in achieving the policy objectives

This subsection assess the effectiveness of achieving policy option 2.

4.3.4.1.1 Achievement of specific objectives

By creating a certification scheme for data altruism mechanisms and obligating Member States to implement it, the Commission would achieve to build a common data space with regard to data altruism.

Considering that the certification would be issues by a private organisation would not limit trust and the fact that it is voluntary still ensure data interoperability across sectors.

To facilitate the effectiveness of this policy option, the Commission could additionally **create working sessions** on for example data portability rights, harmonisation of standards and how to set-up and manage personal data spaces in addition to setting-up a certification scheme. This would help Member States to understand the benefits of certification schemes for data altruism mechanisms.

4.3.4.1.2 Achievement of general objectives

Considering the current crisis and focus on data altruism schemes to contain the COVID-19 pandemic, the Commission should take this opportunity to make urgent changes and align across the EU27 on data altruism schemes. Thereby it would also take a great step toward building the **Single Market for Data.**

4.3.4.2 Efficiency: Costs and benefits of the option

This subsection assess the efficiency of achieving policy option 2.

4.3.4.2.1 Costs of the option

Costs for re-users:

- The obligation to set up certification schemes will most likely be carried by the data re-users because most often the data will she made available to the public sector, who then also re-uses the data for analysis. Therefore the obligation to set-up a scheme under option 2 is carried by the public sector, who simultaneously will also re-use the data. However, it has to be noted that the costs and re-use of data could be, and most likely, will be carried by different parties in the public sector. For instance, the ministry of interior affairs could be responsible for setting-up such a scheme, whereas various research institutes then re-use the data for their analysis or research.
- Organisational, legal and technical costs would be associated with such a certification scheme

Costs for data holders and intermediaries:

- Certification will likely have associated costs for the data holders and intermediaries (if relevant), however it would be voluntary

Table 27 – Overview of costs for subtasks 1.2 | PO 2

Overview of costs – PO 2[161]

	Data holders		Data intermediaries		Data re(users)	
	One-off	Recurrent	One-off	Recurrent	One-off	Recurrent

---

[161] Further data collection will be performed to provide an estimate.

A certification scheme for data altruism mechanisms	*Direct costs*	-Giving consent to make data available - Negotiating data altruism conditions	- Giving consent to make data available (could be recurrent if it is revoked)	- Establish infrastructure to facilitate data altruism 30,000 EUR - Becoming certified	- Maintain data altruism certification	- Establishing scheme -	- Maintaining the scheme/certification -Reviewing public authorities authorized to certify schemes
	Indirect costs	-	-	-	-	-	-

4.3.4.2.2 Benefits of the option, including reductions in some of the costs as well as other positive effects on (some of) the stakeholders

Benefits for data re-users:

- Could choose to only reuse data that has been shared through certified data altruism schemes, thereby being secured that it is legally compliant to reuse the data
- Save costs because a private company would issue the certification
- The EU would be a first-mover and possibly set a global certification standard considering that this does not yet exist and non-EU27 might copy the EU, as has happened with GDPR
- Increase transparency and trust among data holders, which in return provides data re-users with larger data sets and or insights
- Reduce organisational, technical and legal costs in the long-run because one certification would eliminate the create a new process for every new data altruism project

Benefits for data holders and intermediaries:

- Increase of trustworthiness, security and possibly awareness of data altruism schemes
- A privately-run certification scheme will likely operate more efficiently thus make is more attractive for data holders and intermediaries to become certified
- In the long-term, the data altruism scheme could possible run on the voluntary scheme due to its success
- Save on legal costs to due to absence of legal fragmentation in various Member States, the certification would provide an automatic and straightforward process for data altruism schemes

Table 28 – Overview of benefits – Subtask 1.2 | PO 2

Type of action	Description	Amount	Stakeholders
A certification scheme for data altruism mechanisms	**Direct benefits**		
	Effect on Gross Domestic Product (GDP)	*Not quantifiable due to lack of data*	**Data holders**
	New business relationships with other stakeholders (e.g. data holders)	*Not quantifiable due to lack of data*	**Data intermediaries**
	Easy and transparent way to access data of various fields, contributing to research and development as well as improved decision-making	*22 Eur million*	**Data (re)users**
	Indirect benefits		

R&I and competition advancement Impact on policy- and decision-making	*Not quantifiable due to lack of data*	**Data holders, Data producers**
Contribution to a societal goal If donated to research, development of new scientific insights (including medical research, leading to lives saved)	*Not quantifiable due to lack of data*	**Data donors**
Value of personal data, assumption: the higher the amount of data, the greater the benefit Possibility to enter new business sectors, research fields, generation of new correlation of data -> new insights	*Not quantifiable due to lack of data*	**Data (re)users**

4.3.4.2.3 Findings of the Cost-Benefit Analysis

This policy option carries voluntary costs for the data intermediaries and brings a relative amount of benefits for those, EUR 22 million. The Cost-Benefit Analysis for PO2 finds that the voluntarily certification scheme has a cost-benefit ratio of only 2.3.

4.3.4.3 Coherence of the option

The study team is unaware of any relevant data altruism certification legislation, however points out the importance to align with or build on any other existing certification mechanisms to avoid conflicting mechanisms that would create additional costs for Member States and society. In addition, the study team urges the Commission to think of incentives for voluntary certification. What would encourage the private sector to voluntarily become certified? The Commission could encourage authorities to provide funding for certification or bring experts together to encourage the community to become certified. An example from the banking sector could be the PCI-Payment Card Certification- which is a voluntarily credit card verification certification that has been adopted voluntarily globally and even become contractual requirements because the financial ecosystem took responsibility to ensure consumer trust and greater transparency. Overall, this option would be aligned with the Digital Single Market and help build the European Data Economy, while still giving Member States the liberty to control the certification process and not overburdening the private sector and providing additional security measures for personal data.

4.3.5 **Policy option 3: Regulatory intervention with high intensity**

This section assesses the third policy option of Task 1.2.

4.3.5.1 Effectiveness in achieving the policy objectives

This subsection assess the effectiveness of this policy option.

4.3.5.1.1 Achievement of specific objectives

While an administrative approval scheme would increase **trust** of the society in data altruism schemes, it could slow down the building of a common data space in case the public sector does not have sufficient resources to authorise data altruism schemes. This could create a negative impact on the data economy because it could discourage data holders to seek authorisation to participate in data altruism schemes. This could then limit the **social impact** of data altruism for public good, because if less data is available for the data re-users, the lesser the impact will be.

4.3.5.1.2 Achievement of general objectives

A strong and regulated data altruism scheme could either accelerate the creation of a **Single Market for Data,** or could completely put it at a halt should Member States not have the capacity to set-up and manage an authorisation authority.

4.3.5.2 Efficiency: Costs and benefits of the option

This subsection assesses the efficiency of this policy option

4.3.5.2.1 Costs of the option

Costs for data (re)users

- High cost to create a national authorisation body
- High organisational, legal and technical costs
- High human resources costs to find and or skill people to manage data altruism schemes in the Member States
- Could lead to increased costs if the public body does not have the capacity to handle and process the certification requests
- Could limit the amount of data for re-users

Costs for data holders and intermediaries:

- Increased costs due to mandatory authorisation
- Increased costs if the data altruism authorisation process is slow or poorly managed by the public sector
- In the worst case scenario, this could deter data holder and intermediaries to contribute to data altruism schemes which would lead to less data being shared

Table 29 – Overview of costs for subtasks 1.2 | PO 3

Overview of costs – PO 3[162]

		Data holders		Data intermediaries		Data re(users)	
		One-off	*Recurrent*	*One-off*	*Recurrent*	*One-off*	*Recurrent*
An authorisation scheme for data altruism mechanisms	*Direct costs*	-Giving consent to make data available	- Giving consent to make data available (could be recurrent if it is revoked)	Becoming authorized (if applicable 3,800-10,500 EUR depending on the size of the organisation Establish scheme/authorisation process and national oversight body\n\nNon-quantifiable, however every EU27 state has a data authority (or equivalent) that could	- Maintain data altruism authorisation (if relevant) 5,000 EUR	- Establishing scheme/certification and national oversight body N.a. because this will be covered by the established authority under Task 1.1.	- Maintaining the scheme/authorisation

---

[162] Further data collection will be performed to provide an estimate.

	implement this.					
Indirect costs	-	-	-	-	-	-

4.3.5.2.2 Benefits of the option, including reductions in some of the costs as well as other positive effects on (some of) the stakeholders

Benefits for data re-users:

- Increase of trustworthiness, security and possibly awareness of data altruism schemes
- Possibly increased data sharing by SMEs, NGOs and private citizens, although this could be limited if the authorisation process is slow

Benefits for data holders and intermediaries:

- Increase of trustworthiness, security and possibly awareness of data altruism schemes
- Possibly increased data sharing by SMEs, NGOs and private citizens

Table 30 – Overview of benefits – Subtask 1.2 | PO 3

Type of action	Description	Amount	Stakeholders
A auhtorisation scheme for data altruism mechanisms	**Direct benefits**		
	Effect on Gross Domestic Product (GDP)	*Not quantifiable due to lack of data*	**Data holders**
	New business relationships with other stakeholders (e.g. data holders)	*Not quantifiable due to lack of data*	**Data intermediaries**
	Easy and transparent way to access data of various fields, contributing to research and development as well as improved decision-making	*300 EUR million*	**Data (re)users**
	Indirect benefits		
	R&I and competition advancement Impact on policy- and decision-making	*Not quantifiable due to lack of data*	**Data holders, Data producers**
	Contribution to a societal goal If donated to research, development of new scientific insights (including medical research, leading to lives saved)	*Not quantifiable due to lack of data*	**Data donors**
	Value of personal data, assumption: the higher the amount of data, the greater the benefit Possibility to enter new business sectors, research fields, generation of new correlation of data -> new insights	*Not quantifiable due to lack of data*	**Data (re)users**

The study team calculated that the high intensity option, compulsory authorisation, would lead to benefits of EUR 300 million because a compulsory certification will lead to more data altruism, due to increased trust

in the certified mechanisms, and therefore more shared data. In this calculation all 500 companies are certified and all 5 million data holders (of the ones willing to participate in data altruism) provide their data altruistically. In the first year this already leads to a benefit of EUR 50 million, the value of a data set remains EUR 10, and increases annually by 10% as was assumed for PO2 as well. This increase is again due the impact that more companies will enter the market, seeing the benefits, and more data holders will be willing to share their data considering the increased benefits reaped. Therefore PO3, the high intensity option, will achieve a EUR 300 million benefit. This benefit is so much larger, than compared to PO2, because data altruism is still very new and calls for many privacy concerns. Mandatory certification could, likely will, increase the trust of data holders to share data and more shared data, which has a market value although not being used commercially considering it is for research purpose, will lead to higher benefits. In the case of healthcare for example, more shared data could save lives or be the solution to a pandemic.

The key argument is that while the costs are initially high, the recurring costs are very low, and the benefits of the volume of data shred strongly outweigh the costs.

4.3.5.2.3   Findings of the Cost-Benefit Analysis

Policy Option 3 entails significant benefits including an acceleration of data altruism mechanism in the European Union due to the mandatory certification mechanisms. This will especially establish an increased trust into data altruism mechanism which will increase the amount of data shared by data holders. This will then lead to the substantial benefit of EUR 300 million, which translates to a Cost-Benefits Ratio of 6.3, as referenced in Annex C .

4.3.5.3   Coherence of the option

The option is coherent with the Digital Single Market strategy and aim to build the data economy, however obliging Member States to create a national authorisation authority could create additional costs for them and thereby society. In addition, it could potentially increase the burden of regulation on Member States and businesses, thereby slowing down the process of data altruism scheme implementation in Europe.

4.3.6   **Summary of the impacts**

The following table summarises the possible impacts of the policy options:

Table 31 - Summary of impacts for subtask 1.2

Economic impacts	(Depending on the shared data) • Increased GDP due to improved policy making • Fairer access to data altruism schemes • Increased transparency and security of data altruism schemes • Enable all organisations and companies to participate in data altruism, not just the firms that have sufficient funds to navigate the ICT and legislative labyrinth
Social impacts	(Depending on the shared data) • Improved social policy due to improved policy making based on data (insights) • Empower data holders
Environmental impacts	(Depending on the shared data) • Improved environmental policy making due to better data insights
Fundamental rights impacts	(Depending on the shared data) • Improved policies concerning fundamental rights due to better/more data insights • Empower data holders and increase citizens ability to make their data available for the public good

4.4   Subtask 1.3: Establishing a European structure for governance aspects of data sharing

Under Subtask 1.3, we started by identifying and selecting data points related to the indicators about the costs and benefits of deciding and implementing data standards that effectively contribute to foster data sharing within and across sectors in Europe. These indicators - linked to the causes of the cost occurring due to not sharing data - were categorised and quantified to the extent possible.

Because of the nature of the policy option, the costs and benefits are very indirectly related to the policy measure. An expert group or even the "hard" regulatory option of a legal body has very limited influence over the development and adoption of data standards, as ultimately the decision on how the standard is designed and whether to use it lies with companies.

As the problem analysis (reproduced in the table below) shows, the logical links are clear.

1. Data sharing is opening up huge opportunities for efficiency gains. In the manufacturing sector alone, data sharing accounts for 80% of the potential efficiency gains, equal to 1.4 trillion Euros by 2027.
2. There is consensus that the main barriers to data sharing are trust and interoperability. Companies do not want to lose the control over the data and are wary of potential risks from misuse. The costs for establishing ad hoc agreements is too high.[163]
3. These barrier is precisely the target of initiatives such as IDSA, IHAN and iShare. They provide scalable standardised tool to implement data sharing on a large scale while ensuring the control of the data holder over the data. These tools include technological protocols, process templates and legal agreements. However, these initiatives are at an early stage and being applied still at limited scale – there is no robust evidence about their traction and impact.
4. The four policy options do not envisage the establishment of a standard or standardisation activities, but a set of accompanying activities to support the development and adoption of data standards.

Table 32 - overview of the problem analysis

Ongoing initiatives	Causes	Problem	Effects
Standardisation and coordination initiatives	Lack of data and metadata standards , data schemes within sectors Lack of technical interoperability across sector	Lack of data sharing within/across sector	Lower productivity and innovation

Because policy option 1, 2 and 3 are similar in nature and vary only by the level of formality of the group, the impact will be similar across option. The main difference lies in:

- Increased costs for the more institutionalized options
- Increased benefits from a greater possibility to foster the adoption of standards.

The benefits framework is reproduced below.

Table 33 - overview of the cost and benefits framework

Types of stakeholders (economic impact)

---

[163] The Everis study on data sharing places technical interoperability as the most mentioned obstacle, by 73% of companies. Legal uncertainty about data ownership is the second, with 54%, and control over usage the third with 42%. The Deloitte studies reports costs of normalizing data, lack of standard protocols, cumbersome legal procedures, involuntary disclosure of commercial secrets as the main barriers. The WEF "Share to gain" report identifies standars, trust and legal arrangement as the key enablers.

Types of Impact		Data holders	Data re-user same sector	Data re-user other sectors	Data intermediaries	Society	Environment
Direct Benefits	**Costs Savings**	Easier reuse of data, lower cost of data processing and management	Easier reuse of data, lower cost of data processing and management	Easier reuse of data, lower cost of data processing and management	Easier reuse of data, lower cost of data processing and management	Easier reuse of data, lower cost of data processing and management	
	Efficiency gains	Greater access to data	Greater access to data	Greater access to data			
	Time savings	n.a					
	Resource / energy savings	n.a					
	Revenues/user charges	n.a.					
Indirect Benefits	**New products and services**	Faster development of new products and services Anticipate disruption by players from different sectors	Faster development of new products and services Anticipate disruption by players from different sectors	New markets	Greater demand for data intermediaries	New products and services	
	Digital single market	Greater access to data from other countries Reduced friction in reusing data	Greater access to data from other countries Reduced friction in reusing data	Greater access to data from other countries Reduced friction in reusing data	Greater access to data from other countries Reduced friction in reusing data		
	Other non-monetisable benefits					New discoveries for health	Environmental efficiency and new products
	Health, Safety & Security improvement						
Direct Costs	**CAPEX**			✓	Cost of setting up governance bodies	n.a	n.a
	OPEX	Cost of participating to standardization activities	Cost of participating to standardization activities	Cost of participating to standardization activities		n.a	n.a

	R&D Costs	n.a.	n.a.	n.a.	n.a.	n.a	n.a
	Implementat ion	n.a.	n.a.	n.a.	n.a.	n.a	n.a
	Training	n.a.	n.a.	n.a.	n.a.	v	n.a
	Compliance costs	Cost of complying to standards	Cost of complying to standards	Cost of complying to standards			
	Administrati ve burden	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
	Compliance costs	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
Indire ct Costs	**Digital single market**	Greater competition within and across sectors and countries	Greater competition within and across sectors and countries				

4.4.1 Stakeholders affected

The following table provides an overview of the key stakeholders affected by the possible policy options and how:

Table 34 – Overview of stakeholders affected by subtask 1.3 policy options

Who?	How?
Data holders	Companies would benefit from the possibility to share data and reap the opportunities of data driven innovation
Intermediary initiatives	European initiatives, data spaces and data standardisation initiatives would benefit from increased awareness and standardisation
Data (re)users	Same as data holders, as data sharing happen between companies in a peer to peer fashion. But with the addition of data companies who would benefit from access to data

4.4.2 Policy option 0 – no action at EU level

4.4.2.1 Effectiveness in achieving the policy objectives

4.4.2.1.1 Achievement of specific objectives

The current status is slowly progressing towards the achievement of the objectives. Traditional standardization efforts are ongoing within sectors. Several initiatives (illustrated in the problem analysis) are addressing the full architecture of data sharing, such as iShare, IHAN, IDSA. These initiatives are mostly national but are trying to grow at European level. They encounter different rates of success in achieving the buy in of business not only in the standard definition (where they managed to achieve momentum, for instance IDSA involves 100 leading EU companies) but also in the adoption of the standards.

This will also result in limited access to data, because of the additional cost of agreement on data standards and most importantly on the legal and trust framework, which have to be developed ad hoc in the absence of such protocols and standards.

According to interviewed experts, the present rate of growth in developing and adopting data schemes and standards is slow and unlikely to achieve the promised 1.4 trillion euros benefits by 2027.

4.4.2.1.2 Achievement of general objectives

Because of the slow progress towards the digital single market for data. There is a clear risk that the slow progress towards data sharing could reduce the capacity of European industry to guarantee the sovereignty over their data and standards, opening the way for competitors from third countries to enter the European market and potentially achieve a gatekeeper role for industrial data – just as platforms did for personal data.

In other words, the policy option 0 allows for a slow progress that might expose the European economy to strong risks.

4.4.2.2 Efficiency: Costs and benefits of the option

4.4.2.2.1 Costs of the option

Traditional business would spend limited time in developing standards as well as in adopting them, hence face no additional costs.

Similarly, there would be no additional costs for technological companies, intermediaries as well as for the European Commission

4.4.2.2.2 Benefits of the option, including reductions in some of the costs as well as other positive effects on (some of) the stakeholders

There would be no benefits in terms of reducing costs for data sharing and reuse. For traditional business, the current situation entails very high costs for data sharing and reusing, because of the lack of standards. Taking the analogy of scientific data, the application of FAIR principles would reduce the time spent integrating data by 30%.

Similarly, trust would continue to be managed on a bilateral level, hindering large scale initiatives such as the data spaces.

As a result, data would be insufficiently shared. There would be only a moderate increase in data access and reuse. This would limit the capacity for productivity gains in traditional sectors that are described in the problem analysis. For instance, the gains seen in the iShare project would only slowly be apparent in other sectors and countries. Existing standardization initiatives will scale very slowly.

Intermediaries will then grow slowly, based on individual negotiations.

In the short term, however, it would protect business from increased competition by new entrants as well as competitors.

On a similar note, data companies would have less opportunities to develop new products and services because of limited data sharing.

The specific lack of standard on metadata for machine learning would not improve the risk for bias and incorrect decisions, as well as exposing to human rights violation, as shown by recent examples. It would potentially lead to a limited adoption of machine learning.

4.4.2.2.3 Findings of the Cost-Benefit Analysis

Option zero entails, by definition, no additional costs for any stakeholder. The option would maintain the current level of adoption for data schemes for data sharing between companies, which is marginal, thereby leading to no efficiency increases.

4.4.2.3 Coherence of the option

As there would be no intervention, coherence with existing legislation can largely be confirmed. However, there is a case for arguing that the issues identified as part of the problem assessment would hinder the achievement of a Single Market for Data. In particular, there seems to be a missing link with the creation of data spaces included in the data strategy.

4.4.3   Policy option 1 – Informal expert group

4.4.3.1 Effectiveness in achieving the policy objectives

4.4.3.1.1 Achievement of specific objectives
The policy option is suitable to the achievement of the specific objectives, notably:

- Ensuring trust, by facilitating the definition and adoption of standard data sharing schemes and reference architectures that include easy legal and trust arrangements for data sharing
- Facilitating interoperability, by promoting the activities for the definition of data and metadata standards, and principles for interoperability between sectors

It is clear that the creation of an informal expert group will only provide an indirect impact on the achievement of the objectives, as it will merely facilitate the activities of the existing standardization initiatives. The informal nature of the group raises questions on its capacity to achieve an impact, as it represents the weakest of the possible options. It will however allow for a more varied set of expertise, which is appropriate in view of the fast-evolving nature of data sharing.

The agile nature of the expert group is also more complementary with the large ecosystem of existing instruments and institution for fostering standard creation and adoption.

4.4.3.1.2 Achievement of general objectives
On a similar tone, the present option is certainly in line with the objective to increase data sharing and create a European Single Market for data, although the relation is indirect and heavily mediated.

Moreover, the expert group nature might allow for closer collaboration with industry than a legal body, thereby helping with the variable that will determine the ultimate success: guaranteeing the participation of companies in the definition and in the adoption of the standard.

4.4.3.2 Efficiency: Costs and benefits of the option

4.4.3.2.1 Costs of the option
Cost for business:

- Standardization is costly, in particular to ensure wide adoption. Full interoperability based on the FAIR principles requires 5 to 10% of a research project budget. So there will be additional costs as in the case of all standards. However, these additional costs will be ultimately part of the decision process that is controlled by industry as the policy option does not entail any mandatory standards.
- Another important cost category refers to participating in the expert groups activities, which is expected to be additional to other effort in the field of standardization. One interviewee summarizes the effort as the involvement of 1 to 7 people in a company team, engagement in 1 phone call per week (average 3 h), 1 meeting each 3 months that spans from 3 to one full week. This entails high travel expenses via continental flight (overseas because combines US - EU - Asia travels) and 1-week hotel every 3 months in average. The IDSA experience suggest the need for 20% of a full time equivalent per company. These costs will have to be covered by the companies.

Cost for intermediaries:

- Cost of paying for the expert group. Expert groups are typically only reimbursed for expenses, but only in exceptional cases for work. The costs are therefore limited to around 200 Euros per day per person – so a group of 10 experts would cost about 24,000 Euros per year based on the engagement above (four three-days meetings a year)..
- Cost of standard documentation and education. Costs of guidelines, toolkits, tutorials, webinars etc. It is discretionary to the team that lead the standardization effort and how much they want to invest on the engagement – but there is a direct correlation between the width of the support measures and the level of uptake. These costs are highly variable and are difficult to estimate.

Table 35 – Overview of costs for subtasks 1.3 | PO 1

Overview of costs – PO 1

		Data holders		Intermediaries		Data re(users)	
		One-off	*Recurrent*	*One-off*	*Recurrent*	*One-off*	*Recurrent*
European structure for governance aspects of data sharing	*Direct costs*	-	200,000/year for running the group	-	24,000/year for running the group	-	As data holders
	Indirect costs	-	-	-	-	-	-

4.4.3.2.2 Benefits of the option, including reductions in some of the costs as well as other positive effects on (some of) the stakeholders

For traditional business:

- Increased clarity on the existing standard and standardization initiatives, leading to a more informed choice by business over the standards to adopt.
- Increased adoption of standards, leading to reduction in costs for acquiring, integrating and processing data. Lack of technical standards for data and metadata is considered a major barrier to data sharing in traditional sectors.
- Adoption of standards provides easier access to data. Estimates from individual case studies show that adoption of standards for data sharing results in increased adoption. For instance, in the case of sport activities, the openactive.io standard led to 200,000 new activities being posed by 29 organisations, resulting in 150 to 500.000 new activities carried out every month.
- This will lead to accelerating the progress towards achieving the benefits from data sharing – enable the achievement of the high growth scenario put forward by the EU data market study, or to grasp the opportunities of the Internet of Things outlined in the Deloitte study. IDSA estimate an efficiency gain of 15% by the adoption of the reference architecture. Considering a potential addressable market of 700,000 companies (the data users identified by the EU data market study), and considering that the current number of companies involved in implementing data schemes is below 100, we can estimate conservatively that under this option 700 companies would be involved by 2025.
- For the scientific domain alone, FAIR data introduces efficiency measures by at least 10 billion euros, by reducing the effort for data reuse and avoiding duplication of data collection. When it comes to innovation, the expected gains from scientific data are quantified in 16 billion euros annually.

For intermediaries:

- Increased capacity to scale up standardization initiatives.

- Increased market and demand for standardization from traditional business.
- Reduced costs for the initiatives. The typical annual budget in the initial phase for such initiatives is around 5 million euros (for AboutML and iShare). Replication would allow for economies of scale.

For society:

- Lower costs for consumer and greater competition
- New products and services, particularly important for health.
- Reduction of bias in algorithms

For environment

- Reduction in carbon footprint due to increased efficiency

Table 36 – Overview of benefits – Subtask 1.3 | PO 1

Type of action	Description	Amount	Stakeholders
	Direct benefits		
European structure for governance aspects of data sharing	Costs Savings and efficiency gains - Easier reuse of data, lower cost of data processing and management	*30% reduction in data processing costs*	**Data holders, Data re-users in same sector, Data re-users in other sectors, Data Intermediaries, the Society**
	Greater access to data for faster development of new products and services Anticipate disruption by players from different sectors	*10bn euros/year cost of not having FAIR research data*	**Data holders in traditional industries**
	Possibility to access new markets	*Not possible to predict*	**Data re-users in same sector**
	Greater access to data sources for analytics and machine learning as well as development of new products and services	*16bn euros/year from innovative reuse of FAIR data*	**Researchers, Data holders in traditional industries, Tech companies**
	Indirect benefits		
	New products and services Anticipate disruption by players from different sectors	*1.4 trillion euros by 2027 for manufacturing*	**Data holders, Data re-users in same sector, the Society**
	Greater demand for data intermediaries	*Not possible to predict*	**Data intermediaries**
	Digital single market – greater access to data from other countries Reduced friction in reusing data	*Not possible to predict*	**Data holders, Data re-users in same sector, Data re-users in other**

		sectors, Data Intermediaries
Other non-monetisable benefits: new discoveries for health, environmental efficiency and new products	*Not possible to predict*	**Society**

4.4.3.2.3 Findings of the Cost-Benefit Analysis

Option 1 would lead according to our estimations validated by experts, to a small increase in adoption of such schemas, estimated in 700, or 0,1% of the current number of "data users" in the EU data market study (700.000).

Because of the nature of this measure, which entails the peer to peer sharing of data between companies, there is no distinction between data holders and reusers.

These 700 companies would benefit from an operational efficiency of 15% of an average OPEX of 50M EUR in 5 years, as estimated by interviewed experts. This leads to an average 1050 M EUR benefit per year, or 5250 M EUR in five years (NPV 4.668 M EUR).

As regard the costs, we estimate no additional costs for the European Commission since the participation to the experts group is voluntary, only a reimbursement of expenses for about 24.000 EUR per year. Costs by participating companies would be for a total of 200.000 EUR per year. This is based on an estimate of four three-day meetings a year by ten company representatives.

The final results is a NPV of 4668 Million EUR and a BCR of 200.362 M EUR. However, the BCR is not the most reliable indicator because of the very small scale of expenses.

4.4.3.3 Coherence of the option

The option is fully coherent with the Digital Single Market strategy and will be highly complementary with the creation of data spaces.

4.4.4 Policy option 2 – Formal expert group

4.4.4.1 Effectiveness in achieving the policy objectives

4.4.4.1.1 Achievement of specific objectives

The policy option is suitable to the achievement of the specific objectives, notably:

- Ensuring trust, by facilitating the definition and adoption of standard data sharing schemes and reference architectures that include easy legal and trust arrangements for data sharing
- Facilitating interoperability, by promoting the activities for the definition of data and metadata standards, and principles for interoperability between sectors

It is clear that the creation of a formal expert group will only provide an indirect impact on the achievement of the objectives, as it will merely facilitate the activities of the existing standardization initiatives. The formal nature of the group is likely to provide some kind of authority, while not requiring the investment that is devoted to a legal body, and allowing for a more varied set of expertise, which is appropriate in view of the fast-evolving nature of data sharing.

The agile nature of the expert group is also more complementary with the large ecosystem of existing instruments and institution for fostering standard creation and adoption.

4.4.4.1.2   Achievement of general objectives

On a similar tone, the present option is certainly in line with the objective to increase data sharing and create a European Single Market for data, although the relation is indirect and heavily mediated.

Moreover, the expert group nature might allow for closer collaboration with industry than a legal body, thereby helping with the variable that will determine the ultimate success: guaranteeing the participation of companies in the definition and in the adoption of the standard.

4.4.4.2   Efficiency: Costs and benefits of the option

4.4.4.2.1   Costs of the option

Cost for business

- Standardization is costly, in particular to ensure wide adoption. Full interoperability based on the FAIR principles requires 5 to 10% of a research project budget. So there will be additional costs as in the case of all standards. However, these additional costs will be ultimately part of the decision process that is controlled by industry as the policy option does not entail any mandatory standards.
- Another important cost category refers to participating in the expert groups activities, which is expected to be additional to other effort in the field of standardization. One interviewee summarizes the effort as the involvement of 1 to 7 people in a company team, engagement in 1 phone call per week (average 3 h), 1 meeting each 3 months that spans from 3 to one full week. This entails high travel expenses via continental flight (overseas because combines US- EU- Asia travels) and 1-week hotel every 3 months in average. The IDSA experience suggest the need for 20% of a full time equivalent per company. These costs will have to be covered by the funder of the expert group, usually the European Commission, under the budget below.

Cost for intermediaries:

- Cost of paying for the expert group or scientific committee. Formal expert groups and scientific committees include a daily fee for work in addition to expenses. The costs are higher than for informal expert groups. For instance, the current Scientific Committees cost on average 280.000 euros per year (four three-day meetings a year).[164] This average includes travel costs and miscellaneous costs.
- Cost of standard documentation and education. Costs of guidelines, toolkits, tutorials, webinars etc. It is discretionary to the team that lead the standardization effort and how much they want to invest on the engagement – but there is a direct correlation between the width of the support measures and the level of uptake. These costs are highly variable and are difficult to estimate.

Table 37 – Overview of costs for subtasks 1.3 | PO 2

Overview of costs – PO 2

		Data holders		Intermediaries		Data re(users)	
		One-off	*Recurrent*	*One-off*	*Recurrent*	*One-off*	*Recurrent*
European structure for governance aspects of data sharing	*Direct costs*	-		-	280,000/year for running the group	-	As data holders
	Indirect costs	-	-	-	-	-	-

---

[164] European Commission, 2016. Report on the activity of the scientific committee's term 2013-2016.

4.4.4.2.2 Benefits of the option, including reductions in some of the costs as well as other positive effects on (some of) the stakeholders

For traditional business:

- Increased clarity on the existing standard and standardization initiatives, leading to a more informed choice by business over the standards to adopt.
- Increased adoption of standards, leading to reduction in costs for acquiring, integrating and processing data. Lack of technical standards for data and metadata is considered a major barrier to data sharing in traditional sectors.
- Adoption of standards provides easier access to data. Estimates from individual case studies show that adoption of standards for data sharing results in increased adoption. For instance, in the case of sport activities, the openactive.io standard led to 200,000 new activities being posed by 29 organisations, resulting in 150 to 500,000 new activities carried out every month.
- This will lead to accelerating the progress towards achieving the benefits from data sharing – enable the achievement of the high growth scenario put forward by the EU data market study, or to grasp the opportunities of the Internet of Things outlined in the Deloitte study. IDSA estimate an efficiency gain of 15% by the adoption of the reference architecture. Considering a potential addressable market of 700.000 companies (the data users identified by the EU data market study), and considering that the current number of companies involved in implementing data schemes is below 100, we can estimate conservatively that under this option 800 companies would be involved by 2025.
- For the scientific domain alone, FAIR data introduces efficiency measures by at least 10 billion euros, by reducing the effort for data reuse and avoiding duplication of data collection. When it comes to innovation, the expected gains from reuse of scientific data are quantified in 16 billion euros annually.

For intermediaries:

- Increased capacity to scale up standardization initiatives.
- Increased market and demand for standardization from traditional business.
- Reduced costs for the initiatives. The typical annual budget in the initial phase for such initiatives is around 5 million euros (for AboutML and iShare). Replication would allow for economies of scale.

For society:

- Lower costs for consumer and greater competition
- New products and services, particularly important for health.
- Reduction of bias in algorithms

For environment

- Reduction in carbon footprint due to increased efficiency

Table 38 – Overview of benefits – Subtask 1.3 | PO 2

Type of action	Description	Amount	Stakeholders
	Direct benefits		

European structure for governance aspects of data sharing	Costs Savings and efficiency gains - Easier reuse of data, lower cost of data processing and management	*30% reduction in data processing costs 15% overall efficiency gain*	**Data holders, Data re-users in same sector, Data re-users in other sectors, Data Intermediaries, the Society**
	Greater access to data for faster development of new products and services Anticipate disruption by players from different sectors	*10 bn euros/year cost of not having FAIR research data*	**Data holders in traditional industries**
	Possibility to access new markets	*Not possible to predict*	**Data re-users in same sector**
	Greater access to data sources for analytics and machine learning as well as development of new products and services	*16bn euros/year from innovative reuse of FAIR data*	**Researchers, Data holders in traditional industries, Tech companies**
	Indirect benefits		
	New products and services Anticipate disruption by players from different sectors	*1.4 trillion euros by 2027 for manufacturing*	**Data holders, Data re-users in same sector, the Society**
	Greater demand for data intermediaries	*Not possible to predict*	**Data intermediaries**
	Digital single market – greater access to data from other countries Reduced friction in reusing data	*Not possible to predict*	**Data holders, Data re-users in same sector, Data re-users in other sectors, Data Intermediaries**
	Other non-monetisable benefits: new discoveries for health, environmental efficiency and new products	*Not possible to predict*	**Society**

4.4.4.2.3   Findings of the Cost-Benefit Analysis

Option 2 would lead according to our estimations, validated by experts, to a slightly greater increase in adoption of such schemas, estimated in 800, against the 700 of option 1. Interviewed experts were adamant that we cannot expect a major difference in terms of adoption between an informal and formal expert group, which cannot have any way to directly act (not to mention enforce) companies' choice for adoption of a specific standard. In other words, the general effect of such a soft policy measure as an expert group on adoption by companies is limited, as it is heavily mediated by other factors, and the difference between a formal and informal expert group is expected to be minor.

Because of the nature of this measure, which entails the peer to peer sharing of data between companies, there is no distinction between data holders and reusers.

These 800 companies would benefit from an operational efficiency of 15% of an average OPEX of 45M EUR in 5 years, assuming that the companies joining are typically a bit smaller than in option 1. This leads to an average 1.200,0 M EUR benefit per year, or 6000 M EUR in five years (NPV 4. 5335,6 M EUR).

As regard the costs, we estimate some additional costs for the European Commission with respect to option 2, for the reimbursement of travel expenses and a fee for participation, for a total of 280.000 (based on experience with other formal expert groups). This is based on an estimate of four three-day meetings a year by ten company representatives.

The final results is a NPV of 5.335,3 Million EUR and a BCR of 19.627,3 M EUR. However, the BCR is not the most reliable indicator because of the very small scale of expenses.

4.4.4.3 Coherence of the option
The option is fully coherent with the Digital Single Market strategy and will be highly complementary with the creation of data spaces.

4.4.5 Policy option 3 – Legal body

4.4.5.1 Effectiveness in achieving the policy objectives
4.4.5.1.1 Achievement of specific objectives
The policy option is suitable to the achievement of the specific objectives, notably:

- Ensuring trust, by facilitating the definition and adoption of standard data sharing schemes and reference architectures that include easy legal and trust arrangements for data sharing
- Facilitating interoperability, by promoting the activities for the definition of data and metadata standards, and principles for interoperability between sectors

However, it is clear that the creation of a legal body will only provide an indirect impact on the achievement of the objectives, as it will merely facilitate the activities of the existing standardization initiatives.

There are already well-developed instruments and institution for fostering standard creation and adoption.

4.4.5.1.2 Achievement of general objectives
On a similar tone, the present option is certainly in line with the objective to increase data sharing and create a European Single Market for data, although the relation is indirect and heavily mediated. Moreover, it does not offer particular advantages with regard to the variable that will determine the ultimate success: guaranteeing the participation of companies in the definition and in the adoption of the standard. Indeed, a legal body could even prove counterproductive in a context with multiple actors and initiatives in place, and where leadership sit squarely with business.

4.4.5.2 Efficiency: Costs and benefits of the option
4.4.5.2.1 Costs of the option
Cost for business

- Standardization is costly, in particular to ensure wide adoption. Full interoperability based on the FAIR principles requires 5 to 10% of a research project budget. So there will be additional costs as in the case of all standards. However, these additional costs will be ultimately part of the decision process that is controlled by industry as the policy option does not entail any mandatory standards.
- Another important cost category refers to participating in the expert groups activities, which is expected to be additional to other effort in the field of standardization. One interviewee summarizes the effort as the involvement of 1 to 7 people in a company team, engagement in 1 phone call per week (average 3

h), 1 meeting each 3 months that spans from 3 to one full week. This entails high travel expenses via continental flight (overseas because combines US- EU- Asia travels) and 1-week hotel every 3 months in average. The IDSA experience suggest the need for 20% of a full time equivalent per company. These costs will have to be covered by the legal body as part of the budget below.

Cost for intermediaries

- Cost of setting up and running the legal body which are higher than the expert groups. Based on comparable initiatives, such as the European Data Protection Board, we estimate a yearly budget of 3.5 million euros.
- Cost of standard documentation and education. Costs of guidelines, toolkits, tutorials, webinars etc. It is discretionary to the team that lead the standardization effort and how much they want to invest on the engagement – but there is a direct correlation between the width of the support measures and the level of uptake. These costs are highly variable and are difficult to estimate.

Table 39 – Overview of costs for subtasks 1.3 | PO 3

Overview of costs – PO 3		Data holders		Intermediaries		Data re(users)	
		One-off	*Recurrent*	*One-off*	*Recurrent*	*One-off*	*Recurrent*
European structure for governance aspects of data sharing	*Direct costs*	-		-	3.5 million euros/year	-	As data holders
	Indirect costs	-	-	-	-	-	-

4.4.5.2.2 Benefits of the option, including reductions in some of the costs as well as other positive effects on (some of) the stakeholders

For traditional business:

- Increased clarity on the existing standard and standardization initiatives, leading to a more informed choice by business over the standards to adopt.
- Increased adoption of standards, leading to reduction in costs for acquiring, integrating and processing data. Lack of technical standards for data and metadata is considered a major barrier to data sharing in traditional sectors.
- Adoption of standards provides easier access to data. Estimates from individual case studies show that adoption of standards for data sharing results in increased adoption. For instance, in the case of sport activities, the openactive.io standard led to 200,000 new activities being posed by 29 organisations, resulting in 150 to 500,000 new activities carried out every month.
- This will lead to accelerating the progress towards achieving the benefits from data sharing – enable the achievement of the high growth scenario put forward by the EU data market study, or to grasp the opportunities of the Internet of Things outlined in the Deloitte study. IDSA estimate an efficiency gain of 15% by the adoption of the reference architecture. Considering a potential addressable market of 700.000 companies (the data users identified by the EU data market study), and considering that the current number of companies involved in implementing data schemes is below 100, we can estimate conservatively that under this option 900 companies would be involved by 2025.
- For the scientific domain alone, FAIR data introduces efficiency measures by at least 10 billion euros, by reducing the effort for data reuse and avoiding duplication of data collection. When it comes to innovation, the expected gains from scientific data are quantified in 16 billion euros annually.

For intermediaries:

- Increased capacity to scale up standardization initiatives.
- Increased market and demand for standardization from traditional business.
- Reduced costs for the initiatives. The typical annual budget in the initial phase for such initiatives is around 5 million euros (for AboutML and iShare). Replication would allow for economies of scale.

For society:

- Lower costs for consumer and greater competition
- New products and services, particularly important for health.
- Reduction of bias in algorithms

For environment

- Reduction in carbon footprint due to increased efficiency

Table 40 – Overview of benefits – Subtask 1.3 | PO 3

Type of action	Description	Amount	Stakeholders
European structure for governance aspects of data sharing	**Direct benefits**		
	Costs Savings and efficiency gains - Easier reuse of data, lower cost of data processing and management	*30% reduction in data processing costs*	**Data holders, Data re-users in same sector, Data re-users in other sectors, Data Intermediaries, the Society**
	Greater access to data for faster development of new products and services Anticipate disruption by players from different sectors	*10 bn euros/year cost of not having FAIR research data*	**Data holders in traditional industries**
	Possibility to access new markets	*Not possible to predict*	**Data re-users in same sector**
	Greater access to data sources for analytics and machine learning as well as development of new products and services	*16bn euros/year from innovative reuse of FAIR data*	**Researchers, Data holders in traditional industries, Tech companies**
	Indirect benefits		
	New products and services Anticipate disruption by players from different sectors	*1,4 trillion euros by 2027 for manufacturing*	**Data holders, Data re-users in same sector, the Society**
	Greater demand for data intermediaries	*Not possible to predict*	**Data intermediaries**
	Digital single market – greater access to data from other countries Reduced friction in reusing data	*Not possible to predict*	**Data holders, Data re-users in same sector, Data re-users in other**

			sectors, Data Intermediaries
	Other non-monetisable benefits: new discoveries for health, environmental efficiency and new products	*Not possible to predict*	**Society**

Type of action	Description	Amount	Stakeholders
European structure for governance aspects of data sharing	**Direct benefits**		
	Costs Savings and efficiency gains - Easier reuse of data, lower cost of data processing and management	*30% reduction in data processing costs* *15% overall efficiency gain*	**Data holders, Data re-users in same sector, Data re-users in other sectors, Data Intermediaries, the Society**
	Greater access to data for faster development of new products and services Anticipate disruption by players from different sectors	*10 bn euros/year cost of not having FAIR research data*	**Data holders in traditional industries**
	Possibility to access new markets	*Not possible to predict*	**Data re-users in same sector**
	Greater access to data sources for analytics and machine learning as well as development of new products and services	*16bn euros/year from innovative reuse of FAIR data*	**Researchers, Data holders in traditional industries, Tech companies**
	Indirect benefits		
	New products and services Anticipate disruption by players from different sectors	*1.4 trillion euros by 2027 for manufacturing*	**Data holders, Data re-users in same sector, the Society**
	Greater demand for data intermediaries	*Not possible to predict*	**Data intermediaries**
	Digital single market – greater access to data from other countries Reduced friction in reusing data	*Not possible to predict*	**Data holders, Data re-users in same sector, Data re-users in other sectors, Data Intermediaries**
	Other non-monetisable benefits: new discoveries for health, environmental efficiency and new products	*Not possible to predict*	**Society**

4.4.5.2.3   Findings of the Cost-Benefit Analysis

Option 3 would lead according to our estimations validated by experts, to a slightly greater increase in adoption of such schemas, estimated in 900, against the 700 of option 1 and 800 for option 2. Interviewed experts were adamant that we cannot expect a major difference in terms of adoption between an expert group and a legal body, as the initiative should remain mostly with industry and government should not be in a position of picking winners among standards. The legal body cannot have any way to directly act (not to mention enforce) companies' choice for adoption of a specific standard. In other words, the general effect

of such a soft policy measure as a legal body on adoption by companies is limited, as it is heavily mediated by other factors, and the difference between an expert group and a legal body is expected to be minor.

Because of the nature of this measure, which entails the peer to peer sharing of data between companies, there is no distinction between data holders and reusers.

These 900 companies would benefit from an operational efficiency of 15% of an average OPEX of 40M EUR in 5 years, assuming that the companies joining are typically a bit smaller than in option 1 and 2. This leads to an average 1350 M EUR benefit per year, or 6750 M EUR in five years (NPV 6.002,5 M EUR).

As regard the costs, we estimate additional costs for the European Commission with respect to option 2, related to fixed costs for the organization for a total of 3,5 M EUR (based on analogy with other legal bodies such as the European Data Protection Board. This includes also the budget for travelling and reimbursement of speakers' fees.

The final results is a NPV of 5.999,1 Million EUR and a BCR of 1.766,5 M EUR.

4.4.5.3 Coherence of the option
The option is fully coherent with the Digital Single Market strategy and will be highly complementary with the creation of data spaces.

4.4.6 Summary of the impacts

The following table summarises the possible impacts of the policy options:

Table 41 - Summary of impacts for subtask 1.3

Economic impacts	- costs for business - staff for participating to standardisation meeting - benefits for business - efficiency from data sharing - costs for intermediaries : - staff for participating - benefits for intermediaries : - increased demand and awareness of their solutions - costs for European Commission - Cost of running expert group
Social impacts	Increased social innovation such as new drug discoveries
Environmental impacts	Increased energy efficiency through sharing of consumption data
Fundamental rights impacts	Reduced risk for algorithmic bias and discrimination through dataset metadata

4.5 Subtask 1.4: Establishing a certification framework for data intermediaries

4.5.1 Stakeholders affected

The following table provides an overview of the key stakeholders affected by the possible policy options and how:

Table 42 – Overview of stakeholders affected by subtask 1.4 policy options

Who?	How?
Data holders	Data holders will not be directly affected in terms of costs, but will benefit mainly from data monetisation via the data sharing through certified intermediaries, cost and time savings through digitisation of the transactions, increased control over their data and increased revenue generated from the network growth the increased volume of data sharing.

Data intermediaries	Certification will help data intermediaries to scale up and grow in terms of revenue, resources, client base and volume of data transactions. However, they will be the main stakeholders affected directly from the certification cost.
Data (re)users	Data re-users will benefit mainly from cost and time savings through digitisation of the transactions, easier access to data and the creation of mechanisms to assess the quality of data intermediaries' services. They might also face some indirect certification costs due to potentially increased charges of the certified data intermediaries' services.
Society	Societal benefit will be twofold. On the one hand society will benefit as the potential of the European data market will be unlocked through certification. On the other hand, data flows of intermediaries serving societal purposes (i.e. health, research) will be increased.

4.5.2 Policy option 0: Baseline scenario- No horizontal action at EU level

4.5.2.1 Effectiveness in achieving the policy objectives

This subsection examines the effectiveness of the baseline scenario in achieving the specific and general policy objectives.

4.5.2.1.1 Achievement of specific objectives

The absence of horizontal action at EU level would not allow to create **trust** in common European data spaces, neither to build **common data spaces** that is making more data usable where data holders could agree to it through technical, legal and organisational support. Finally, current data **interoperability** issues across sectors would persist.

Since the data intermediaries would remain uncertified, there would remain a lack of trust and mechanisms for data holders and users to assess their "neutrality", while at the same time there would be no differentiation between neutral and non-neutral data intermediaries. This will lead to further interoperability issues and regulatory fragmentation in the internal market. Ultimately, there will be less data available for reuse across sectors. This will prevent the EU from reaping the full benefits of horizontal data sharing which account for 20% of all the benefits of data sharing in general[165].

In particular, according to the summary report of the open public consultation on the European strategy for data, almost 80% of the 512 respondents have encountered difficulties in using data from other companies, related to technical aspects (data interoperability and transfer mechanisms), denied data access, and prohibitive prices or other conditions considered unfair or prohibitive. A very large share of respondents (87.7%) supported the idea that the EU should make major investments in technologies and infrastructures that enhance data access and use, while giving individuals as well as public and private organisations full control over the data they generate. Around the same proportion of respondents considered that the development of common European data spaces should be supported by the EU in strategic industrial sectors and domains of public interest. [166] This idea cannot be supported by the policy option 0 and the absence of action at EU level.

4.5.2.1.2 Achievement of general objectives

The absence of horizontal action at EU level would not contribute to setting **the foundations of a Single Market for Data, neither to strengthen the EU data economy.** Since the data intermediaries would remain unregulated, such businesses would not be empowered with respect to the data use they generate to create value for the society.

---

[165] Realising the economic potential of machine-generated, non-personal data in the EU, Deloitte Report for Vodafone Group, July 2018
[166] https://ec.europa.eu/digital-single-market/en/news/summary-report-public-consultation-european-strategy-data

4.5.2.2 Efficiency: Costs and benefits of the option

This subsection examines the efficiency of the baseline scenario in terms of costs and benefits.

4.5.2.2.1 Costs of the option

The majority of stakeholders interviewed expressed opinions against the baseline scenario and the lack of action at EU level, due to the costs arising from the lack of trust within the market, which does not allow the potential of data to be unlocked and the data intermediaries to grow. Furthermore, data holder and data users would also have to bear the cost of more expensive transactions, while losing time savings and efficiency gains from transactions that would be otherwise facilitated by the certified intermediaries.

4.5.2.2.2 Benefits of the option, including reductions in some of the costs as well as other positive effects on (some of) the stakeholders

The absence of action at EU level could create cost and time savings for data intermediaries who were against certification, as they would not need to bear the cost of obtaining and maintaining the certification, neither to have a competitive disadvantage compared to certified data intermediaries.

4.5.2.2.3 Findings of the Cost-Benefit Analysis for Policy Option 0

Policy Option O entails no significant costs and benefits for the stakeholders, since no action is taken. The assessment of both costs and benefits under the baseline scenario is linked to the current absence of a certification framework for data intermediaries in the European market.

4.5.2.3 Coherence of the option

The absence of action at EU level would not change the status quo, therefore coherence with the existing EU policy and legal framework is ensured. However, incoherence issues might arise from the fact that the provisions of the European Data Strategy for (i) data flow within the EU and across sectors, for the benefit of all (ii) and the rules for access and the use of data are fair, practical and clear would not be strengthened or promoted by the maintenance of the baseline scenario.

4.5.3 Policy option 1: Coordination at EU level (industry driven self-regulatory certification framework)

4.5.3.1 Effectiveness in achieving the policy objectives

This subsection examines the effectiveness of policy option 1 in achieving the specific and general policy objectives.

4.5.3.1.1 Achievement of specific objectives

The establishment of a self-regulatory certification framework could contribute into the three specific policy objectives however not in a significant way. Stakeholders from both B2B and C2B interviewed expressed concerns that policy option no. 1 might have little added value compared to Policy option 0. There is already a self-regulatory effort being currently conducted in the frame of MyData Community for personal data operators.

Concerns were also raised that big industry players would have a stronger role and could potentially influence the outcome of the discussions taking place in the frame of the stakeholder forum.

4.5.3.1.2 Achievement of general objectives

In line with arguments presented in the above mentioned section, within the limits of its effectiveness this policy option could promote and contribute into setting the foundations of a single market for data as well as strengthen the EU data economy as a broader impact, although in a limited fashion.

4.5.3.2 Efficiency: Costs and benefits of the option

Policy option 1 was not examined for a detailed cost-benefit analysis. During stakeholder interviews, several **data intermediaries in both B2B and C2B markets expressed concerns that big industry players would have a stronger role compared to SMEs** and **could potentially influence in their favour the outcome of the discussions** taking place within the stakeholder forum. The absence of a regulatory measure adopted at EU level would not help ensuring a **fair and well-representative selection of certification criteria/requirements** for all the various types of data intermediaries active in the European market. Furthermore, **there is already a completely industry-driven, self-regulatory certification (self-description) process in place**, initiated in 2020 within the MyData Community for personal data spaces. For these reasons, it was estimated that this policy option would have **limited added-value and impacts** compared to the baseline scenario, and therefore the option was not further considered for detailed cost and benefits analysis. Indicative categories of costs and benefits are provided below.

4.5.3.2.1   Costs of the option

Data intermediaries would have in this case to bear the cost of **obtaining and maintaining the certification**. Depending on the requirements and the criteria decided during the stakeholder forum discussions, some intermediaries would also have to face implementation costs to ensure compliance with the requirements. Given that this is a self-regulatory approach, the requirements are not expected to be hard. This affects the effectiveness and efficiency of the option, but also limits the related costs. In particular the expected costs by the stakeholders interviewed, linked to this policy option, are 15K to 20K EUR one-off costs for obtaining the certification for the first time (including 10K EUR internal preparatory costs and 5K-10K EUR external certification costs) and 10K to 20K EUR/year recurrent costs for renewing it (including 5K-10K EUR internal preparatory costs and up to 5K EUR external certification costs).

In this policy option data intermediaries might have to face also costs of **setting up the scheme** and potentially funding a private **certification agency**, depending on how this framework is being designed. Time and resource costs are also involved for the participation to the stakeholder forum discussions.

In terms of indirect impacts, the **competition on the market** is expected to be between 25%-1% decreased in both C2B and B2B data sharing markets, within the first year after the first operators will obtain the certification.

Data re-users might be affected as well from the certification from **indirect transaction costs**, as the certified intermediaries might increase the user charges to cover the certification cost.

4.5.3.2.2   Benefits of the option, including reductions in some of the costs as well as other positive effects on (some of) the stakeholders

Depending on its effectiveness described in the previous section, policy option 1 could provide benefits in all the stakeholders of the value chain, due to the **trust between the actors** that would be brought in the ecosystem.

This trust could lead into further **efficiency gains and time savings, increase in the client base and data transactions and therefore increase of revenues,** allowing data intermediaries to scale up but also other stakeholders in the value chain. However, if the effectiveness rate of this option remains low, then the benefits would be lower as well, compared to the full certification potential under other policy options. Given that this is a self regulatory approach, the requirements are not expected to be hard. This affects the effectiveness and efficiency of the option, but also limits its associated benefits. In particular, the benefits expected by the interviewed stakeholders under this policy option include:

- 20%-25% business development time acceleration

- Growth in terms of revenues and employee number:
  - between 25-35% growth, within the 1st year after obtaining the certification;
  - between 25-35% growth, from 2 to 5 years' timeframe;
  - up to 25% growth, beyond 5 years' timeframe
- Growth in terms of client base (including both number of clients and number use cases)
  - up to 25% increase within the 1st year after obtaining the certification;
  - 25% increase from 2 to 5 years' timeframe;
  - 25%-35% increase beyond 5 years' timeframe.

Data holders would have the opportunity to monetise more from data sharing while more individuals will be attracted to share their personal data through the certified platforms. Data holders as well as data re-users would also have cost and time savings through digitization of interaction with data re-users by certified intermediaries, while at the same time they would have a kind of mechanisms for assessing the quality of services provided by data intermediaries.

In terms of indirect benefits, while certification is expected to limit market competition in a short-term plan, the opposite impact is expected in a longer-term plan. In particular, the interviewed stakeholders expect a competition increase in both C2B and B2B data sharing markets of 1%-25% from 2-5 years' timeframe after obtaining the certification, and a further 1%-25% increase in a beyond 5 year's timeframe.

4.5.3.2.3 Findings of the Cost-Benefit Analysis for Policy Option 1

Policy Option 1 entails relatively low levels of benefits including business development time acceleration, client base and revenues increases, market competition, as a result of the increased trust between the stakeholders in the market after certification. These benefits are also linked to relatively low levels of recurrent and one-off costs for data intermediaries in order to obtain and maintain the certification. However, the expected benefits significantly outweigh the costs. Details on the costs and benefits for policy option 1 are presented in Annex C.

4.5.3.3 Coherence of the option

This policy option could promote the objectives of the single market for data under the European Data Strategy and therefore remains coherent with the current EU legal and policy framework.

4.5.4 Policy option 2: Regulatory intervention with low intensity (voluntary certification framework)

4.5.4.1 Effectiveness in achieving the policy objectives

This subsection examines the effectiveness of policy option 2 in achieving the specific and general policy objectives.

4.5.4.1.1 Achievement of specific objectives

This policy option could significantly contribute to the three specific objectives, particularly, in creating trust in common European data spaces, building common data spaces as well as ensuring data interoperability across sectors, through the certification framework. The majority of stakeholders interviewed agreed to this policy option as the most effective one, for the current status of the market, given also that the certification criteria would be defined by a legal instrument. Therefore, if a big number of industry players decides to proceed to the certification process, the trust between the stakeholders involved in the market would be increased significantly, allowing the data intermediary market to flourish and bringing various economic and societal benefits.

4.5.4.1.2   Achievement of general objectives

Following the effectiveness in achieving the specific objectives described in the above section, this policy option would also further contribute to set the foundations of a Single Market for Data, and furthermore, strengthen the EU data economy, since the European data market overall will be significantly boosted through certification, increasing the volume of data flows.

4.5.4.2   Efficiency: Costs and benefits of the option

This subsection examines the efficiency of policy option 2 in terms of costs and benefits. The detailed results of the cost-benefit analysis for the sub-task 1.4 can be found in Annex C.

4.5.4.2.1   Costs of the option

Similarly, to the policy option 1, data intermediaries would have in this case as well to bear the cost of **obtaining and maintaining the certification**. Depending on the requirements and the criteria defined in the legal instrument, some intermediaries would also have to face implementation costs to ensure compliance with the requirements. Small industry players of early stage growth pointed out the importance of these costs to remain in an affordable level (not exceeding the 20K as a point of reference), and explained that the costs might also involve the need for additional resources (e.g such as lawyers or developers to ensure compliance) depending on how exactly the framework is being designed. If the certification cost cannot be minimised to a reasonable cost for SMEs then it should be subsidised. However, opinions on this vary depending on the growth stage of a company. In particular, a well-established industry player in the B2B market, raised the argument that the certification cost should not be kept low while the criteria should rather be strict, in order to ensure a proper differentiation between the various players in the market. The interviewed stakeholders estimate that certification costs under this policy option would be of 20K – 50K EUR one-off costs for obtaining the certification for a first time (including 10K-25K EUR internal preparatory costs and 10K-25K EUR external certification costs) and of 20K-35K EUR/year recurrent costs for renewing it (including 10-25K EUR/year internal preparatory costs and 10K EUR/year external certification costs).

In terms of indirect impacts, the **competition in the market** might be slightly affected in this policy option as well, once these first intermediaries will obtain the certification, with an expected decrease of 25% in the B2B market. However, in the long term new competitors will be attracted in the market, due to the acceleration of the market and the speed of adoption. Some stakeholders pointed out that **innovation** might also be decreased in the market depending on how strict the criteria will be. On the contrary, other stakeholders face certification as an opportunity for innovation increase, based on the fact that the certification will be significantly beneficial for the market, providing the opportunity for the development of new products or services. A stakeholder representing the personal data market, expressed the opinion that the way the certification requirements will be written will determine whether the certification will limit innovation or not. If the requirements are only about what needs to be proven but not about the way this will be implemented, then it should not limit innovation. Another stakeholder in growth stage in the B2B market highlighted the importance of the need for the framework to represent and benefit the market as a whole, and not only specific big-industry players, as this would lead to both reduced innovation and reduced competition.

Data re-users might be affected as well from the certification from **indirect transaction and implementation costs**, as the certified intermediaries might increase the user charges to cover the certification cost. However not all stakeholders agree to this argument.

Table 43 – Overview of costs for subtasks 1.4 | PO 2[167]

Overview of costs – PO 2			
	Data holders	Data intermediaries	Data re(users)

---

[167] The table presents the estimated amounts of costs by the interviewed stakeholders

		One-off	Recurrent	One-off	Recurrent	One-off	Recurrent
Certification framework for data intermediaries	*Direct costs*	-	-	**20K – 50K EUR** •10K-25K EUR internal preparatory costs •10K-25K EUR external certification costs	**20K-35K EUR/year** •10-25K EUR/year internal preparatory costs •10K EUR/year external certification costs	-	-
	Indirect costs	-	-	-	Around 25% decreased market competition	-	Non-quantifiable costs due to lack of data

4.5.4.2.2 Benefits of the option, including reductions in some of the costs as well as other positive effects on (some of) the stakeholders

The majority of the interviewed stakeholders expressed views in favour of this policy option as the preferred one, seeing a broad number of benefits. These mainly include the **trust** between the actors that would be brought in the ecosystem leading into **further efficiency gains, time savings, increase of the client base** and data transactions and therefore **increase of revenues**, allowing data intermediaries to scale up but also other stakeholders in the value chain. An aspect linked to the increased trust, where many stakeholders focused on, is the acceleration of the market, speed of adoption and time savings in business development cycle of data intermediaries. Furthermore, certification under this policy option is expected to guarantee uniform and transparent security levels, interoperability, safety and quality of intermediaries' services. In particular, the benefits expected by the interviewed stakeholders under this policy option include:

- 25%-50% business development time acceleration
- Growth in terms of revenues and employee number:
  - between 35%-40% increase, within the 1st year after obtaining the certification;
  - between 40%-50% increase, from 2 to 5 years' timeframe;
  - between 40%-50% increase, beyond 5 years' timeframe
- Growth in terms of client base (number of clients)
  - between 25%-50% increase within the 1st year after obtaining the certification;
  - between 40%-50% increase from 2 to 5 years' timeframe;
  - between 40%-50% increase beyond 5 years' timeframe.
- Growth in terms of client base (number use cases)
  - between 35%-40% increase within the 1st year after obtaining the certification;
  - between 40%-50% increase from 2 to 5 years' timeframe;
  - between 35%-40% increase beyond 5 years' timeframe.

According to a stakeholder of growth stage in the C2B value chain, an increase of approximately 20% in client base and revenues expected after certification, corresponds to additional 30M-50M EUR revenue increase per year.

With regards to the indirect benefits under this policy option, the interviewed data intermediaries expect also increased innovation-related benefits once the market is boosted through the certification, explaining market acceleration will bring new use cases, which wouldn't be viable before certification. Furthermore,

increased trust in the market could also lead to increase of funding, as investors will find it safer to invest in certified companies, A stakeholder representing a data union (C2B value chain) focused on the legitimacy-related benefits, to provided under this policy option which is expected to lead to additional efficiency and trust between the company and its members/clients as well as to strengthen the role of newly established C2B data intermediaries, such as data unions or data cooperatives by setting the scene behind their mission.. As a consequence, the market competition would be boosted for these newly established data intermediaries being in growth phase, empowering them towards the monopoly or oligopoly of big industrial players. Increased competition benefits are expected to be met however in both C2B and B2B value chains under this policy option, especially in a long-term timeframe. In particular, the interviewed stakeholders, expect:

- between 1%-25% increased competition in the B2B market from 2 to 5 years' timeframe;
- between 1%-25% increased competition in the B2B market beyond 5 years' timeframe;
- between 1%-25% increased competition in the C2B market, within the 1st year after obtaining the certification;
- between 1%-25% increased competition in the C2B market, from 2 to 5 years' timeframe;
- between 25%-50% increased competition in the C2B market, beyond 5 years' timeframe.

Data holders would have the opportunity to monetise more from data sharing, while they will have easier access to a bigger network of data re-users. At the same time more individuals will be attracted to share their personal data through the certified platforms. Data holders as well as data re-users would also have cost and time savings through digitization of interaction with data re-users by certified intermediaries, while mechanisms for assessing the quality of services provided by data intermediaries will be created for them through certification.

Table 44 – Overview of benefits – Subtask 1.4 | PO 2[168]

Type of action	Description	Amount	Stakeholders
Certification framework for data intermediaries	**Direct benefits**		
	Time savings and business development cycle acceleration; speed of adoption of the market	25%-50% business development time acceleration after certification	**Data Intermediaries**
	Company growth in terms of revenue and number of employees	• Between 35%-40% increase within the 1st year after obtaining the certification; • between 40%-50% increase from 2 to 5 years' timeframe; • between 40%-50% increase beyond 5 years' timeframe • 20% increase corresponds to 30M-50M EUR/year revenue increase for a growth-stage intermediary	**Data Intermediaries**
	Company growth in terms of client base (number of clients)	• between 25%-50% increase within the 1st year after obtaining the certification; • between 40%-50% increase from 2 to 5 years' timeframe; • between 40%-50% increase beyond 5 years' timeframe	**Data Intermediaries**
	Company growth in terms of client base (number of use cases)	• between 35%-40% increase within the 1st year after obtaining the certification; • between 40%-50% increase from 2 to 5 years' timeframe;	**Data Intermediaries**

---

[168] The table presents the estimated amounts of benefits by the interviewed stakeholders

		• between 35%-40% increase beyond 5 years' timeframe	
Data monetisation via data sharing through certified platforms	*Not quantifiable due to lack of data*		**Data holders/ Data providers**
Cost and time savings, including through digitization of interaction with data (re)users by certified intermediaries and quicker access to data suppliers (in the case of certified intermediaries who provide data pooling services) as well as efficiency gains when sharing supply chain data through a data intermediary	*Not quantifiable due to lack of data*		**Data holders/ Data providers, Data (re)users**
Growing network and additional revenue through easier access to data (re)users (in the case of certified intermediaries who provide data pooling services) and larger client base (especially in the case of intermediaries providing data pooling services).	*Not quantifiable due to lack of data*		**Data holders/ Data providers, (Certified) data intermediaries, Data (re)users**
Data users obtain personal data with legal clarity on usability (consented data). Creation of mechanism for intermediaries' clients to assess the quality of intermediaries' operations (i.e. in terms of compliance with legislation) Additional revenue and increased productivity generated through the increased volume of re-used data (unlocked and facilitated through the certified intermediaries)	*Not quantifiable due to lack of data*		**Data (re)users**
Indirect benefits			
Increased market competition – B2B Market	• between 1%-25% increase from 2 to 5 years' timeframe, • between 1%-25% increase beyond 5 years' timeframe;		**Data Intermediaries**
Increased market competition – C2B market	• between 1%-25% increase within the 1st year after obtaining the certification, • between 1%-25% increase from 2 to 5 years' timeframe, • between 25%-50% increase beyond 5 years' timeframe		**Data Intermediaries**
Additional revenue potentially generated through the "re-use" of the data unlocked and facilitated through the certified intermediaries	*Not quantifiable due to lack of data*		**Data holders/ Data providers**
Enhancing of trust between the main actors involved	*Not quantifiable due to lack of data*		**(Certified) Data Intermediaries, Data holders/ Data providers, Data (re)users**
Effect on Gross Domestic Product (GDP)	*Not quantifiable due to lack of data*		
Economic value of data (both personal and industrial) will be maximised and additional revenue potentially	*Not quantifiable due to lack of data*		**Data (re)users, Data holders/ Data providers**

generated through the re-use of the
data facilitated through the certified
intermediaries

4.5.4.2.3 Findings of the Cost-Benefit Analysis for Policy Option 2

Policy Option 2 entails significantly high levels of benefits including business development time acceleration, client base and revenues increases, market competition, as a result of the increased trust between the stakeholders in the market, after certification. Recurrent and one-off costs are expected as well for data intermediaries in order to obtain and maintain the certification, however, the expected benefits significantly outweigh the expected costs. This policy option presents a great cost-benefit relation, with benefits exceeding more than 10 times the costs. Details on the costs and benefits for policy option 2 are available in Annex C.

4.5.4.3 Coherence of the option

This policy option could promote the objectives of the single market for data under the European Data Strategy and therefore remains coherent with the current EU legal and policy framework.

4.5.5 Policy option 3: Regulatory intervention with high intensity (compulsory certification framework)

4.5.5.1 Effectiveness in achieving the policy objectives

This subsection examines the effectiveness of policy option 3 in achieving the specific and general policy objectives.

4.5.5.1.1 Achievement of specific objectives

This policy option is expected to contribute to some of the specific objectives, namely creating trust in common European data spaces as well as contributing to data interoperability across sectors. However, concerns were raised regarding its effectiveness to build common data spaces, due to the fact that a compulsory certification process with hard neutrality requirements is likely to prevent small industry players from getting into the market due to the potentially prohibitive certification cost.

4.5.5.1.2 Achievement of general objectives

In the same line of argumentation, there are doubts whether the overall impact of a compulsory certification framework would be positive by boosting the market, since it could create significant burdens for new players to get into the market. It is therefore doubtful whether this policy option could further contribute to set the foundations of a Single Market for Data, and furthermore, strengthen the EU data economy.

4.5.5.2 Efficiency: Costs and benefits of the option

This subsection examines the efficiency of policy option 3 in terms of costs and benefits. The detailed results of the cost-benefit analysis for the sub-task 1.4 can be found in Annex C.

4.5.5.2.1 Costs of the option

The types of costs under this policy option are similar to those of the previous policy options. However, the hard neutrality requirements of a compulsory certification framework are expected to increase the level of such costs. In terms of direct costs, data intermediaries would have in this case as well to bear the cost of **obtaining and maintaining the certification**. Depending on the requirements and the criteria defined in the legal instrument, some intermediaries would also have to face implementation costs to ensure compliance with the requirements. In particular, the interviewed stakeholders estimate that certification costs under this policy option would be of 35K – 75K EUR one-off costs for obtaining the certification for a first time (including 25K-50K EUR internal preparatory costs and 10K-25K EUR external certification costs)

and of 20K-50K EUR/year recurrent costs for renewing it (including 10K-25K EUR/year internal preparatory costs and 10-25K EUR/year external certification costs).

In terms of indirect costs, concerns are raised by data intermediaries in relation to the innovation in the market that might be limited, due to hard neutrality requirements defined by such a compulsory framework. Competition in the market might also be reduced in a short-term timeframe, as some companies might fail to enter and remain in the market due to the high certification cost while others in early stage might fail to afford the high certification cost. In particular, within the first year timeframe after the establishment of the certification framework, the interviewed stakeholders expect a decrease of competition between 50% - 25% in B2B market and of 25%-1% in the C2B market. Data re-users are also likely to be affected by **indirect transaction and implementation costs**, as some of the certified intermediaries might increase the user charges to balance the certification cost.

Finally, stakeholders expressed concerns that a compulsory certification framework would be prohibitive for data intermediaries which are not necessarily legal entities, but the intermediation is provided through the form of a project or product. Under this policy option, this type of data intermediation initiatives would have to be locked out of the market.

Table 45 – Overview of costs for subtasks 1.4 | PO 3[169]

Overview of costs – PO 3		Data holders		Data intermediaries		Data re(users)	
		One-off	*Recurrent*	*One-off*	*Recurrent*	*One-off*	*Recurrent*
Certification framework for data intermediaries	*Direct costs*	-	-	35K – 75K EUR •25K-50K EUR internal preparatory costs •10-25K EUR external certification costs	20K-50K EUR/year •10-25K EUR/year internal preparatory costs •10-25K EUR/year external certification costs	-	-
	Indirect costs	-	-	-	• Between 50% and 25% decrease within the 1st year after obtaining the certification in B2B market; • between 25%-1% decrease within the 1st year after obtaining the certification in C2B market	-	Non-quantifiable costs due to lack of data

---

[169] The table presents the estimated amounts of costs by the interviewed stakeholders

4.5.5.2.2 Benefits of the option, including reductions in some of the costs as well as other positive effects on (some of) the stakeholders

Even though the majority of stakeholders involved was not in favour of this policy option, due to the additional burdens it might create in a yet not mature market, this option still presents a broad number of benefits similar to the policy options 1 and 2, even at higher levels in some cases. These are interwoven to the increased trust between the actors that would be brought in the ecosystem through compulsory certification. It would lead to further time and cost savings for data holders and re-users as well as efficiency gains, increased client base and volumes of data transactions. This would increase revenues for data intermediaries, allowing them to scale up. Additionally, it will guarantee uniform and transparent security levels, interoperability as well as safety and quality of data intermediaries' services. The role of newly established data intermediaries in the market such as data unions or data cooperativ es would still be empowered, within their ecosystem, through this legal framework which would provide legitimacy to their work. Additional benefits are expected for data intermediaries who already respect certain neutrality requirements, as they would have a competitive advantage towards the other players, without having to bear a high level of costs for the certification. In particular, the benefits expected by the interviewed stakeholders under this policy option include:

- 45%-50% business development time acceleration
- Growth in terms of revenues and employee number:
  - between 25%-35% increase, within the 1st year after obtaining the certification;
  - between 40%-50% increase, from 2 to 5 years' timeframe;
  - between 40%-50% increase, beyond 5 years' timeframe
- Growth in terms of client base (number of clients)
  - between 25%-50% increase within the 1st year after obtaining the certification;
  - approximately 50% increase from 2 to 5 years' timeframe;
  - approximately 50% increase beyond 5 years' timeframe.
- Growth in terms of client base (number use cases)
  - between 25%-50% increase within the 1st year after obtaining the certification;
  - between 40%-50% increase from 2 to 5 years' timeframe;
  - between 40%-50% increase beyond 5 years' timeframe.

With regards to the indirect benefits expected under this policy option, increased competition benefits are expected to be met in both C2B and B2B value chains, especially in a long-term timeframe. In particular, the interviewed stakeholders expect:

- between 1%-25% increased competition in the B2B market from 2 to 5 years' timeframe;
- between 25%-50% increased competition in the B2B market beyond 5 years' timeframe;
- between 1%-25% increased competition in the C2B market, from 2 to 5 years' timeframe;
- between 25%-50% increased competition in the C2B market, beyond 5 years' timeframe.

Data holders would still have the opportunity to monetise more from data sharing, while they will have easier access to a bigger network of data re-users. At the same time increased number of individuals will be attracted to share their personal data through the certified platforms. Data holders as well as data re-users would also have cost and time savings through digitization of interaction with data re-users by certified intermediaries, while mechanisms for assessing the quality of services provided by data intermediaries will be created for them through certification.

Several interviewed stakeholders, mainly in the C2B market, expressed the view that, even though the European data market is not mature at this stage for the establishment of a compulsory certification framework, in a long-term timeframe this policy option might become the preferred one. This is also based

on the idea that compulsory certification will initially decrease the number of suitable data intermediaries, however it will significantly increase confidence in the market and fair market competition in the long-term, with same rules applying to everyone.

Table 46 – Overview of benefits – Subtask 1.4 | PO 3[170]

Type of action	Description	Amount	Stakeholders
Certification framework for data intermediaries	**Direct benefits**		
	Time savings and business development cycle acceleration; speed of adoption of the market	45%-50% business development time acceleration *expected after certification*	**Data Intermediaries**
	Company growth in terms of revenue and number of employees	• between 25%-35% increase within the 1st year after obtaining the certification; • between 40%-50% increase from 2 to 5 years' timeframe; • between to 40%-50% increase beyond 5 years' timeframe • 20% increase corresponds to 30M-50M EUR/year revenue increase for a growth-stage intermediary	**Data Intermediaries**
	Company growth in terms of client base (number of clients)	• between 25%-50% increase within the 1st year after obtaining the certification; • 50% increase from 2 to 5 years' timeframe; • 50% increase beyond 5 years' timeframe	**Data Intermediaries**
	Company growth in terms of client base (number of use cases)	• between 25%-50% increase within the 1st year after obtaining the certification; • between 40%-50% increase from 2 to 5 years' timeframe; • between 40%-50% increase beyond 5 years' timeframe	**Data Intermediaries**
	Data monetisation via data sharing through certified platforms	*Not quantifiable due to lack of data*	**Data holders/ Data providers**
	Cost and time savings, including through digitization of interaction with data (re)users by certified intermediaries and quicker access to data suppliers (in the case of certified intermediaries who provide data pooling services) as well as efficiency gains when sharing supply chain data through a data intermediary	*Not quantifiable due to lack of data*	**Data holders/ Data providers, Data (re)users**
	Growing network and additional revenue through easier access to data (re)users (in the case of certified intermediaries who provide data pooling services) and larger client base (especially in the case of intermediaries providing data pooling services).	*Not quantifiable due to lack of data*	**Data holders/ Data providers, (Certified) data intermediaries, Data (re)users**
	Data users obtain personal data with legal clarity on usability (consented data).	*Not quantifiable due to lack of data*	**Data (re)users**

---

[170] The table presents the estimated amounts of benefits by the interviewed stakeholders

Creation of mechanism for intermediaries' clients to assess the quality of intermediaries' operations (i.e. in terms of compliance with legislation) Additional revenue and increased productivity generated through the increased volume of re-used data (unlocked and facilitated through the certified intermediaries)		
Indirect benefits		
Increased market competition – B2B Market	• between 1%-25% increased competition from 2 to 5 years' timeframe, • between 25%-50% increased competition beyond 5 years' timeframe;	**Data Intermediaries**
Increased market competition – C2B market	• between 1%-25% increase from 2 to 5 years' timeframe, • between 25%-50% beyond 5 years' timeframe	**Data Intermediaries**
Additional revenue potentially generated through the "re-use" of the data unlocked and facilitated through the certified intermediaries	*Not quantifiable due to lack of data*	**Data holders/ Data providers**
Enhancing of trust between the main actors involved	*Not quantifiable due to lack of data*	**(Certified) Data Intermediaries, Data holders/ Data providers, Data (re)users**
Effect on Gross Domestic Product (GDP)	*Not quantifiable due to lack of data*	
Economic value of data (both personal and industrial) will be maximised and additional revenue potentially generated through the re-use of the data facilitated through the certified intermediaries	*Not quantifiable due to lack of data*	**Data (re)users, Data holders/ Data providers**

4.5.5.2.3 Findings of the Cost-Benefit Analysis for Policy Option 3

Policy Option 3 entails significantly high levels of benefits including business development time acceleration, client base and revenues increases, market competition, as a result of the increased trust between the stakeholders in the market, after certification. Increased recurrent and one-off costs are expected as well for data intermediaries in order to obtain and maintain the certification, however, the expected benefits significantly outweigh the expected costs. This option presents a great cost-benefit relation as well, as the benefits exceed more than 10 times the costs. Details on the costs and benefits for policy option 2 are presented in Annex C.

4.5.5.3 Coherence of the option

This policy option could promote the objectives of the single market for data under the European Data Strategy and therefore remains coherent with the current EU legal and policy framework.

4.5.6 Summary of the impacts

The following table summarises the possible impacts of the policy options:

Table 47 - Summary of impacts for subtask 1.4

Economic impacts	• Data intermediaries would be affected from the certification cost. However, certification would enable them to grow in terms of revenue, employees and client base. Competition and innovation in the data intermediaries market might also be affected. • Data holders and re-users will mainly benefit from efficiency gains including cost and time savings through digitisation of the transactions and facilitation of data sharing via the certified intermediaries.
Social impacts	• Societal benefit will be twofold. On the one hand society will benefit as the potential of the European data market will be unlocked through certification. On the other hand, data flows of intermediaries serving societal purposes (i.e. health, research) will be increased
Environmental impacts	• N/A
Fundamental rights impacts	• Protection of privacy and personal data will be promoted through the certification of data intermediaries, especially of personal data spaces.

5 Comparison of the policy options

The aim of this chapter is to compare of the policy options in order to identify the preferred policy option for each of the subtask.

The following MCA has been performed in line with the European Commission's *Better Regulation Guidelines*[171] and its toolbox[172], most importantly tool 63[173]. The assessment builds on the prior analysis of each individual option.

It has been concluded in the previous chapter that for none of the area under investigation the baseline will be able to achieve the desired results and resolved identify problems. The assessment concludes that a policy intervention is needed. It remains to be seen the type (regulatory vs. non regulatory) and the intensity (low vs. high) of intervention. The MCA will assess which of the three policy options under each area is the most adequate:

- Subtask 1.1: Measures facilitating secondary use of sensitive data held by the public sector
  - PO 1: Guidelines
  - PO 2: One-Stop-Shop
  - PO 3: Single Data authorisation Body
- Subtask 1.2: Establishing a certification/authorisation scheme for data altruism mechanisms
  - PO 1: Coordination at EU level
  - PO 2: Voluntary certification scheme
  - PO 3: Mandatory authorisation
- Subtask 1.3: Establishing a European structure for certain governance aspects of data sharing
  - PO 1: Informal expert group
  - PO 2: Formal expert group
  - PO 3: Legal body
- Subtask 1.4: Establishing a certification framework for data intermediaries
  - PO 1: Industry driven certification framework
  - PO 2: Voluntary certification framework
  - PO 3: Compulsory certification framework

The MCA was carried out in the following three distinct steps:
- *Step 1:* Establish indicators or assessment criteria against which the policy options are assessed and compared. This includes establishing the performance of a policy option (i.e. the magnitude of its impact), the weight of the criteria in relation to each other, as well as the direction of the impact (negative/positive). The indicators are established in an analytical grid;
- *Step 2:* Build an outranking matrix in which the scores for all policy options and criteria are provided in order to summarise how the policy options compare with each other in relation to established criteria; and
- *Step 3:* Prepare a permutation matrix that enables the selection of a final ranking of all the possible policy options against each other for each subtask. This means that it is possible not only to select a preferred policy option but also a ranking of all other options against each other.

---

[171] http://ec.europa.eu/smart-regulation/guidelines/toc_guide_en.htm
[172] http://ec.europa.eu/smart-regulation/guidelines/toc_tool_en.htm
[173] https://ec.europa.eu/info/sites/info/files/file_import/better-regulation-toolbox-63_en_0.pdf

5.1 Step 1: Establishing assessment criteria and indicators

The following assessment criteria were agreed with the European Commission for the assessment of the impacts of the options. A weight has been defined for each criterion. The direction of the change desired are all positive. The proportionality assessment criteria is considered as an exclusion criteria, and is therefore not included in the MCA.

Table 48 – Weight, direction and performance value allocated to the assessment criteria

Assessment criterion	Weight	Direction	Performance value
Effectiveness	0.3	1	Qualitative +/-4 scale
Efficiency	0.3	1	Benefit/Cost-ratio (BCR)
Coherence	0.25	1	Qualitative +/-2 scale
Legal and political feasibility	0.15	1	Qualitative +/-2 scale
Proportionality	This exclusion criteria will not be assessed as part of the MCA	N/A	N/A

Based on the results of the Cost-Benefit analysis and the qualitative assessment of each individual options, we have drafted an **input grid** for each subtask in which the scores for all policy options are collected and compared in relation to each criterion towards each other.

Table 49 – Input Matrix

Input matrix			Subtask 1.1					
			PO 1 - Non regulatory		PO 2 - Low intensity		PO 3 - High intensity	
Critera	Weight	Direction	Performance	Weighted performance	Performance	Weighted performance	Performance	Weighted performance
Effectiveness of the policy options in reaching the specific and general policy objectives	0,3	1	1	0,3	2	0,6	3	0,9
Efficiency (BCR)	0,3	1	0	0	6	1,8	2,4	0,72
Coherence of the policy options	0,25	1	0,5	0,125	1	0,25	-1	-0,25
Legal and Political feasibility	0,15	1	1	0,15	0,5	0,075	-1	-0,15
			Subtask 1.2					
			PO 1 - Non regulatory		PO 2 - Low intensity		PO 3 - High intensity	
Critera	Weight	Direction	Performance	Weighted performance	Performance	Weighted performance	Performance	Weighted performance
Effectiveness of the policy options in reaching the specific and general policy objectives	0,3	1	1,5	0,45	2	0,6	2	0,6
Efficiency	0,3	1	0	0	2,7	0,81	6,3	1,89
Coherence of the policy options	0,25	1	1	0,25	1	0,25	1	0,25
Legal and Political feasibility	0,15	1	1	0,15	1	0,15	1	0,15
			Subtask 1.3					
			PO 1 - Non regulatory		PO 2 - Low intensity		PO 3 - High intensity	
Critera	Weight	Direction	Performance	Weighted performance	Performance	Weighted performance	Performance	Weighted performance
Effectiveness of the policy options in reaching the specific and general policy objectives	0,3	1	1	0,3	2	0,6	2	0,6
Efficiency	0,3	1	200 362,19	60108,657	19627,32	5888,196	1766,46	529,938
Coherence of the policy options	0,25	1	-1	-0,25	1	0,25	1	0,25
Legal and Political feasibility	0,15	1	1	0,15	1	0,15	-1	-0,15
			Subtask 1.4					
			PO 1 - Non regulatory		PO 2 - Low intensity		PO 3 - High intensity	
Critera	Weight	Direction	Performance	Weighted performance	Performance	Weighted performance	Performance	Weighted performance
Effectiveness of the policy options in reaching the specific and general policy objectives	0,3	1	1	0,3	2	0,6	1	0,3
Efficiency	0,3	1	5,21	1,563	2,68	0,804	2,68	0,804
Coherence of the policy options	0,25	1	0,5	0,125	1	0,25	1	0,25
Legal and Political feasibility	0,15	1	1	0,15	1	0,15	1	0,15

5.1.1 **Subtask 1.1: Measures facilitating secondary use of sensitive data held by the public sector**

The analysis above shows that of all policy options, the single data authorisation body (Policy Option 3) is the option likely to achieve best the policy objectives, as it is also the most ambitious. This policy option combines the establishment of a one-stop shop and of secure data processing environments (Policy Option 2) with additional tasks designed to facilitate secondary re-use of data the use of which is subject to the rights of others. As a consequence, it brings more benefits, but also additional costs.

However, the ambition of policy option 3 to enable the re-use of data the use of which is subject to the rights of others for commercial purposes is incompatible with the national legislation of some Member States. As noted above, a single secure data processing environment combining data the use of which is subject to the rights of others from different holders may not be permitted by some national laws. The introduction of a single data authorisation body can also damage public trust. In one Member States, such establishment has been stranded because of trust-related considerations.

Policy option 2 – establishing a one-stop shop and data processing environments is both feasible and proportionate: the tasks linked with data discovery and with advisory services performed by the one-stop shops appear feasible and proportionate in relation to the policy objectives (although the provision of advice specific to datasets should remain the responsibility of data holders), while limiting sensitive data re-use to research and to commercial purposes serving a public interest also appears proportionate (and appears to correspond to the legal status quo in a large number of Member States). Likewise, allowing Member States to design more than one data processing environment as needed, appears proportionate in view of the policy objectives, and is likely to be feasible/compatible with existing national laws. As pointed out by participants to the workshop held on 8 July 2020, the very different nature of datasets in scope (i.e. in health, and other sectors) militate for several data processing environments coexisting. In addition, policy-option 2 is the most efficient, with a Benefit-Cost ratio higher than policy option 3.

When asked their view, **participants to the 8 July workshop expressed a strong preference for PO2** over the other policy options, and a slight preference for PO3 over PO1. The following table includes short description of how the low/high intensity options compare in terms of efficiency, effectiveness, coherence, legal/political feasibility and proportionality.

Table 50 – Summary comparison table between low/high intensity policy options for subtask 1.1

	Regulatory intervention with low intensity (PO2)	Regulatory intervention with high intensity (PO3)
Efficiency	This option would bring costs likely ranging in millions of euros for each Member State. However, it would also result in very significant savings that are likely to significantly outweigh the costs.	This option would create higher costs than PO2 (linked to processing data access applications), but would also yield higher benefits for both data holders and data reusers. Opening the re-use of sensitive data to commercial purposes would likely produce innovation and growth, but comes with risks linked to trust – risks which could undermine these benefits. Overall, the Benefit-Cost ratio of PO3 is lower than that of PO2.
Effectiveness	Contributes to achieving the specific and general objectives.	Contributes more than PO2 to achieving the specific and general objectives, as it also includes centralisation of data access applications.
Coherence	Coherent with EU law and with national laws.[174]	Coherent with EU law. Enabling re-use of data for strictly commercial purposes would be incompatible with national laws of several Member States. A single data processing

---

[174] The national laws of stakeholders interviewed.

		environment would be incompatible with national law of some Member States.[175]
Legal/political feasibility	This option appears to be feasible.	There are doubts as to the political feasibility of this option in some countries (e.g. where public trust is government is lower).
Proportionality	This option appears proportionate.	A single data processing environment would be disproportionate, particularly in view of the very different types of datasets concerned.

As regards guidelines/recommendations (policy option 1), it remains unclear whether these would result in a larger number of Member States setting up structures to enhance the reuse of data the use of which is subject to the rights of others compared to a no intervention scenario (i.e. whether many Member States would do so *as a result of* these recommendations).

5.1.2 Subtask 1.2: Establishing a certification/authorisation scheme for data altruism mechanisms

The following table includes short description of how the low/high intensity options compare in terms of efficiency, effectiveness, coherence, legal/political feasibility and proportionality.

Table 51 – Summary comparison table between low/high intensity policy options for subtask 1.2

	Regulatory intervention with low intensity (PO2)	Regulatory intervention with high intensity (PO3)
Efficiency	This option would bring substantial costs to Member States, however also increasing benefits.	This option would bring substantial costs to Member States, however also increasing benefits.
Effectiveness	This option would increase transparency for and decrease security concerns of citizens.	This option would increase transparency for and decrease security concerns for citizens.
Coherence	Coherent with EU law as there is currently no EU law on data altruism.	Coherent with EU law as there is currently no EU law on data altruism.
Legal/political feasibility	This option appears to be feasible.	This option appears to be feasible.
Proportionality	This option appears to be proportionate.	This option appears to be proportionate.

The preliminary analysis of the different policy options shows there is great potential for data altruism mechanisms in the European Union considering that the COVID-19 crisis catapulted data altruism in the limelight, it also uncovered the national differences amongst Member States on core discussion points such as data privacy and the discussion of costs and benefits of such mechanisms. This was highlighted by the different approaches Member States took to implement data altruism mechanisms, if at all.

Of the four policy options, the baseline scenario is the least effective and efficient to achieve the general objectives. It would likely exclude SMEs and organisations from data altruism because only large cooperation's would have the necessary resources to handle legal fragmentation across member states and a European data economy including data altruism would be difficult to achieve. This would be an economic and societal loss.

Policy option 1 would be a positive step towards achieving the specific and general objectives, however considering the fast-paced developments in this field, it could also have a very limited effectiveness since cooperation or expert groups and voluntary soft regulation can be lengthy process.

Policy option 2 and 3 appear to be the most effective and efficient options to achieve the general and specific outcomes. Whereas option 2 and 3 would both cost member states, option 3 is less favourable because the

---

[175] The national laws of stakeholders interviewed.

costs would be significantly higher and could create additional regulatory burden for member states. Nevertheless, in the light of the COVID-19 crisis, the benefits of data altruism schemes have been highlighted in the public and political debate and the Commission could use this momentum to highlight the expected benefits and how this could outweigh costs.

5.1.3 Subtask 1.3: Establishing a European structure for governance aspects of data sharing

The analysis of the different options allows to draw a preliminary comparison of the different options, based on the current available evidence.

The following table includes short description of how the low/high intensity options compare in terms of efficiency, effectiveness, coherence, legal/political feasibility and proportionality.

Table 52 – Summary comparison table between low/high intensity policy options for subtask 1.3

	Regulatory intervention with low intensity (PO2)	Regulatory intervention with high intensity (PO3)
Efficiency	The costs of a formal expert group are limited, while the benefit are similar to PO3: an expert group will have a limited positive effect over the development and adoption of standardisation initiatives	The cost of a legal body are typically higher than a formal expert group (5 to 10 times higher), while the benefits are similarly positive.
Effectiveness	The overall effect on data sharing are moderately positive – there is a very mediated relation between an expert group and the adoption of data sharing standards	Equivalent to PO2. It is unlikely that a legal body status would visibly make a difference when it comes to business adoption of the standards
Coherence	Strong coherence with digital strategy, notably with the creation of data spaces.	Equivalent to PO2
Legal/political feasibility	High. Positive reception by industry interviewees, who are reluctant to strong EU role in standardisation but calls for soft support and stakeholders engagement	Moderate. Industry as well as standardisation initiatives are suspicious towards government led efforts in standardisation. Also, the creation of yet another legal entity would have to be strongly justified internally.
Proportionality	High. The intervention is very limited and proportionate to the stakes at hand	High. The intervention is stronger and proportionate to the stakes at hand

First and foremost, the specific and general objectives are very demanding and ambitious. They concern the internal decision of how companies behave when it comes to data standardization and sharing. In addition, the options have to take into account the multiplicity of actors already carrying out activities in this area.

As such, it can be expected that no single option can deliver the specific and general objective, as their achievement is based on how companies react to the policy option. It will take time and a multiplicity of intervention to achieve them.

Option 0 appears as the worse option, not just because it will be less efficient and effective, but it has possible negative effects because of the fast-evolving nature of the economic context. Many other countries worldwide are active in data standardization, and there is the risk that Europe follows the lead of others, with all the negative consequences in terms of competitiveness. In other words, there is the risk that after personal data, also industrial data become controlled by non-European players.

Furthermore, option 0 is inconsistent with the initiatives in place on data spaces. The absence of activity on data standardization will weaken the effectiveness of the data spaces.

Option 1, 2 and 3 have limited differences – basically around an informal, formal expert group or a legal body. An informal expert group is unlikely to have an effect at all, while the legal body appears overly heavy from a bureaucratic point of view, considering the wealth of existing bodies working on standardization. In

other words, option 1 has limited costs, but also very little impact. Option 3 on the other hand seem to achieve a comparable impact to option 2 in terms of standardisation, but with increased costs related to the creation of a new legal entity or the expansion of an existing one and lower political feasibility.

5.1.4 Subtask 1.4: Establishing a certification framework for data intermediaries

This section provides intermediary conclusions with regard to the policy options for Subtask 1.4. Whereas this measurement unit cannot be added, a qualitative summary assessment is provided in the Average row of the table.

The following table includes short description of how the low/high intensity options compare in terms of efficiency, effectiveness, coherence, legal/political feasibility and proportionality.

Table 53 – Summary comparison table between low/high intensity policy options for subtask 1.4

	Regulatory intervention with low intensity (PO2)	Regulatory intervention with high intensity (PO3)
Efficiency	PO 2 presents the best balance between costs and benefits for the stakeholders and appears to be the preferred policy option by the interviewed stakeholders.	PO 3 presents a broad number of benefits for the stakeholders, at similar levels to the benefits of PO 2. However, it is doubtful whether the European market is mature enough for the establishment of a compulsory certification framework. There are concerns that the latest might impose unnecessary burdens to data intermediaries, leading in the end to opposite results from the desired ones.
Effectiveness	PO 2 could significantly contribute to the three specific objectives, particularly, in creating trust in common European data spaces, building common data spaces as well as ensuring data interoperability across sectors, through the certification framework. The majority of stakeholders interviewed agreed to this policy option as the most effective one given that the certification criteria would be defined by a legal instrument. Therefore, if a big number of industry players decides to proceed to the certification process, the trust between the stakeholders involved in the market would be increased significantly, allowing the data intermediary market to flourish and bringing various economic and societal benefits. Following the effectiveness in achieving the specific objectives described in the above section, this policy option would also further contribute to set the foundations of a Single Market for Data, and furthermore, strengthen the EU data economy, since the European data market overall will be significantly boosted through certification, increasing the volume of data flows.	PO3 is expected to contribute to some of the specific objectives, namely creating trust in common European data spaces as well as contributing to data interoperability across sectors. However, concerns were raised regarding its effectiveness to build common data spaces, due to the fact that a compulsory certification process with hard neutrality requirements is likely to prevent small industry players from getting into the market due to the potentially prohibitive certification cost. In the same line of argumentation, there are doubts whether the overall impact of a compulsory certification framework would be positive by boosting the market, since it could create significant burdens for new players to get into the market. It is therefore under question whether this policy option could further contribute to set the foundations of a Single Market for Data, and furthermore, strengthen the EU data economy.
Coherence	This policy option could promote the objectives of the single market for data under the European Data Strategy and therefore remains coherent with the current EU legal and policy framework.	This policy option could promote the objectives of the single market for data under the European Data Strategy and therefore remains coherent with the current EU legal and policy framework.
Legal/political feasibility	This policy option appears to be legally and politically feasible to be adopted and implemented at European level,	This policy option appears to be legally and politically feasible to be adopted and implemented at European level,

Proportionality	This policy option is proportionate as its intensity matches the identified problem and objectives of this study	This policy option is not proportionate as its intensity is deemed too strong for the identified problem and objectives of this study.

The majority of stakeholders involved expect benefits and costs in all three policy options while they are not in favour of policy option 0 meaning no action is taken at EU level.

The preferred policy option seems to be Policy Option 2 establishing a voluntary certification framework through a legal instrument by the European Commissions. Even though the level of costs and benefits remain common and similar for all the three policy options, stakeholders expressed concerns that policy option 1 might be too weak compared to the other two, in which case it would place it close to the effectiveness levels of policy option 0. On the other side, according to the stakeholders, it is doubtful whether the European market is mature enough for the establishment of a compulsory certification framework. There are concerns that the latest might impose unnecessary burdens to data intermediaries, leading in the end to opposite results from the desired ones.

5.2   Step 2: Building an outranking matrix to compare policy options

In relation to Step 2, the following table provides an **outranking matrix** in which all the weights indicated in the table under step 1 are totalled for the criteria in relation to which a policy option is favoured over another policy option (abbreviated e.g. as "PO1/PO2") as indicated by the weighted performance of each criterion.

This means that the outranking matrix provides an overview of the overall scores of the policy options compared to each other (i.e. the differences between them).

Table 54 – Outranking matrix

Outranking matrix			
Subtask 1.1	PO 1	PO 2	PO 3
PO 1	**0**	**0,15**	**0,4**
PO 2	**0,85**	**0**	**0,7**
PO 3	**0,6**	**0,3**	**0**
Subtask 1.2	PO 1	PO 2	PO 3
PO 1	**0**	**0**	**0**
PO 2	**0,6**	**0**	**0**
PO 3	**0,6**	**0,3**	**0**
Subtask 1.3	PO 1	PO 2	PO 3
PO 1	**0**	**0,3**	**0,45**
PO 2	**0,55**	**0**	**0,45**
PO 3	**0,55**	**0**	**0**
Subtask 1.4	PO 1	PO 2	PO 3
PO 1	**0**	**0,3**	**0,3**
PO 2	**0,55**	**0**	**0,55**
PO 3	**0,25**	**0**	**0**

Naturally, the grey combinations received a score of 0 as it does not make sense to compare these. In essence, the table shows that the impacts of the policy options outrank those of the baseline scenario and that policy options with a higher score outrank those with a lower score.

The differences between the overall rankings of each policy option between each other as presented above are derived from the sum of the individual scores per policy option and assessment criterion in the analytical grid.

5.3 Step 3: Preparing a permutation matrix to identify the order of preferences

The table below present the six different combination of policy options for the four areas under investigation.

Table 55 – Policy ranking permutation

Policy ranking permutation	Policy packages	Coefficients of policy packages	Final score
Subtask 1.1			
PO1/PO2/PO3	PO1/PO2 + PO1/PO3 + PO2/PO3	0,15 + 0,4 + 0,7	1.25
PO1/PO3/PO2	PO1/PO3 + PO3/PO2 + PO1/PO2	0,4 + 0,3 + 0,15	0.85
PO2/PO1/PO3	*PO2/PO1 + PO1/PO3 + PO2/PO3*	*0,85 + 0,4+ 0,7*	*1.95*
PO2/PO3/PO1	PO2/PO3 + PO3/PO1 + PO3/PO2	0,7 + 0,6 + 0,3	1.6
PO3/PO1/PO2	PO3/PO1 + PO1/PO2 + PO3/PO2	0,6 + 0,15 + 0,3	1.05
PO3/PO2/PO1	PO3/PO2 + PO3/PO1 + PO2/PO1	0,3 + 0,6 + 0,85	1.75
Subtask 1.2			
PO1/PO2/PO3	PO1/PO2 + PO1/PO3 + PO2/PO3	0 + 0 + 0	0
PO1/PO3/PO2	PO1/PO3 + PO3/PO2 + PO1/PO2	0 + 0,3 + 0	0.3
PO2/PO1/PO3	PO2/PO1 + PO1/PO3 + PO2/PO3	*0,6 + 0 + 0*	*0.6*
PO2/PO3/PO1	PO2/PO3 + PO3/PO1 + PO3/PO2	0 + 0,6 + 0,3	0.9
PO3/PO1/PO2	PO3/PO1 + PO1/PO2 + PO3/PO2	0,6 + 0 + 0,3	0.9
PO3/PO2/PO1	*PO3/PO2 + PO3/PO1 + PO2/PO1*	*0,3 + 0,6 + 0,6*	*1.5*
Subtask 1.3			
PO1/PO2/PO3	PO1/PO2 + PO1/PO3 + PO2/PO3	0,3 + 0,45 + 0,45	1.2
PO1/PO3/PO2	PO1/PO3 + PO3/PO2 + PO1/PO2	0,45 + 0 + 0,3	0.75
PO2/PO1/PO3	*PO2/PO1 + PO1/PO3 + PO2/PO3*	*0,55 + 0,45 + 0,45*	*1.45*
PO2/PO3/PO1	PO2/PO3 + PO3/PO1 + PO3/PO2	0,45 + 0,55 + 0	1
PO3/PO1/PO2	PO3/PO1 + PO1/PO2 + PO3/PO2	0,55 + 0,3 + 0	0.85
PO3/PO2/PO1	PO3/PO2 + PO3/PO1 + PO2/PO1	0 + 0,55 + 0,55	1.1
Subtask 1.4			
PO1/PO2/PO3	PO1/PO2 + PO1/PO3 + PO2/PO3	0,3 + 0,3 + 0,55	1.15
PO1/PO3/PO2	PO1/PO3 + PO3/PO2 + PO1/PO2	0,3 + 0 + 0,3	0.6
PO2/PO1/PO3	*PO2/PO1 + PO1/PO3 + PO2/PO3*	*0,55 + 0,3 + 0,55*	*1.4*
PO2/PO3/PO1	PO2/PO3 + PO3/PO1 + PO3/PO2	0,55 + 0,25 + 0	0.8
PO3/PO1/PO2	PO3/PO1 + PO1/PO2 + PO3/PO2	0,25 + 0,3 + 0	0.55
PO3/PO2/PO1	PO3/PO2 + PO3/PO1 + PO2/PO1	0 + 0,25 + 0,55	0.8

This means the following:

- For subtask 1.1, **policy option PO2 – One-stop-shop** is the preferred option as it providers the most combination of effectiveness, efficiency and coherence.
- For subtask 1.2, **policy option PO3 – mandatory authorization scheme** for data altruism is the preferred option;
- For subtask 1.3**, policy option PO2 – formal expert group** is the preferred option;
- For subtask 1.4**, policy option PO2 – Voluntary certification framework** for intermediaries is the preferred option.

It must be noted that although PO1s for the four areas, rank rather well in comparison to the other policy options, the actual amounts of costs and benefits expected by a non-regulatory intervention are extremely limited. Additionally, it is assumed that the schemes foreseen under subtask 1.2 and 1.4 would rely on either the one-stop shop mechanisms set up by Member States (PO2) or a private conformity assessment body, accredited by the European Data Innovation Board (PO3) to issue the respective certifications and authorizations to data intermediaries.

6 Assessment of macro-economic impacts

This section consists in comparing the expected macro-economic impacts of a low intensity, high intensity regulatory intervention and the preferred association of policy options on the overall economy and society compared to the baseline scenario.

6.1 Definition of policy options packages

The assessment of impacts on the overall data economy and society can only be performed at an aggregated level, by creating policy options packages composed of one policy option per subtasks (area). Based on the multi-criteria analysis performed under Chapter 5, the fact that the PO1 (non-regulatory options) were deemed to create low impacts in terms of costs and benefits, and taking into account the interdependences between the policy options (reliance on PO2 or PO3 under subtask 1.1 to support the certification mechanisms linked to data altruism and data intermediaries), four policy option packages where identified:

- **Policy Package 0 – Baseline**: the baseline scenario consists in applying no policy changes to the four areas for which problems could be identified: sensitive data held by the public sector, data altruism schemes, certification of data intermediaries, governance and standards. The EU economy will not be able to reap the benefits of data sharing.
- **Policy Package 1** – Low intensity regulatory options: this package includes the creation of a one-stop shop to foster the sharing of (sensitive) data whose use is subject to the rights of others held by the public sector. A voluntary certification scheme would be established by EU Member States for data altruism mechanism and organisations offering such schemes. Data intermediaries will also be able to obtain a certification to demonstrate their neutrality and absence of conflict of interest (e.g. absence of competition with data users) on a voluntary basis. Finally, the European Data innovation Board would take the form of a formal expert group created by the European Commission, including Member States representatives and industry representatives.
- **Policy Package 2** – High intensity regulatory options: Under this package, Member States will be required to set up a Single Data Authorisation body in charge of providing the authorisation to enable the further use of data that is subject to the rights of others contained held by the public sector. This entity will also be in charge of delivering the compulsory authorisation required from organisations offering data altruism schemes, as well as mandatory certification scheme for data intermediaries. Under this package, the European Data Innovation Board would consist of an independent European body with legal personality, supported by a secretariat.
- **Policy Package 3** – Preferred policy options: this package is similar to Policy Package 1, with the exception that a compulsory authorisation mechanism is set-up for organisations offering data altruism schemes.

Table 56 – Policy Packages composition

		Subtask 1.1	Subtask 1.2	Subtask 1.3	Subtask 1.4
Policy Package 1	Low regulatory options	PO2	PO2	PO2	PO2
Policy Package 2	High regulatory options	PO3	PO3	PO3	PO3

| Policy Package 3 | Preferred options | PO2 | PO3 | PO2 | PO2 |

6.2　Methodological approach

This section provides a brief explanation about the methodological approach for the macroeconomic analysis.

For the analysis of the economic impact both a top-down and a bottom-up analysis is conducted. The top-down approach is the primary method of analysis, whereas the bottom-up approach serves to validate the results. However, both approaches can be used to estimate a range of results and to calculate averages.

The top-down approach will be based on the estimation of a broader baseline for the size of the relevant data economy in a first step. The estimation of the baseline is described in chapter 6.2.1**Error! Reference source not found.**. In a second step, the potential is calculated by estimating a high growth scenario and calculating the difference to the baseline. The core of the top-down analysis is the estimation of the (positive) impact of each policy option per sub task in terms of contributing to reach the high-growth scenario. In this regard, ratios are estimated. The top-down approach is summarized in the figure below and described in more detail in chapter 6.2.2.

Figure 3 – Approach to the top-down analysis

| Determination of baseline | Potential to be addressed by PO | Impact of Policy Options |

First step

Estimate the value of the relevant economy (to be addressed) for the baseline. The **share in GDP/GVA** will be used as **main measurement indicator** in this regard.

Second step

The overall potential is defined as the **difference** between the **baseline** and the **high-growth forecast**.We will use this data as a baseline.

Third step

The **economic impact** of the different **policy options/packages (on GDP)** will be measured for each indicator on an incremental basis (vs. the baseline).

Based on assumptions for each PO and task, the potential gap in the data economy that could be addressed has been estimated.

The bottom-up approach, on the other hand, is based on the micro-analysis of estimated impacts conducted for each subtask. Within the CBA, certain benefits and costs are assessed. As far as possible, the impact on GDP is estimated based on the CBA results and/or case studies. The results and estimations of the micro-analyses are extrapolated and scaled in this regard. The bottom-up approach is described in more detail in chapter 6.2.3.

6.2.1　Calculation of the baseline

The baseline has been calculated based on the forecast of the European Data Market Monitoring Tool.[176] The monitoring tool provides three forecasts at the 2025 Horizon: a baseline scenario, a challenge scenario and a high growth scenario.

---

[176] Data landscape, The European Data Market Monitoring Tool see: http://datalandscape.eu/european-data-market-monitoring-tool-2018

Our study aims at understanding to what extent the introduction of a regulatory intervention will aim at reaching the expected levels of the high growth forecast from the European Data Market Monitoring Tool.[177]

In the year 2020, the outbreak of Covid-19 massively affected the European economy. Expected figures have been corrected to take into account the impact of this crisis. The European Data Market Monitoring Tool already provides a Covid19 correction of the 2025 forecast. We have made further annual adjustments according to The Economist intelligence Unit data forecast of GDP, which includes Covid19 corrections for 2020.

The impacts are calculated until 2025 on the basis of the value of the data economy as projected by the EU Data Monitoring Tool, which is the basis (or baseline) of our analysis. The EU Data Monitoring Tool forecast projects a growth of the data economy of approx. 8% p.a.. This forecast for the growth of the EU data economy, however, ends in 2025. In order to calculate impacts beyond 2025 we have taken a conservative approach and calculated the impacts on the basis of the GDP growth rate forecast of the OECD (1.5%-1.6% p.a.). For this reason the impacts are based on a much lower per annum growth rate for the period 2026-2028.

6.2.2 Top-Down analysis

In order to obtain the economic impact of data sharing, in relation to its contribution to GDP, a top-down analysis has been performed.

Figure 4 – Approach to the top-down analysis
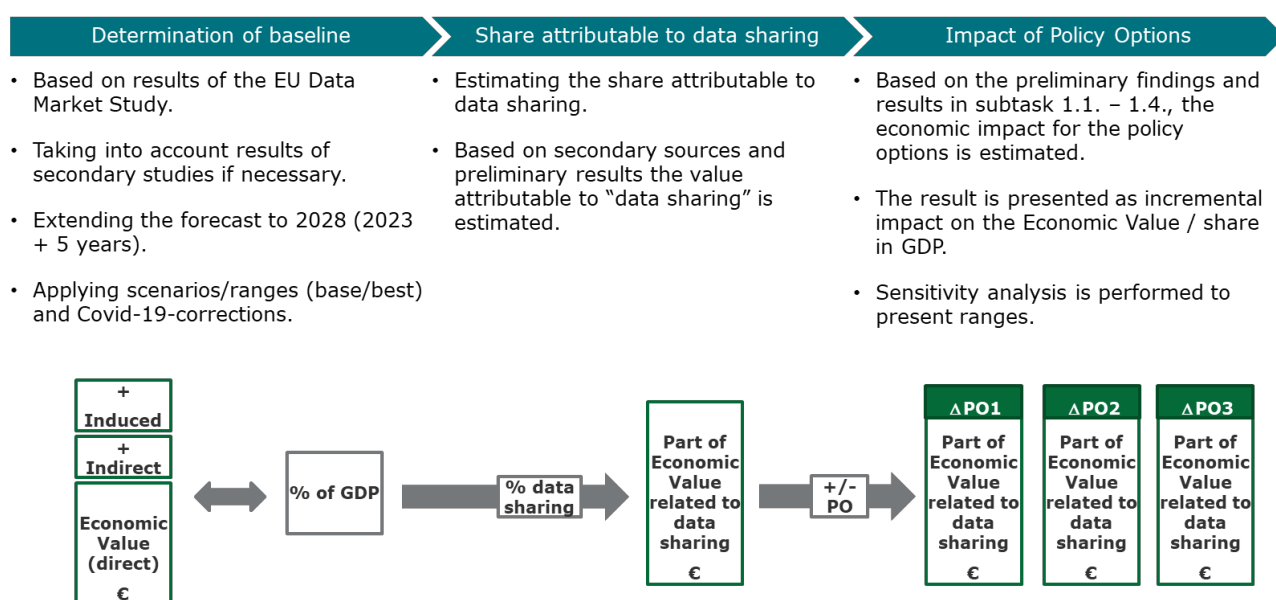
The data from the European Data Monitoring Tool provides a baseline for the economic value of the data economy and relates it to GDP. We have used this data to calculate the baseline. Adjustments with regard to Covid-19 outbreak macroeconomic impact have been included. The overall potential is defined as the difference between the baseline and the high-growth forecast.

---

[177] According to the European Data Market Study, "The High Growth scenario is characterised by a high level of data innovation, low data power concentration, an open and transparent data governance model with high data sharing, and a wide distribution of the benefits of data innovation in the society".

In order to define the relevant part of the value of the data economy that the policy intervention foreseen as part of this study can address, two conservative assumptions were made:

- The full potential of data sharing represent 80% of the total data economy;[178]
- 50% of barriers to data sharing are linked to interoperability and trust related issues, which are the specific problems the policy options under scrutiny in this study aim to address:
  - The open public consultation on the European Strategy for Data revealed that almost 80% of the participants of the public consultation have encountered difficulties in using data from other companies. These obstacles mainly relate to technical aspects (data interoperability and transfer mechanisms), denied data access, and prohibitive prices or other conditions considered unfair or prohibitive;[179]
  - A study from the World Economic Forum also estimates that the trust and technical issues are the most pregnant barriers to data sharing.[180]
  - At the same time, there is a general consensus (91% of the participants of the open public consultation) that standardisation is necessary to improve interoperability and ultimately data re-use across sectors.[181]

Based on these assumptions, the potential gap in the data economy that could be addressed if these problems were resolved has been estimated.

For each area, further assumptions have been made to understand in more details the magnitude of impact of the specific policy options on this potential gap. The experts' assumptions are based on the findings of the research, interviews carried out and the literature studies.[182] The baseline and the potential gap to be addressed are presented in the table below. The growth rates of the reference period 2026-2028 are based on the OECD long-term GDP forecast highlighted in dark green.

---

[178] Deloitte study for Vodafone group, Realising the economic potential of machine-generated, nonpersonal data in the EU, see: https://www.vodafone.com/content/dam/vodcom/files/public-policy/Realising_the_potential_of_IoT_data_report_for_Vodafone.pdf

[179] European Commission, 2020, Open public Consultation on the European Strategy for Data. Summary Report on the open public consultation on the European Strategy for Data.

[180] WEF, Share to Gain: Unlocking Data Value in Manufacturing, see : https://www.weforum.org/whitepapers/share-to-gain-unlocking-data-value-in-manufacturing

[181] European Commission, 2020, Open public consultation on the European Strategy for Data.

[182] Accordingly, the assumptions to a certain extent reflect the results of the bottom-up approach.

Figure 5 - Baseline estimates

Data sharing | Economic Impact

M€	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028
	forecast based on EU Data Monitoring Tool →							*OECD GDP forecast* →		
EU Data Monitoring Tool 2020 - baseline										
Data revenues	64 262	71 050	75 866	81 008	86 499	92 362	98 623	100 144	101 711	103 321
Data market value	58 214	62 244	65 795	69 584	73 628	77 948	82 564	83 837	85 149	86 497
Value of Data Economy										
Direct Impact	58 214	54 081	58 481	63 239	68 385	73 948	79 965	81 198	82 469	83 775
Indirect Backward Impact	3 197	3 105	3 324	3 559	3 811	4 081	4 369	4 436	4 506	4 577
Indirect Forward Impact	155 389	150 887	161 556	172 979	185 209	198 305	212 326	215 600	218 975	222 441
Induced Impact	108 058	98 853	115 213	134 280	156 502	182 402	212 589	215 867	219 246	222 717
Total Impact	**324 858**	**306 926**	**338 574**	**374 057**	**413 907**	**458 736**	**509 249**	**517 101**	**525 197**	**533 510**
EU Data Monitoring Tool 2020 - high growth										
Data revenues	64 262	71 050	80 943	92 215	105 055	119 684	136 350	138 453	140 620	142 846
Data market value	58 214	62 244	69 320	77 236	86 097	96 020	107 139	108 791	110 494	112 243
Value of Data Economy										
Direct Impact	58 214	54 081	62 005	71 090	81 505	93 447	107 139	108 791	110 494	112 243
Indirect Backward Impact	3 197	3 105	3 622	4 224	4 928	5 748	6 704	6 808	6 914	7 024
Indirect Forward Impact	155 389	150 887	176 002	205 296	239 467	279 324	325 817	330 840	336 020	341 339
Induced Impact	108 058	98 853	129 651	170 044	223 023	292 506	383 638	389 553	395 652	401 915
Total Impact	**324 858**	**306 926**	**371 279**	**450 655**	**548 922**	**671 026**	**823 298**	**835 992**	**849 081**	**862 521**
EU Data Monitoring Tool 2020 - potential										
Data revenues	-	-	5 078	11 207	18 556	27 322	37 727	38 309	38 909	39 525
Data market value	-	-	3 525	7 653	12 469	18 072	24 575	24 954	25 345	25 746
Value of Data Economy										
Direct Impact	-	-	3 524	7 850	13 121	19 499	27 174	27 593	28 025	28 468
Indirect Backward Impact	-	-	297	665	1 116	1 667	2 335	2 371	2 408	2 447
Indirect Forward Impact	-	-	14 446	32 317	54 257	81 020	113 491	115 241	117 045	118 898
Induced Impact	-	-	14 438	35 765	66 520	110 104	171 049	173 686	176 406	179 198
Total Impact	**-**	**-**	**32 705**	**76 598**	**135 015**	**212 290**	**314 049**	**318 891**	**323 884**	**329 011**
Data sharing [% of total Data Economy]		80.0%	80.0%	80.0%	80.0%	80.0%	80.0%	80.0%	80.0%	80.0%
- share linked to trust [% of total Data Economy]		50.0%	50.0%	50.0%	50.0%	50.0%	50.0%	50.0%	50.0%	50.0%
Data sharing [% linked to trust]		40.0%	40.0%	40.0%	40.0%	40.0%	40.0%	40.0%	40.0%	40.0%
Data sharing linked to trust - potential gap	-	-	13 082	30 639	54 006	84 916	125 620	127 556	129 553	131 604

6.2.3 Bottom-up analysis

A Bottom-up validation of these estimates has been performed based on the results of the Cost-Benefit analysis for each subtask.

Figure 6 – Results of the CBA per policy options and Policy Packages

Data sharing | Economic Impact

M€	2023	2024	2025	2026	2027	2028
Policy impact - bottom up (based on CBA result						
Policy Option 1 - direct						
1.1	-	-	-	-	-	-
1.2	-	-	-	-	-	-
1.3	(0.0)	1 050.0	1 050.0	1 050.0	1 050.0	1 050.0
1.4	(2.3)	23.6	4.8	4.8	4.8	4.2
Policy Option 2 - direct						
1.1	(286.3)	709.2	709.2	709.2	709.2	709.2
1.2	(3.8)	0.1	0.6	0.6	0.6	0.6
1.3	(0.3)	1 200.0	1 200.0	1 200.0	1 200.0	1 200.0
1.4	(5.3)	30.9	4.6	4.6	4.6	3.4
Policy Option 3 - direct						
1.1	(572.7)	1 090.8	1 090.8	1 090.8	1 090.8	1 090.8
1.2	(13.7)	43.7	48.7	53.7	58.4	63.3
1.3	(3.5)	1 350.0	1 350.0	1 350.0	1 350.0	1 350.0
1.4	(6.0)	30.6	(1.3)	(1.8)	(2.7)	(4.9)
Policy Package 1 (low intensity) - direct		1 940	1 914	1 914	1 914	1 913
Policy Package 2 (high intensity) - direct		2 515	2 488	2 493	2 497	2 499
Policy Package 3 (mixed option) - direct		1 984	1 963	1 968	1 972	1 976

In order to fully reflect on the reality of the impact, the indirect impacts have been added to the estimates based on the CBA results. A coefficient of 2.6 has been used, in line with the results of the European Data Monitoring Tool.[183]

Figure 7 – Overview of direct, indirect and induced impacts.

Value chain

Indirect Impact (backward)

-usually measured via Input-Output model

- Captures impacts on supplier industry

- Usually high for traditional industries (e.g. automotive)

- low for sectors at early stage of production

Direct Impact

-usually measured in terms of GDP

to consider:

- revenues not necessarily GDP

- societal impacts from CBA such as Value of Time or Life not GDP

Indirect Impact (forward)

- e.g. increased activities in downstream industries

- to be measured topic-specific

- avoid double counting: e.g. impacts on-data re-users probably already captured as direct impact

Induced Impact

- usually measured via Input-Output model

- captures the additional GDP induced by additional Consumer Spending: GDP of direct, indirect impact increases Wages, Salaries, Household Income (as part of GDP) → this increases consumer spending

6.3   Macroeconomic impacts of the policy packages

A **Single Market for Data with common data spaces** has the potential to bring **immense benefits** to the economy, notably through increased innovation and opportunities, as well as to society and the environment through new insights informing and improving decision-making.

As noted in section 3.1, the **objective of this intervention** is to set **the foundations** of a Single Market for Data by **enabling a range of actors** to make data available for access and secondary use, by ensuring greater interoperability across sectors, and by fostering trust. Therefore, this intervention is a **necessary first step in the process of creating these common data spaces**. The full range of benefits incurred by the latter heavily rely on **other actors seizing the opportunities** offered by these building blocks to complete these data spaces. Without this intervention, however, these actors would have no incentives *ceteris paribus* to change their behaviour.

For instance, the low-intensity **regulatory intervention under Task 1.1** – requiring Member States to set up one-stop shops to facilitate the secondary use of sensitive data held by the public sector – contributes to

---

[183] The European Data Monitoring implicitly includes several types of multipliers, including indirect and induced impacts, which estimate impacts on the supplier industries and the overall economy generated through additional income and consumption (both could be classically estimated using e.g. Input-Output models), as well as indirect forward impacts, which estimate the effects downstream in the economy. To stay conservative, the later one have been considered here based on the European Data Monitoring Tool, since those impacts are expected to be of major interest. The European Data Monitoring Tool in this regard estimates coefficients between 2.6 in the baseline as a lower bound and 3.0 in the high growth scenario as an upper bound.

fostering trust and to increasing the amount of data available for reuse. By this study's estimates, this low-intensity intervention would bring annual benefits of EUR 725 million for the EU27. Yet, it is by itself **insufficient to reap the full benefits of increased secondary use of such data**. It requires **subsequent action, particularly by public sector data holders** that will need to make their data available via this one-stop shop, **as well as by reusers** who will need to familiarise themselves and make use of this new service. This is a **long-term process**, in part because building trust is a lengthy process, and due to potential path dependence within public sector data holders or research organisations.

As regards **Task 1.2**, the **regulatory intervention with high intensity** – requiring Member States to establish a compulsory authorisation of data altruism mechanisms, administered by the one-stop shop established under Task 1.1 – could lead to the most beneficial outcome for Member States. It would increase the **trust** of data holders in certified data altruism mechanisms leading to an increase of shared data and thereby available data for data reusers. However, the **data altruism certification scheme** is merely the **first step** to this goal. **Member States** who have done so would **need to set up data altruism schemes**. Data reusers would still have to continue working on reaping all benefits such as analysing the data and utilising the data for e.g. new policy initiatives for the public good. In addition, data reusers will continuously have to work on **building trust** with data holders/subjects for these to share data for altruistic purposes. Only when data reusers can adequately present the benefits of data altruism to encourage data holders to share data, will this succeed in the long-term.

Likewise, as part of **Task 1.3**, the creation of a **formal expert group** with the low-intensity regulatory intervention would **reinforce trust** by facilitating the definition and adoption of standard data sharing schemes and reference architectures that include easy legal and trust arrangements for data sharing. It would also **facilitate interoperability**, by promoting the activities for the definition of data and metadata standards, and principles for interoperability between sectors. Yet, this will **merely facilitate the activities of the existing standardisation initiatives**, and is therefore only a **first step in achieving the objectives of this intervention**. Realising the full benefits of interoperability requires these existing initiatives to flourish, and private sector actors to increase data sharing and reuse.

The low-intensity regulatory intervention under **Task 1.4**, creating a **voluntary labelling/certification framework for data intermediaries** administered by the one-stop shop established under Task 1.1 is also expected to significantly **increase trust** between the stakeholders in the European data market. As most of these novel data intermediaries have recently made their appearance in the market, certification is expected to provide legitimacy to their operations, functionalities offered and business models, while it would also provide mechanisms for data holders and data reusers to assess the quality and neutrality of data intermediaries' services. However, since certification will be voluntary under this framework, the **positive impacts** of this regulatory intervention **depend also on the number of data intermediaries who will decide to proceed** to the certification and comply with the certification requirements.

This study, including the **cost-benefit analysis and the macroeconomic analysis focuses solely on the direct and indirect impacts of this initial first step** taken by the Commission. It does not assess the overall benefits that the EU's economy and societies would reap following the development of data spaces by other actors. Other studies estimate that in manufacturing alone, data sharing of IoT data is expected to generate - if fully implemented – 1.3 trillion euros in increased productivity by 2027. The policy options of this intervention are a necessary first step to encourage increased data sharing in the EU.

The Impact Assessment support study took as the baseline the total economic value of the data economy for the EU27 of 306.93 billion EUR in 2020 (2.7% of the GDP).[184] These numbers take into account a correction linked to Covid-19 impact on the overall EU economy.

The baseline scenario foresees an autonomous growth to 533.51 billion EUR (+74%) in 2028.

In 2028, the value of the data economy could increase from 533.51 billion EUR to between 540.5 billion EUR and 544.04 billion EUR if the lower intensity regulatory intervention was introduced (from 3.87% to between 3.92% and 3.94% of the GDP).

In 2028, the value of the data economy could increase to between 542.65 million EUR to 547.33 million EUR if the high intensity regulatory intervention was introduced (from 3.87% to between 3.93% and 3.97% of the GDP);

In 2028, the value of the data economy could increase from 540.73 billion EUR to 544.43 billion EUR if the mixed regulatory intervention was introduced (from 3.87% to between 3.92% and 3.95% of the GDP).

Figure 8 – Results of the top-down and bottom-up macroeconomic impact calculations

Data sharing | Economic Impact

M€	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028
	forecast based on EU Data Monitoring Tool							*OECD GDP forecast*		
Impact on the Economic Value of the Data Economy compared to GDP [m€]										
Baseline	324 858	306 926	338 574	374 057	413 907	458 736	509 249	517 101	525 197	533 510
% Baseline to GDP	*2.60%*	*2.66%*	*2.79%*	*3.00%*	*3.25%*	*3.54%*	*3.87%*	*3.87%*	*3.87%*	*3.87%*
Policy Package 1 (top-down)	324 858	306 926	338 574	374 057	413 907	465 529	519 299	527 305	535 561	544 039
% Policy Package 1 to GDP	*2.60%*	*2.66%*	*2.79%*	*3.00%*	*3.25%*	*3.59%*	*3.94%*	*3.94%*	*3.94%*	*3.94%*
Policy Package 2 (top-down)	324 858	306 926	338 574	374 057	413 907	467 652	522 439	530 494	538 800	547 329
% Policy Package 2 to GDP	*2.60%*	*2.66%*	*2.79%*	*3.00%*	*3.25%*	*3.60%*	*3.97%*	*3.97%*	*3.97%*	*3.97%*
Policy Package 3 (top-down)	324 858	306 926	338 574	374 057	413 907	465 784	519 675	527 688	535 950	544 433
% Policy Package 3 to GDP	*2.60%*	*2.66%*	*2.79%*	*3.00%*	*3.25%*	*3.59%*	*3.95%*	*3.95%*	*3.95%*	*3.95%*
Baseline	324 858	306 926	338 574	374 057	413 907	458 736	509 249	517 101	525 197	533 510
% Baseline to GDP	*2.60%*	*2.66%*	*2.79%*	*3.00%*	*3.25%*	*3.54%*	*3.87%*	*3.87%*	*3.87%*	*3.87%*
Policy Package 1 (bottom-up)	324 858	306 926	338 574	374 057	413 907	465 879	516 247	524 099	532 195	540 504
% Policy Package 1 to GDP	*2.60%*	*2.66%*	*2.79%*	*3.00%*	*3.25%*	*3.59%*	*3.92%*	*3.92%*	*3.92%*	*3.92%*
Policy Package 2 (bottom-up)	324 858	306 926	338 574	374 057	413 907	467 996	518 344	526 212	534 322	542 645
% Policy Package 2 to GDP	*2.60%*	*2.66%*	*2.79%*	*3.00%*	*3.25%*	*3.61%*	*3.94%*	*3.94%*	*3.93%*	*3.93%*
Policy Package 3 (bottom-up)	324 858	306 926	338 574	374 057	413 907	466 040	516 423	524 293	532 406	540 732
% Policy Package 2 to GDP	*2.60%*	*2.66%*	*2.79%*	*3.00%*	*3.25%*	*3.59%*	*3.92%*	*3.92%*	*3.92%*	*3.92%*

---

[184] It must be noted that the European Data Market Monitoring Tool uses the "Value of the Data Market" as a proxy for the direct economic value. The Value of the Market is calculated based on revenues of data companies, excluding exports and including imports. It should at least be noted, that imports usually do not contribute directly to GDP, those will affect foreign GDP (whereas exports contribute to domestic GDP).

Figure 9 – Impact of the Economic Value of the Data economy compared to GDP (top-down calculation)

Impact on the Economic Value of the Data Economy (top-down calculation)

Baseline ■ Policy Package 1 (top-down) ■ Policy Package 2 (top-down) ■ Policy Package 3 (top-down)
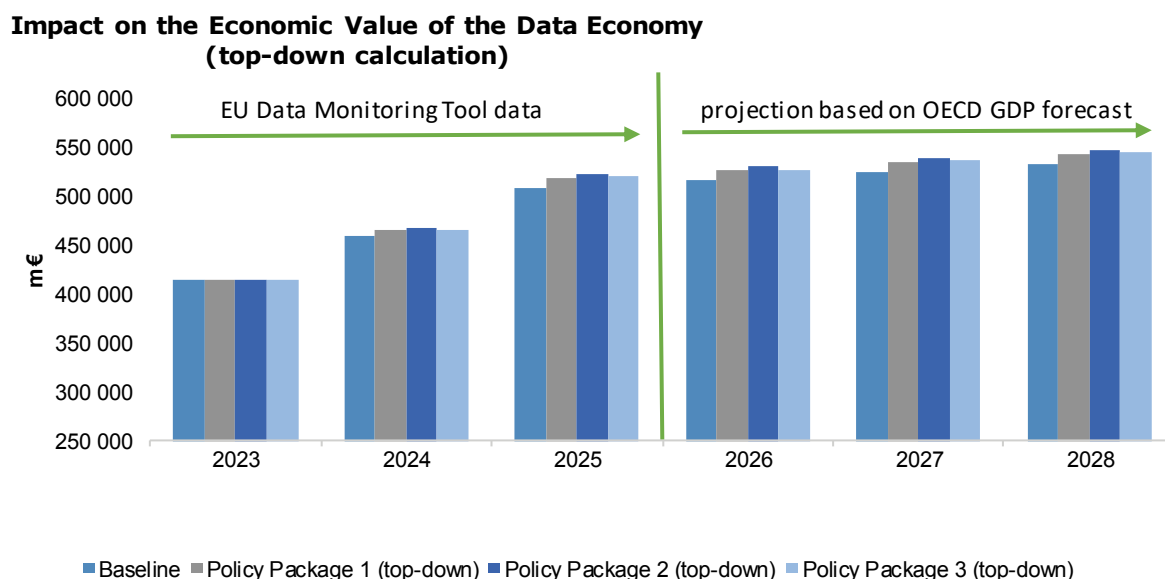
Figure 10 – Impact of the Economic Value of the Data economy compared to GDP (bottom-up calculation)
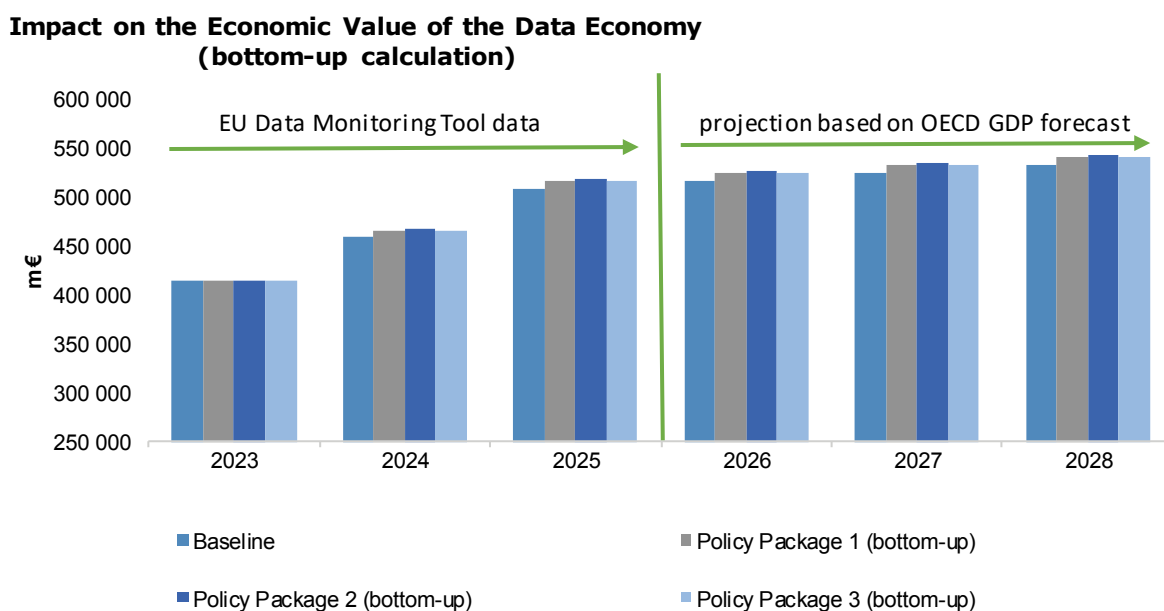
Impact on the Economic Value of the Data Economy (bottom-up calculation)

■ Baseline ■ Policy Package 1 (bottom-up)
■ Policy Package 2 (bottom-up) ■ Policy Package 3 (bottom-up)

The impacts are calculated until 2025 on the basis of the value of the data economy as projected by the EU Data Monitoring Tool, which is the basis (or baseline) of our analysis. The EU Data Monitoring Tool forecast projects a growth of the data economy of approx. 8% p.a.. This forecast for the growth of the EU data economy, however, ends in 2025. In order to calculate impacts beyond 2025 we have taken a conservative approach and calculated the impacts on the basis of the GDP growth rate forecast of the OECD (1.5%-1.6% p.a.). For this reason the impacts are based on a much lower per annum growth rate for the period 2026-2028. Due to a lack of specific growth rates for the data industry, the overall OECD GDP long-term forecast was chosen as a conservative alternative. Even though it could be expected, that growth rates for the data industry might exceed the general GDP growth, for the calculation of the impact in this analysis, the

incremental impact of each policy option/ policy package compared to the baseline is considered to be of most relevance, rather than the growth rate of the baseline itself.

The results regarding the economic value have been compared to the overall GDP for the EU27. For 2019, a ratio of 2.60 % compared to GDP has been estimated. This ratio was estimated to increase to 3.87% in 2028 in the baseline scenario. With regard to policy package 1 (low intensity intervention), the ratio was estimated to increase to between 3.94% (top-down calculation) and 3.92% (bottom-up calculation). For policy package 2 (high intensity intervention), an increase to 3.97% (top-down) and 3.93% (bottom-up) has been estimated. In the mixed intensity intervention (policy package 3), an increase to 3.95% (top-down) and 3.92% (bottom-up) was forecasted respectively. However, with regard to the ratios of the economic value compared to GDP, as presented in Figure 8, it should be noted, that the baseline for the years 2026-2028 has been forecasted based on growth rates of the long-term GDP forecast of the OECD.

In 2028, the total impact of the lower intensity regulatory intervention is estimated between billion EUR 6.993 and billion EUR 10.528 (bottom-up vs. top-down estimation).

In 2028, the total impact of the high intensity regulatory intervention is estimated between billion EUR 9.135 and billion EUR 13.818.

In 2028, the total impact of the preferred policy option package is estimated between billion EUR 7.222 and billion EUR 10.923.

Figure 11 - Total impact of Policy Packages (top-down calculation)
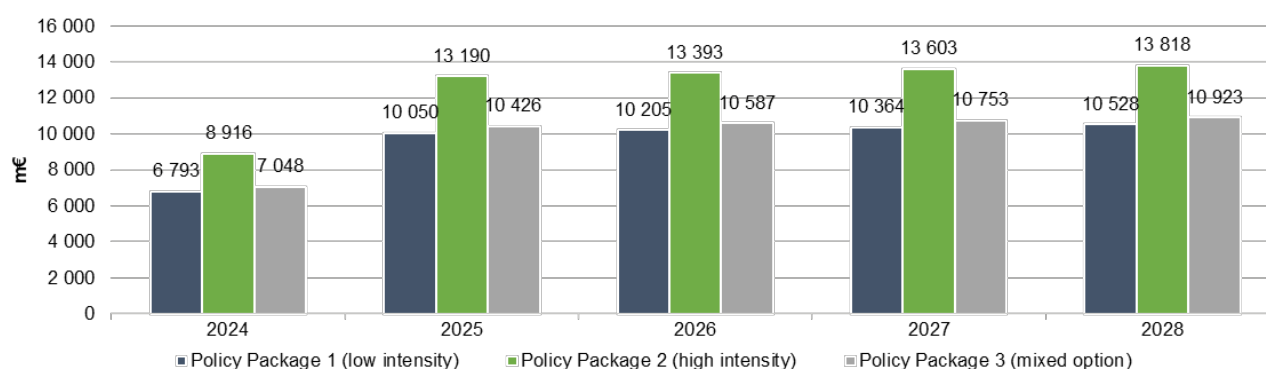

Figure 12 - Total impact of Policy Packages (bottom-up calculation)
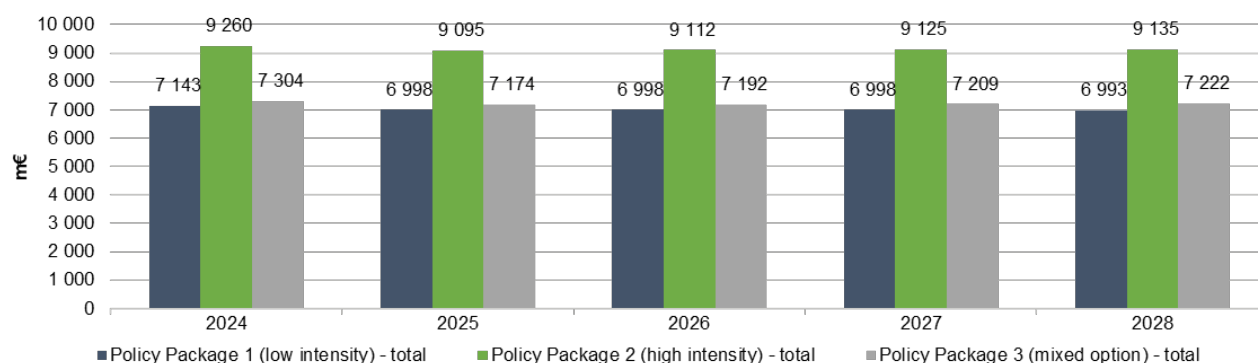
The following figures provide an overview of the share of contribution of the policy options to each Policy Option Package. The qualitative analysis and assumptions that were made for the top-down calculation takes into account total impacts according to the European Data Monitoring Tool. The weight of each policy option was assessed in detail to understand the magnitude of its contribution to the total effect. In the case of the bottom-up calculation, a multiplier was applied to the overall package in order to integrate the most relevant indirect impacts. As a consequence, the results of the CBA, on which the bottom-up approach is based, is equally amplified across all policy options. Taken this into account, we consider that in this case, the share of contribution provided by the top-down calculation is the most relevant.

For all Policy Packages, it is the setting up of a voluntary certification scheme for data intermediaries that realises the most benefits followed by the creation of a European Data Innovation Board aiming at improving coordination in the domain of data interoperability, standards and governance. This order of share of contribution is logical, as these policy options create cross-sectorial effects. The authorization scheme for data altruism and increased sharing of sensitive data held by the public sector are also expected to yield impacts, but at a lower scale considered the smaller scope of the domain at stake compared to the overall economy.

Figure 13 – Economic impact Package 1 by Policy Option
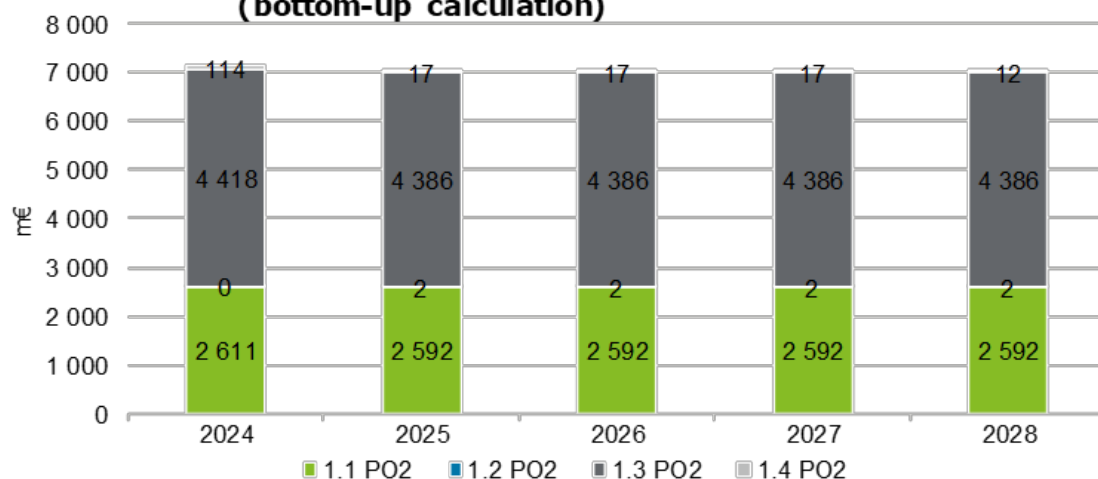
Figure 14 - Economic impact Package 2 by Policy Option

Economic Impact Policy Package 2 by Policy Option (bottom-up calculation)

Figure 15 – Economic impact Package 3 by Policy Option

Economic Impact Policy Package 3 by Policy Option (top-down calculation)

Economic Impact Policy Package 3 by Policy Option (bottom-up calculation)
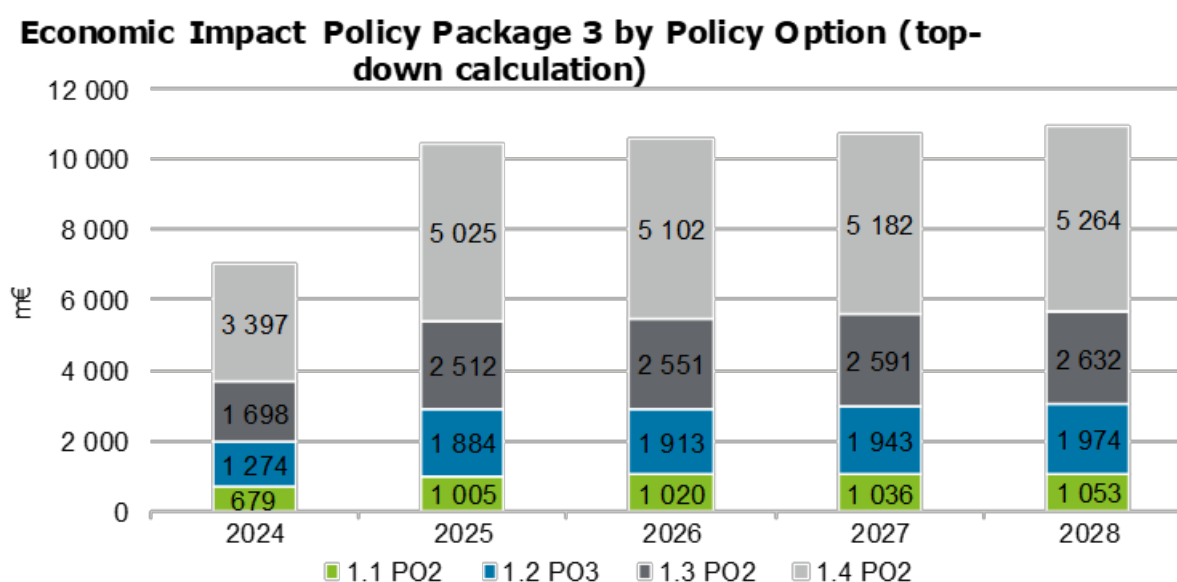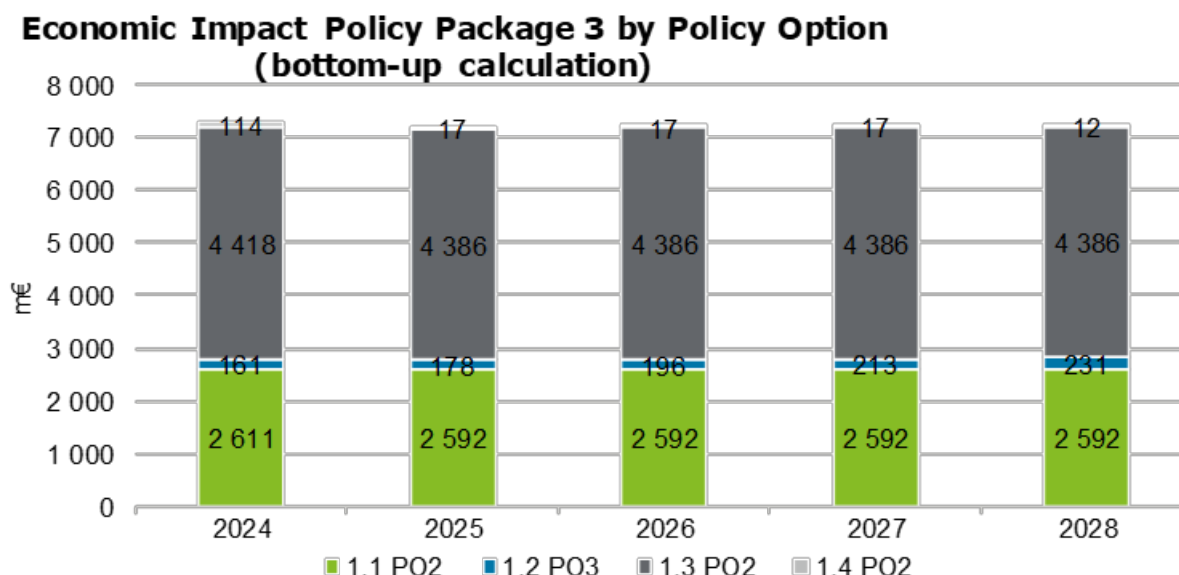
Our assessment is that all Policy Packages (low/high intensity regulatory intervention and preferred policy option packages) are creating positive impacts at the macroeconomic level, by boosting the value of the total economic value of the data economy for the EU27 from a projected autonomous growth from million EUR 306 930 in 2020 (2.7% of the GDP) to million EUR 533 510 (3.87% of the GDP) to between million EUR 540 504 and million EUR 547 329 in 2028 (3.92% to 3.97% of the GDP).

The Policy Package 2 (High intensity regulatory intervention) creates the highest impact on the total economic value of the data economy. This result is logical, as a more stringent intervention will affect more stakeholders (compared to voluntary approaches) and possibly more costs (that contribute to the European GDP at the macroeconomic level).

Policy Package 3 (Preferred Policy option package) creates important impacts on the total economic value of the data economy: more than the baseline and low intensity regulatory option but less than the high intensity intervention. It has been assessed as part of Chapter 5 that the combination of PO2 for subtask 1.1, PO3 for subtask 1.2, PO2 for subtask 1.3 and PO2 for subtask 1.4 hits the best score when all assessment criteria (effectiveness, efficiency, coherence, legal/political feasibility and proportionality) are taken into account.

6.4   Additional indicators

Based on the macroeconomic impacts we have estimated the impact of the policy options and policy packages on the following economic and socio-economic indicators:

- Employment (total number of additional persons employed, direct and indirect)
- Number of additional companies created statistically
- Additional governmental revenues (total gross as % of GDP incl. SSC, taxes, subsidies, governmental revenues etc.)
- Additional investment activity

To estimate the impact on these indicators, coefficients in terms of GDP-ratios have been used based on official data provided by Eurostat. With regard to numbers of person's employment, the number of additional companies and additional investment activities the GDP-ratios of the ICT-sector have been applied.

Governmental revenues were calculated based on the data on tax revenue and its relationship to gross domestic product (GDP) for the EU27 in general.

6.4.1 Employment

The first indicator, employment, indicates the total number of additional persons employed (directly and indirectly) in the case the respective Policy Package will be implemented. To calculate the total number of additionally employed people, the coefficient of employment as per mEUR gross value added (GVA) was determined. This coefficient was determined to be a weighted coefficient of the EU27 per mEUR GDP/GVA in the ICT services sector. Proceeding these calculations, a constant coefficient of 10.6 for the years 2024-2028 was applied.[185] The employment coefficient indicates the per-ratio increase in employment (number of persons employed) throughout the economy which result from an increase in GDP/GVA. On average (combining the bottom-up and the top-down approaches), for the low intensity Policy Package, an additional number of employed persons of 93 045 in 2028 is expected, for the high intensity Policy Package, an incremental of 121 890 persons and for the mixed Policy Package, an additional number of 96 357 persons is expected.

The following two figures provide a detailed overview of the employment impact incremental for the three Policy Packages, based on the top-down and a bottom-up calculation of the GDP impact.

Figure 16 – Employment impact incremental (top-down) in 2024-2028 for the different Policy Packages

---
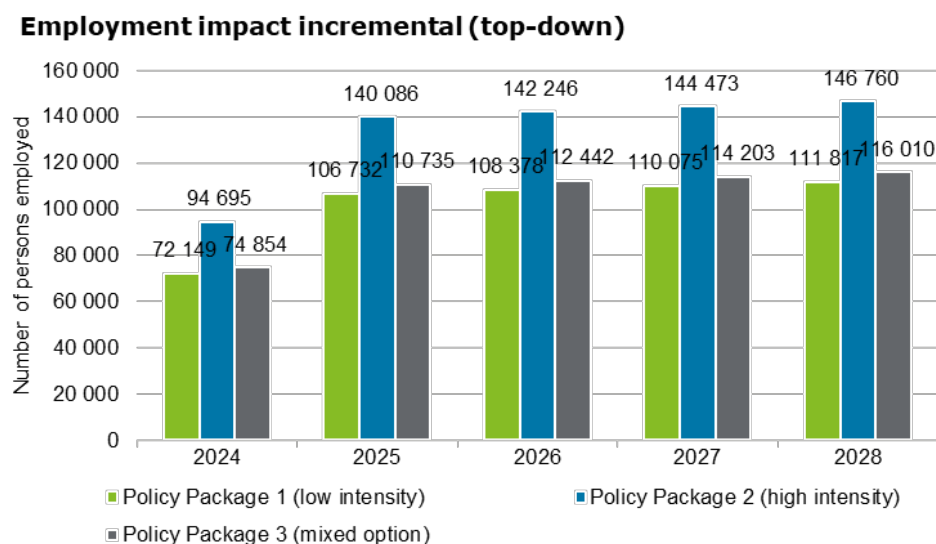
[185] The coefficient has been calculated as average of the years 2013 – 2017 for the total ICT-services sector in the EU27. With regard to the forecast period, the employment ratio should usually be adjusted, according to projected inflation. However, for the ICT industry in total, the HICP index has even been decreasing steadily in the recent years. Against this background we used a constant employment ratio for the forecast period.

Figure 17: Employment impact incremental (bottom-up) in 2024-2028 for the different Policy Packages
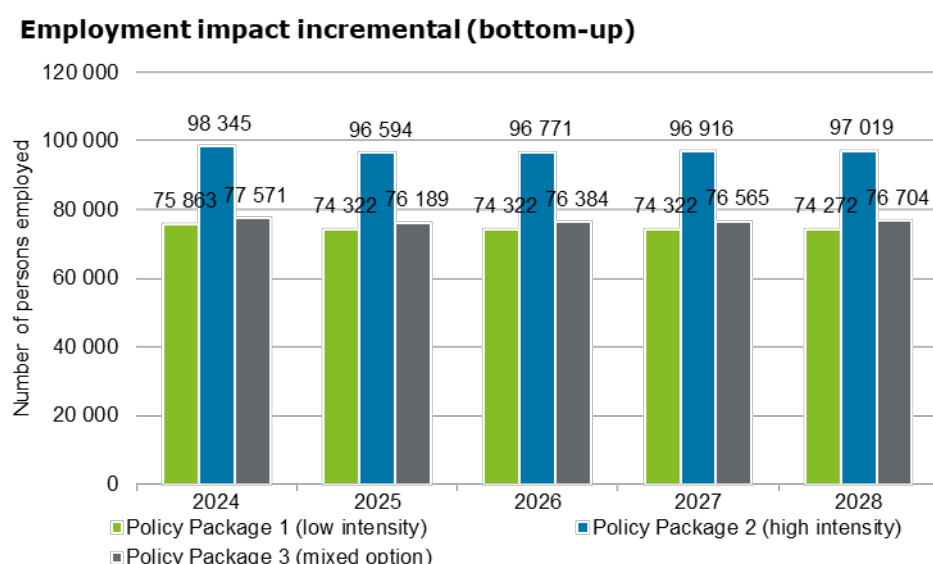
6.4.2 Number of additional enterprises

The second indicator to be included in the next stage of the impact assessment is the number of additional enterprises, which would be created statistically. This variable was calculated with a weighted coefficient for the EU27 ICT-service sector, representing the ratio of enterprises per GVA. For the weighted coefficient it was assumed that an average firm in the EU27 ICT-sector has 6 employees, respectively a statistical ratio of ca. 2 enterprises per 1 million € GVA.[186] Averaging between the top-down and the bottom-up approach, in 2028 18.585 new firms will exist with the low intensity Policy Package, 24 363 additional firms with the high intensity Policy Package and 19 260 with the mixed Policy Package.

However, the results should be interpreted as a proxy and represent a statistical value. It should also be noted, that a part of the impact refers to indirect impacts, which are linked to downstream activities in other industries. The economic impact can also occur in existing companies in the form of expanding their activities.

---

[186] The coefficient has been calculated as average of the years 2013 – 2017 for the total ICT-services sector in the EU27. With regard to the forecast period, the ratio should usually be adjusted, according to projected inflation. However, for the ICT industry in total, the HICP index has even been decreasing steadily in the recent years. Against this background we used a constant employment ratio for the forecast period.

Figure 18 - Number of incremental enterprises from 2024-2028 (top-down approach)
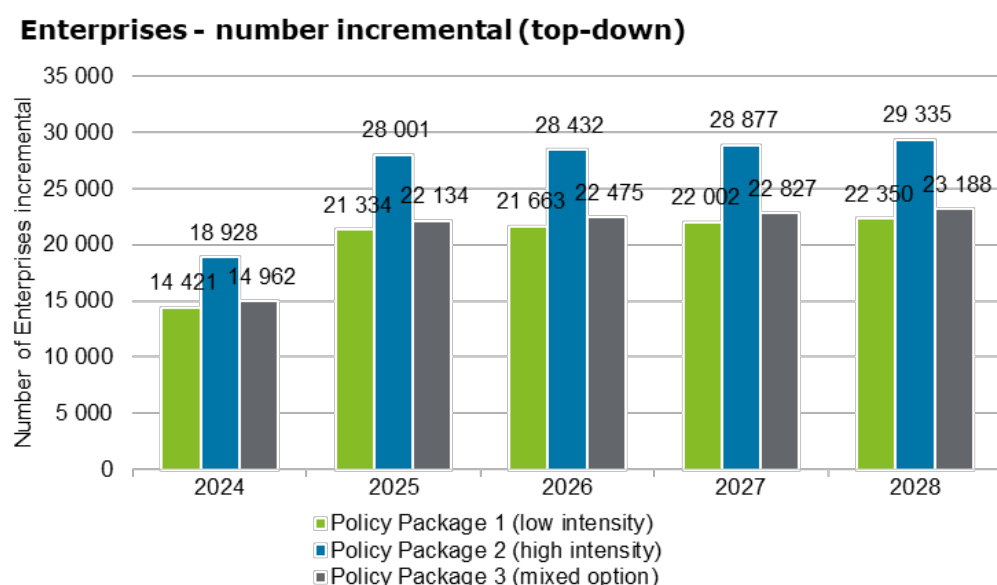
Figure 19 - Number of incremental enterprises from 2024-2028 (bottom-up approach)
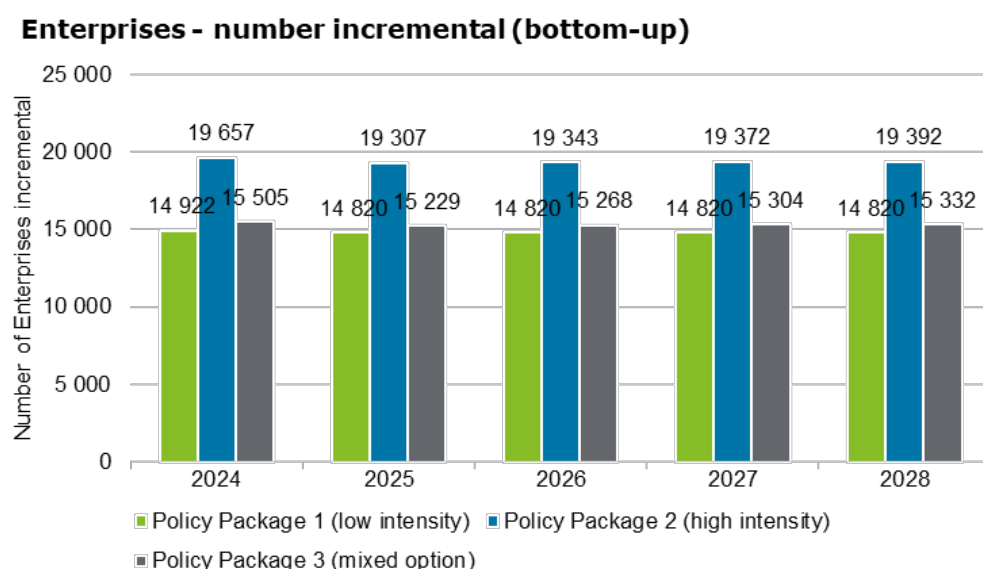
6.4.3 Governmental revenues

The third indicator to be included is the governmental revenues. According to the definition of Eurostat[187], the governmental revenue is the sum market output, of taxes, net social contributions, sales, other current revenues and capital transfer revenues. Total taxes are composed of taxes on production and imports, current taxes on income and wealth and capital taxes. The net social contribution is composed of actual social contributions by employers and households and the imputed social contributions, households' social contribution supplements and social insurance scheme service charges. Other current revenues consist of the categories property income earned, other subsidies on production received and current transfers. Combining these categories of governmental revenue, a weighted coefficient of EU27 by GDP is obtained.

---

[187] Eurostat 2020, Statistics Explained, Glossary: government revenue and expenditure.
https://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:Government_revenue_and_expenditure

Following the calculations of Eurostat, this coefficient has the value of 46% of GDP for the EU27. It should be noted, that part of this is related to governmental output, including market output, output for own final use and payments for non-market output, which could be linked to increased economic activity, but does not represent governmental inflows from taxes, social security payments or similar revenues.

For the governmental revenues, the top-down and the bottom-up calculations for the years 2024-2028 were averaged. This yields to an average governmental revenue of 4 030 mEUR in 2028 with the Policy Package 1. The highest governmental revenue on average can be obtained with an implementation of Policy Package 2, yielding to 5 279 mEUR in 2028. The mixed option, Policy Package 3, yields to an average amount of governmental revenue in 2028 of 4 173 mEUR.

Figure 20 – Governmental revenue from 2024-2028 (top-down approach)
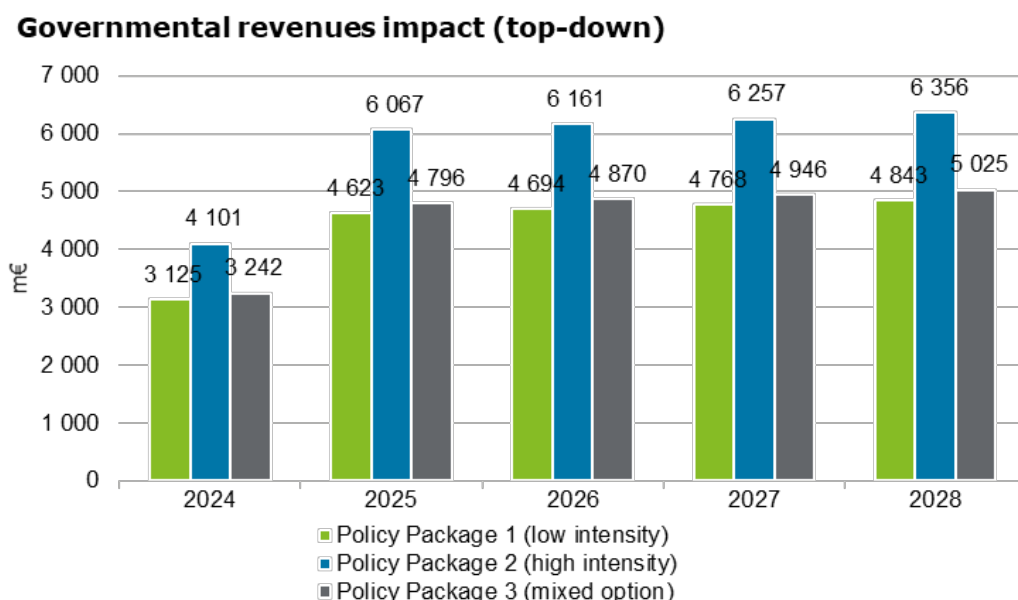
Figure 21 – Governmental revenue from 2024-2028 (bottom-up approach)
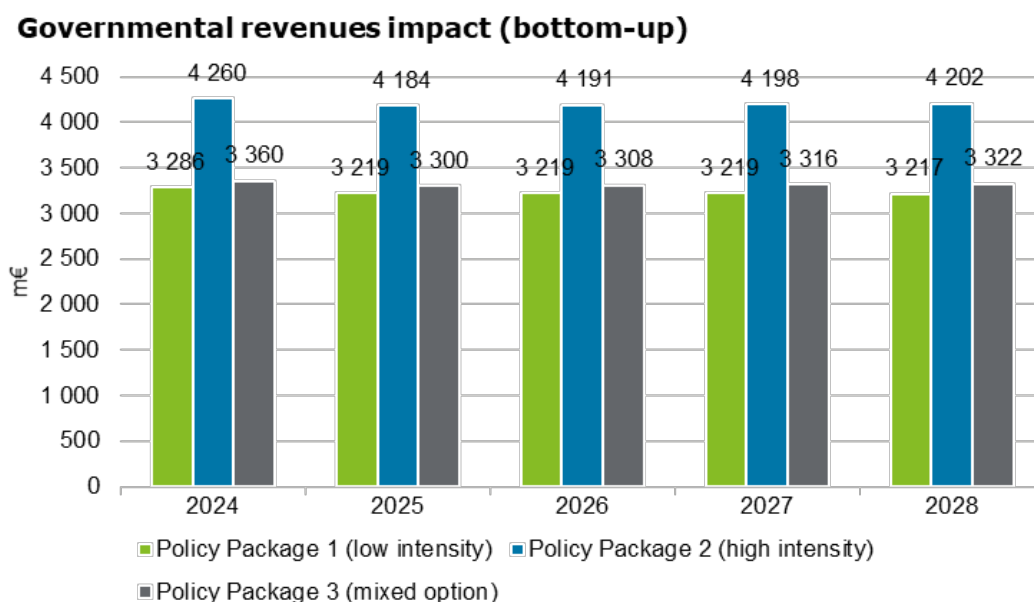
It must be noted, however, that this total governmental revenue includes – as defined in the European System of Accounts 2010 – also the market output, output for own final use and payments for non-market production. As this definition is a rather broad concept and as the macroeconomic effect of the introduction of the Policy Packages depends on a lot yet unknown factors, market output, output for own final use and payments for non-market production cannot be predicted as precisely as the other variables of governmental revenues. Excluding the categories mentioned, the adjusted governmental revenues would lower to approximately 38% of GDP according to OECD estimates.[188]

6.4.4 **Investment activity**

As a fourth indicator to be added we suggest to include investment activity. The investment rate is defined as the investment per value added at factor costs and is indicated as a percentage of the GDP of the EU27. The investment rate which was obtained by Eurostat[189] is at 14.4% of the GDP of the EU27 ICT-sector.

On average, an investment of 1 264 mEUR in 2028 for the EU27 can be expected with the Policy Package 1. The high intensity Policy Package 2 yields on average an investment of 1 657 mEUR in 2028, whereas with the mixed Policy Package 3 an investment of 1 310 mEUR in 2028 is expected.

Figure 22 - Investment activities (top-down) for 2024-2028

Investment activities - incremental (top-down)

---
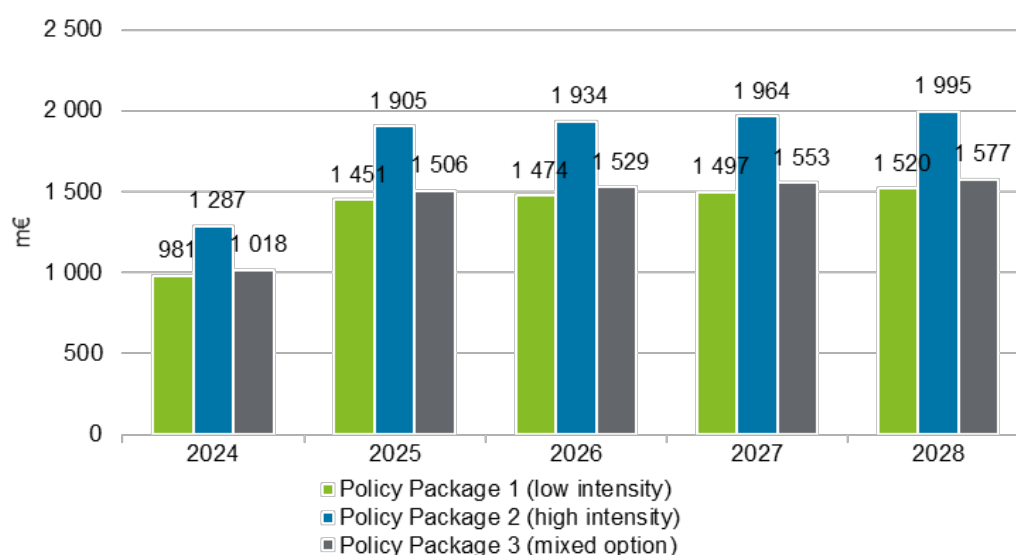
[188] OECD, 2020, Comparative Statistics: Governmental Revenue. https://stats.oecd.org/Index.aspx?DataSetCode=REV
[189] Eurostat, 2020, Investment share of GDP. See: https://ec.europa.eu/eurostat/web/products-datasets/product?code=sdg_08_11

Figure 23 - Investment activities (bottom-up) for 2024-2028

Investment activities - incremental (bottom-up)
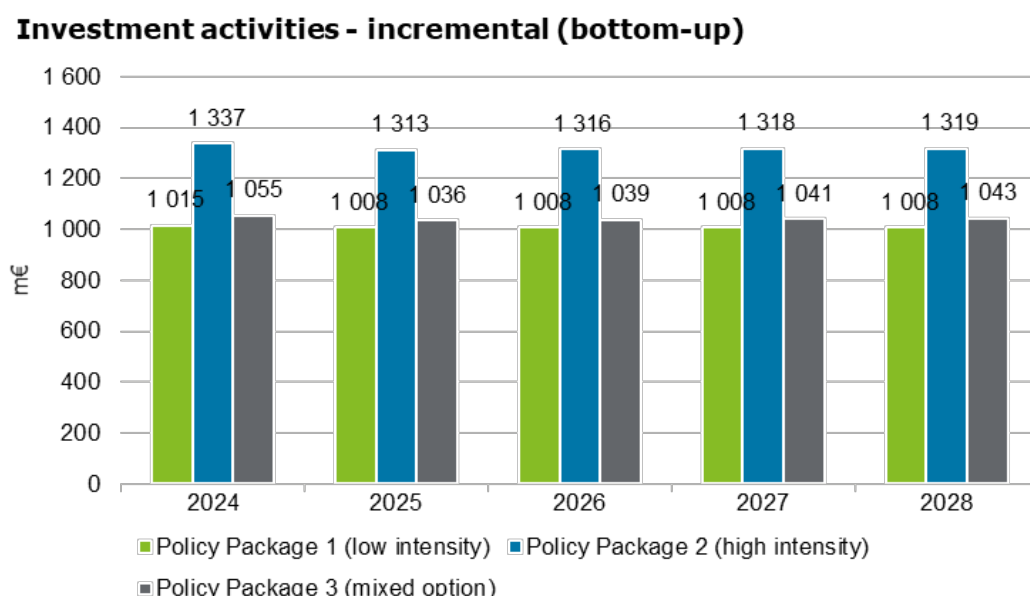
6.5   Impacts linked to specific elements of preferred policy package

6.5.1   Impacts linked to the additional sharing of sensitive data by the public sector

When looking at **health data** specifically, the **potential economic benefits of increased data re-use may be very large**. A 2019 Ernst & Young report estimates that the **UK's 55 million patient records may have a value of 'several billion pounds to a commercial organisation'**. Through increased efficiency, enhanced patient outcomes and wider economic benefits (e.g. big data, AI and personalised medicine), the curated **NHS dataset could deliver benefits worth as much as GBP 9.6 billion** (~EUR 10.7 Billion).[190] Likewise, a 2013 McKinsey study estimated that **increased re-use of health data** by both the public and private sectors would lead to **12% to 17% reduction in healthcare spending** in the United States, representing between **USD 300 billion and USD 450 billion in savings** (~EUR 266 to 399 billion).[191]

---

[190] EY, How we can place a value on health care data. See: https://www.ey.com/en_gl/life-sciences/how-we-can-place-a-value-on-health-care-data

[191] McKinsey & Company, The big-data revolution in US health care: Accelerating value and innovation. See: https://www.mckinsey.com/~/media/mckinsey/industries/healthcare%20systems%20and%20services/our%20insights/the%20big%20data%20revolution%20in%20us%20health%20care/the_big_data_revolution_in_healthcare.pdf

Currently, the extent to which the preferred Policy Package that includes establishing one-stop shops and secure data processing environments would contribute to achieving these economic benefits is unknown, since many **other factors have an impact** on these. For instance, the NHS' single medical market resulting in a large pool of unified data contributes to the potential value of the patient records datasets.

Nevertheless, the establishment of one-stop shops and secure data processing environments, by facilitating the reuse of health-related sensitive data and by improving interoperability between datasets, would make it easier for smaller research organisations, as well as for foreign researchers, to reuse sensitive data. It could also facilitate research re-using sensitive data from more than one Member State – and thus reusing larger datasets leading to potentially **better research outcomes and new insights**. This would contribute to **unlocking part of the value identified above**.

In addition to financial benefits, one-stop shops facilitating the reuse of sensitive health data may result in studies with a potential to **improve the lives of EU citizens**. For instance, the CASD in France facilitated the reuse by a private company of datasets related to home hospitalisation, follow-up/readaptation care, and medicine, surgery, obstetrics and odontology. This reuse resulted in the publication of several publications, including one on post-stroke spasticity and BoNT treatment in French hospitals, with a potential to improve the lives of post-stroke patients by reducing the occurrence of spasticity following strokes.[192]

Other studies resulting from reuse of data via the CASD have concerned wage inequality in different types of private companies,[193] impacts of territorial policies in France,[194] or productivity gains arising from agglomeration economies in Greater Paris.[195] While specific effects from these studies cannot be precisely estimated, they are nonetheless likely to have an **indirect societal impact** by improving awareness of the issues they tackle, and by informing policy-making.

6.5.2 Impacts linked to the establishment of a certification/authorisation scheme for data altruism mechanisms

To date, the expected impact linked to the establishment of certification/authorisation scheme for data altruism mechanisms is expected to be largest for healthcare related mechanisms. This is because data holders, citizens and companies, appear to be the most willing to share data for the public good when the direct impact, better healthcare or mitigation of a health crisis, can be achieved by sharing data. Again, while this argument has been repeated multiple times throughout this study, the COVID-19 pandemic accelerated the impact data altruism can have on a society.

To achieve this, PO3 is expected to be the most effective and impactful because data holder trust is the most important to achieve a high volume of data altruism to achieve a positive outcome for society and on the economy.

Considering that the European Union currently has 446 million inhabitants the future potential of data altruism, at least considering this group of data holders, is very large and a large resource for the public good. The most important factor however is that Member States coordinate their efforts on data altruism and that citizen trust and awareness of the benefits of data altruism are continuously increased. Scientists and governments, such as the German government, are already in favour of data altruism and the benefits

---

[192] Value in Health Journal, Patient Care Pathway for Post-Stroke Spasticity and Bont Management in French Hospitals Through the Prism of PMSI Data. See: https://www.valueinhealthjournal.com/article/S1098-3015(17)31733-3/fulltext
[193] CASD, Qualité de l'emploi dans les coopératives de travailleurs. See: https://www.casd.eu/project/qualite-de-lemploi-dans-les-cooperatives-de-travailleurs/
[194] CASD, Effets des dispositifs ZUS, ZRU, ZFU. See: https://www.casd.eu/project/effets-des-dispositifs-zus-zru-zfu/
[195] CASD, Impact des économies d'agglomération sur la productivité. See: https://www.casd.eu/project/impact-des-economies-dagglomeration-sur-la-productivite/

it can have for greater society, now data holders need to be educated and empowered, while sufficient mechanisms are created to enable data altruism.

6.5.3 Impacts linked to further governance of data and data standards

The economic benefits of greater adoption of data and metadata standards and schemes by companies are very significant: just for manufacturing, 1,4 trillion Euros in benefits of data sharing are estimated by 2027, and the few available studies convene that the main obstacles lie in lack of standards and trusted legal models for data sharing – hence we attribute a conservative estimate of 50% of the gains to the solution of these barriers.[196] There are many initiatives already today in place, but have reached limited traction so far, and there is certainly a need for increased European activity to foster the development and adoption of such initiatives. It is estimated that effective adoption of such instruments could lead to a reduction of 15% in operational expenditure.

Obviously, most of these benefits depend on decision by companies to adopt and comply with such standard, which lies entirely upon their business decision. Any European intervention, while necessary, will have only a very indirect effect on the ultimate adoption of such instruments by companies.

On the other hand, the economic benefits are only one side of the coin. Ultimately, the massive economic benefits deriving from increased efficiency will also be reflected in environmental benefits, namely through increased energy efficiency. These benefits are massive in size: the industrial sector consumes about 54% of the world's total energy according to the International Energy Outlook 2016. Existing cases show a 15% energy reduction thanks to improved IoT based controls.[197] Last but not least, data sharing schemes are crucial for the efficiency of the energy sector itself.[198]

On the other hand, because the main application domain of such instruments is industry, it is difficult to quantify any form of societal impact. Of course, data sharing in domains such as pharmaceutical is already seen as key to develop new drugs, ultimately leading to gains in health,[199] but these benefits are very indirectly related to the measures under discussion.

6.5.4 Impacts linked to projects that would have benefited from the certification of intermediaries

Currently, several data intermediaries in Europe have launched initiatives that encourage and facilitate both B2B and C2B data sharing, aiming to tackle the COVID19 global health crisis and restore the economy faster through data sharing. Non-exhaustive examples of such companies include Digi.me and CitizenMe in the United Kingdom, MIDATA in Switzerland, Dawex in France, de Volksbank in the Netherlands, Polypoly in Germany and many others. In particular, a "COVID19 Hub" has been created in the Digi.me application which counts approximately 700,000 users over time across 140 countries, enabling personal data sharing in order to flatten the COVID curve faster and help restore economy for business and citizens.[200] CitizenMe has launched a research project enabling people to share information in order to tackle COVID19. CitizenMe

---

[196] As previously illustrated, the Everis study on data sharing places technical interoperability as the most mentioned obstacle, by 73% of companies. Legal uncertainty about data ownership is the second, with 54%, and control over usage the third with 42%. The Deloitte Vodafone study reports costs of normalizing data, lack of standard protocols, cumbersome legal procedures, involuntary disclosure of commercial secrets as the main barriers. The WEF "Share to gain" report identifies standars, trust and legal arrangement as the key enablers.

[197] See Us Energy Information Administration, 2016, International Energy Outlook 2016 and https://www.emersontopquartile.com/z-featureditems/featured-2/industrial-internet-of-things-empowering-big-time-energy-savings

[198] Douwe Lycklama et al, Data sharing: a new source for the Energy Transition in Smart Energy International 5-2019

[199] Mugdha Khaladkar and others, 'Uncovering Novel Repositioning Opportunities Using the Open Targets Platform', Drug Discovery Today, 22.12 (2017), 1800–1807 <https://doi.org/10.1016/j.drudis.2017.09.007>.

[200] https://digi.me/covid19/

platform has 250,000 users worldwide who use the app to share data, information and answers to questions anonymously. The results are shared openly with institutions, health organisations, researchers, journalists, charities, and the general public.[201] MIDATA's Corona Science project aims to make available to the public, as quickly as possible, a collection of anonymized/aggregated health and symptom data (stored in the MIDATA platform) in a semantic standard defined with eHealth Suisse as Open Data.[202] Finally, Polypoly's GDPR compliant Corona Protector for corporates is helping them to manage the crisis, without harming the privacy of the employees, while also enabling the trade unions to monitor the data behavior of the employer.

Given that there are currently approximately 150 data intermediaries in the European Market, with thousands of users each of them, it is estimated that the increased trust between the market stakeholders, after certification of data intermediaries, and the resulting increase in the volume of data sharing could lead to the acceleration of the time needed for resolving a global health crisis and restore the economy of approximately 25% under PO1 (industry driven self-certification scheme); 40% under PO2 (voluntary certification scheme) and 45% under PO3 (compulsory certification scheme), as significantly more stakeholders would be eager to share and use data, through data sharing platforms. This assumption is made, based on the on expected benefits by the interviewed stakeholders to arise after certification (including business development time acceleration, client base and revenue increase), described in detail under chapter 4.5.

---

[201] https://covid19.citizenme.com/public/wp/ and https://www.citizenme.com/public/wp/covid19/covid19/
[202] https://coronascience.ch/en/

7 Conclusion

This section concludes on this second interim report.

This study sustains the development of policy measures concerning the areas defined by the Communication on a European Strategy for data[203], which was published in February and constitutes the mandate for the new Commission. The study focused in particular on eight key issues which are outlined in the strategy, divided in two tasks. This current D2 report focused on topics under **Task 1**, namely:

- The question of access and **reuse of sensitive public sector data** which are currently not disclosed by public sector bodies and not covered by the Public Sector Information (PSI)/Open data directive[204] (e.g. health data, statistical microdata, company ownership data, microdata from public transport systems and others).[205]
- The possibility of **establishing "data altruism" schemes** in Europe, defined as means of making data available (whether anonymised or non-anonymised) without expecting anything (not even services) in exchange.
- The question of facilitating data sharing through the establishment of **metadata standards** across or within sectors and including both technical and legal standards.
- The relevance of building a **certification framework** for European data intermediaries or data marketplaces which help data demand and supply to match through independent platforms.

For each of these key aspects, the study explored the state of play in Europe and determines the impact of a number of possible policy options.

Concerning the geographical scope, the study focused on the **27 European Union Member States** but it also covered case studies, examples and literature coming from third countries when relevant (i.e. experiences of B2G data sharing). Furthermore, for specific subtasks (i.e. Subtask 1.2 on data altruism) the data collection and analysis activities focused on a sample of Member States. From a stakeholder perspective, the study focuses on the relevant stakeholders in the data value chain for each of the topics in scope, meaning on data holders, data intermediaries and data re-users.

This study collected data from a range of sources, including desk research, stakeholder interviews, workshops and case studies. The data collection was hampered by the fact that the public and private sector are still relatively new to navigating the data economy and could only share insights regarding costs and benefits to a very limited extent. While this study was able to collect qualitative feedback from the public and private sector on the different policy interventions discussed for each subtask, it was more difficult to quantify their costs and benefits, e.g. because case numbers are still small or the data sharing practices are just emerging and stakeholders themselves do not yet know their scale and/or costs of making data available. In addition, the stakeholders consulted do not yet have a final and consolidated perception on for example the potential benefits they could draw from increased data use and availability in their respective domain, besides speculative thoughts. This report should be considered as a first attempt at examining this topic and gathering the existing data on these subjects. This analysis is therefore based on the limited data

---

[203] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, A European strategy for data, COM/2020/66 final, https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1582551099377&uri=CELEX:52020DC0066

[204] Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information, https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1561563110433&uri=CELEX:32019L1024

[205] In agreement with the Commission, this study focuses on the former two.

available and provides a preliminary (mainly qualitative) overview of the costs and benefits for the different topics under scrutiny. The conclusions reached are based on independent judgement and specific to this study.

The assessment of the policy options for each Task enabled the study to formulate several policy packages combining one policy option per subtask. These policy packages are:

- **Policy Package 0 – Baseline**: the baseline scenario consists in applying no policy changes to the four areas for which problems could be identified: sensitive data held by the public sector, data altruism schemes, certification of data intermediaries, governance and standards. The EU economy will not be able to reap the benefits of data sharing.
- **Policy Package 1** – Low intensity regulatory options: this package includes the creation of a one-stop shop to foster the sharing of (sensitive) data whose use is subject to the rights of others held by the public sector. A voluntary certification scheme would be established by EU Member States for data altruism mechanism and organisations offering such schemes. Data intermediaries will also be able to obtain a certification to demonstrate their neutrality and absence of conflict of interest (e.g. absence of competition with data users) on a voluntary basis. Finally, the European Data innovation Board would take the form of a formal expert group created by the European Commission, including Member States representatives and industry representatives.
- **Policy Package 2** – High intensity regulatory options: Under this package, Member States will be required to set up a Single Data Authorisation body in charge of providing the authorisation to enable the further use of data that is subject to the rights of others contained held by the public sector. This entity will also be in charge of delivering the compulsory authorisation required from organisations offering data altruism schemes, as well as mandatory certification scheme for data intermediary. Under this package, the European Data Innovation Board would consist of an independent European body with legal personality, supported by a secretariat.
- **Policy Package 3** – Preferred policy options: this package is similar to Policy Package 1, to the exception that a compulsory authorisation is set-up for organisations offering data altruism schemes.

In a last step, this study conducted a macroeconomic analysis of these packages. To do so, the team has calculated the baseline scenario using the forecasts of the European Data Market Monitoring Tool, corrected for the impact of the Covid-19 crisis. The study team then performed a top-down analysis of the policy packages, as well as a bottom-up analysis based on the cost-benefit results of the policy options. It found that by 2028, the value of the data economy could increase from EUR 533.51 billion:

- To EUR 540.5 billion – 544.04 billion with the lower intensity regulatory intervention;
- To EUR 542.65 billion – 547.33 billion with the higher intensity regulatory intervention; and
- To EUR 540.73 billion – 544.43 billion with the mixed regulatory intervention.

Yet, while Policy package 2 yields the highest impact on the total economic value of the data economy, Policy package 3 ranks highest when including other assessment criteria such as particularly coherence, and legal/political feasibility and proportionality.

Annex A – Subtask 1.1 Case Studies

This Annex contains the case studies carried out by the study team for Subtask 1.1, namely Findata, RatSWD, the Scottish national Safe Haven, Statistics Denmark, OpenSAFELY, and Statbel.

Findata

Introduction

Finland has a long history of collecting extensive data in registers but making use of the data has been difficult and inefficient. In 2019 a new **Act on Secondary Use of Health and Social Data** entered in force in Finland. With the new enabling legislation, Finland has become the first country in the world to successfully enact a law on the secondary use of well-being data that meets the requirements of the European General Data Protection Regulation (GDPR).

The new legislation **enables and expands the use of social and healthcare data from the traditional areas of scientific research and statistics** to those of management/control of social welfare and healthcare, development and innovations, knowledge management, education, authorities' planning and forecasting tasks, and steering and supervision of work. The new Act facilitates the establishment of a **new central data permit authority** in Finland, Findata.

The objectives of establishing a new centralised body (Findata as Data Permit Authority) devoted to the implementation of the secondary use of health and social data have been mainly:

- **To enable efficient and secure processing of personal data collected during the provision of social and health care** as well as personal data collected for the purpose of steering, supervision, researching and collecting statistics on the social and health care sector, in full compliance with GDPR prescriptions;
- **To allow the collected personal data to be combined** with the personal data held by Social Insurance Institution of Finland, Population Register Centre, Statistics Finland and Finnish Centre for Pensions;
- **To secure the legitimate expectations, rights and freedoms of individuals** when processing personal data.

Findata is the **one-stop-shop** responsible for streamlining and securing the secondary use of social and health data. It guarantees a flourishing ecosystem around the secondary use of social and health data streamlining the processes for the **issuing of research permits** and data collection and ensuring that data is being used in **secure environments**, thereby maintaining the trust that the general public have in authorities and the public sector.

This case study details how Findata operates in practice, using the GOFA model.

Governance

Findata operates directly under the **Ministry of Social Affairs and Health** and is a separate legal entity functioning as part of the **National Institute of Health and Welfare** (THL). Findata's operations are supervised by the **Parliamentary Ombudsman and the Data Protection Ombudsman**, among others.

The **National Supervisory Authority for Welfare and Health Valvira** monitors Findata's data secure user environments. In addition, Findata must give an annual report to the Data Protection Ombudsman regarding the processing of health and social data and the related log data.

To steer the operations of Findata and to develop the cooperation, the Ministry of Social Affairs and Health organises a **steering committee** every three years and elects a **chair person** for the commitee. The members of the steering group have been choosen from:

- the Ministry of Social Affairs and Health,
- the Finnish Institute for Health and Welfare,
- the Social Insurance Institution of Finland,
- the Finnish Centre for Pensions,
- the Populaton Register Centre,
- Statistics Finland,
- the Finnish Institute of Occupational Health,
- the Finnish Medicines Agency Fimea, and
- representatives of social welfare and health care service providers.

The **task of the steering committee** is to process and make proposals to the National Institute for Health and Welfare and the Ministry of Social Affairs and Health on:

- the annual action plan of the Data Permit Authority and the associated budget;
- the report on operations and financial statements as applicable to the Data Permit Authority;
- the joint development of controllers and the resources allocated to the task;
- the resources allocated for the development of information systems and cooperation;

Additionally, the steering committee is responsible to:

- set goal indicators for the processes of the Data Permit Authority and initiate external audits on the processes;
- if necessary, make a proposal to the National Institute for Health and Welfare and the Ministry of Social Affairs and Health on the improvement of the Data Permit Autority's operations;

The Ministry of Social Affairs and Health established a **high-level expert group** for Findata. The task of the group is to create guidelines on anonymisation, data protection and data security for the Data Permit Authority's operations. The expert group must have an expert on each of the following fields: artificial intelligence, data analytics, data security, data protection, suitable research, statistics and statistical service as well as a representive of the Data Permit Authority.

Operations

Findata grants permits to allow the secure and easy use of social and health data for the purposes laid down by the Act on the Secondary Use of Health and Social Data. As such, Findata is a one-stop-shop for data, centralising the decision procedures and access to the data.

Thanks to Findata, retrieving combined health and social data from different sources is easier, faster and possible with just one permit application, removing the need to approach each authority and data source separately. Previously, obtaining the permits and data has taken as long as up to two or three years. The Act guarantees the provision of a permit within just three months. For exceptionally complex data requests that can cover several data registries, the data permit authority can extend the time it takes to obtain a

permit by a maximum of three extra months. In addition, the data is provided with little delay, no later than within 60 business days after the permit has been approved.

Findata is responsible for ensuring the ethically sustainable use of data. It makes decisions on data permits concerning data held by other controllers, and is responsible for the collection, combination, pre-processing and disclosure of data for secondary use, in accordance with the Act. Furthermore, the data permit authority maintains a data request management system to forward and process data requests and permit applications. Findata also maintains a secure hosting service for receiving or disclosing personal data and a secure operating environment, in which the permit holder may process the personal data he/she has been disclosed on the basis of data permit. It also supervises compliance with the terms and conditions of the permit it has issued. The data permit may be revoked if the permit holder fails to comply with the law or the terms and conditions of the permit. Lastly, the data permit authority is responsible for the pseudonymisation and the anonymisation of personal data.

The secondary use of health and social data means that the data generated during health and social services are also used for other purposes, in addition to the primary purposes for which they were originally saved. Health and Social data were initially only used for traditional **scientific research and statistic in the health care domain**. Thanks to the Findata approach, it is possible now to activate new data usage, such as:

- **Development and innovations activities (R&D):** not only researchers but many diverse worldwide health technologies and life-science companies benefit from this new approach. Thanks to the access to social and healthcare data reserves, these companies can start to see opportunities in Finland and expand their R&D activities to the country. These activities however must be aimed at promoting national health or social security, at developing social welfare and healthcare services or service systems, or at protecting the health and well-being of individuals or securing for them the related rights and freedoms.
- **Knowledge management:** thanks to Findata, each organisation can improve its knowledge-management opportunities in social welfare and healthcare sectors with easier access to comprehensive data sources and new services around high-quality registered data.
- **Planning and forecasting of the activities and initiatives performed by social and health care Finnish authorities**: to transform the Finnish authorities in a data-driven organisations, the data collected from Findata can be used as a basis for the planning of central initiatives and programmes.
- **Governance and supervision of social and health care organisations:** the governance and control of organisations by social and health care **Finnish authorities** based on personal data and statistics and/ or on data received from case-studies, such as the National Institute for Health and Welfare or the Population Register centre.
- **Education:** higher education institutions, such as biomedical campus universities, can benefit from the data stored in Findata using data for the development of projects, publications, preparation of seminaries and other materials.

In 2019 a temporary steering group was put in place to prepare the launch of Findata operations. In the summer of 2019 the Findata director and staff were recruited. The Findata website has been opened since August 2019. From the 1 November 2019, the Findata help desk (website, e-mail and phone service) is available. Currently, it Findata is open to receive data requests for anonymised statistical data (since 1 January 2020). The system also collects data permit applications for individual-level data from 1 April 2020.

As regards future developments, Findata aims to guarantee a secure remote use environment for customers and improve data set descriptions, data management and methods.

In 2020 Findata counts 15 people. A first investment has been made to hire profiles with legal expertise and administrative skills. Additionally, several ICT profiles have been hired. The hiring process is still on-going, and in 2021, Findata expects to have 20 professionals. The hiring strategy aims to guarantee a good mix of skills that can enable the use of the new technology and methods, analytics skills and achieve a good understanding of research practices.

In addition to the citizens, the users who can access Findata also include:

- **Authorities:** among others Social Insurance Institution of Finland (KELA), Social Welfare Office and National Supervisory Authority for Welfare and Health – Valvira.
- **Institutes:** research institutes, universities and biomedical campuses.
- **Companies:** pharmaceutical companies, health technology and life-science companies.
- **Professionals:** Healthcare and social welfare professionals, professors and PhDs.

Findata is responsible for data permits and data requests when the data is combined from the following holders:

- Social and health care operating units;
- Finnish Institute for Health and Welfare (does not apply to data collected for statistical purposes);
- Social Insurance Institution of Finland Kela (benefits and prescriptions);
- Data saved in Kanta Services;
- Finnish Centre for Pensions (work and earnings data, benefits and the bases for them);
- National Supervisory Authority for Welfare and Health Valvira;
- Finnish Medicines Agency Fimea;
- Finnish Institute of Occupational Health (occupational illnesses, exposure tests);
- Regional state administrative agencies (matters related to social welfare and health care);
- Population Register Centre (individual's basic details, family relations, places of residence and building information);
- Statistics Finland (to the extent that access is required to data covered by the Act on Establishing the Cause of Death 459/1973).

Figure 24 - Data holders and related data[206]
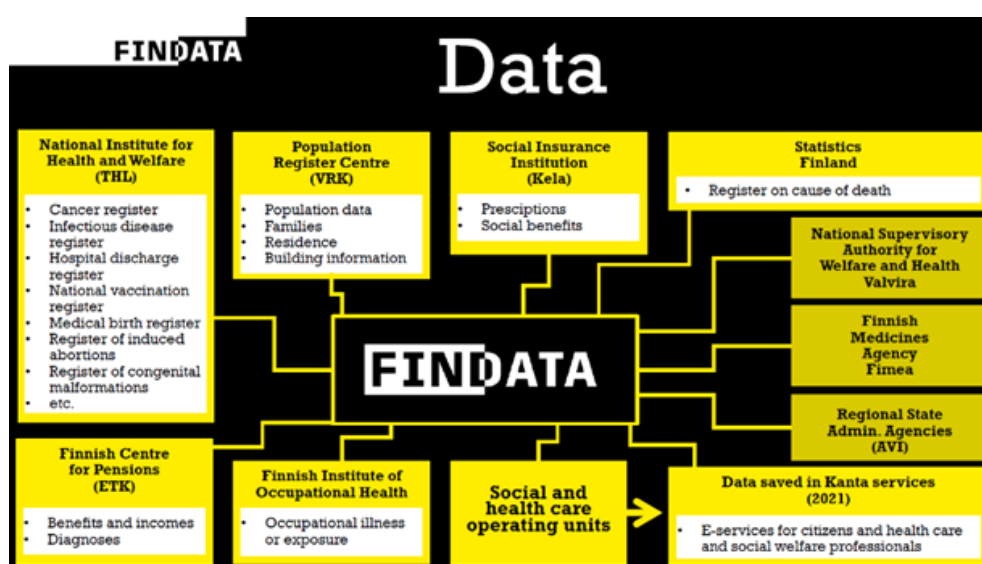
Financing

Findata's 2020 budget, funded from public funds, is **EUR 5.2 million** – in addition to its Isaacus precursor project that had a budget of EUR 14 million (of which an approximate three quarters were linked to the establishment of Findata). This budget is likely to decrease in time, as Findata's operations gain in maturity and as fee-based revenues increase.

These fees are of a triple nature and consist of:

- A **fixed fee of EUR 1,000** for a data permit or request decision (including for EEA re-users), or a EUR 350 fee to amend an existing permit; and
- A **processing** fee (for combining, pre-processing, pseudonymisation and anonymisation) of **EUR 115 per hour worked**;
- A fixed fee for the remote access environment package that ranges between EUR 2,250 and EUR 8,500 (excluding customisation of the environment for an additional fee of EUR 115/hour worked).

These fees are intended, on the long-term, to cover Findata's cost but not to make a profit.

Currently, Findata employs 15 people and aims to hire an additional 10 on the long-term. A rough estimate of costs for one FTE is EUR 75,000 per year for Findata (i.e. currently EUR 1.9 million).

Architecture

The data the use of which is subject to the rights of others is handled in a safe and secure environment. Access to data is controlled, and only the results of the analytics can be used externally.

---

[206] Source: Implementation of the national Social and Health Data permit authority Findata. Johanna Seppänen, PhD, Director

Figure 25 - The evolution of the data provisioning[207]
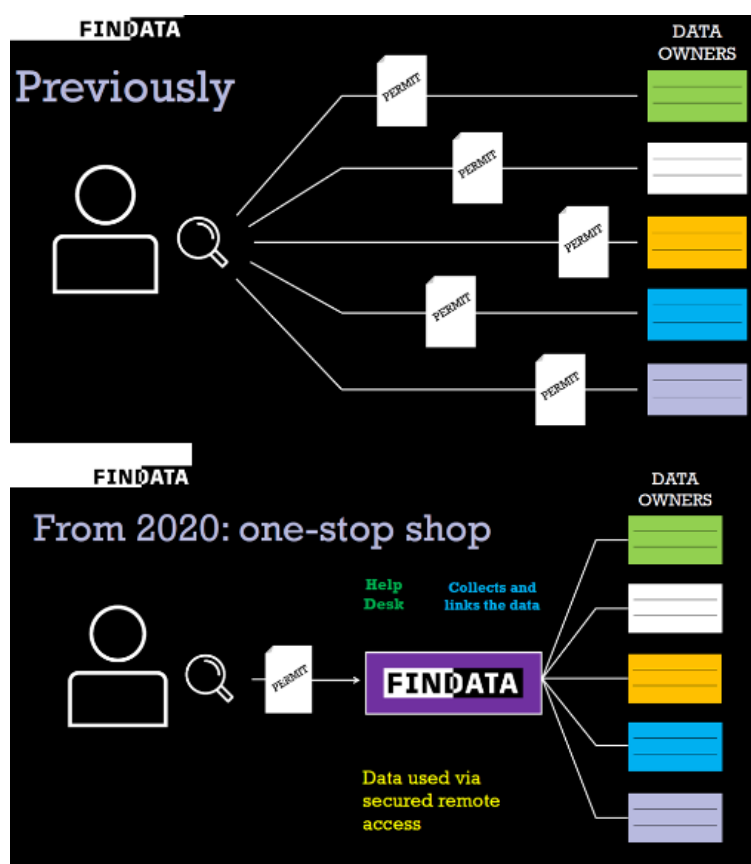
As described in figure 2, previously, obtaining the permits and data was a difficult and expensive process in terms of time. Indeed the user needed to approach each authority and data source separately. Today instead of having to apply for separate permissions from several different data owners, a single central operator/ service operator (Findata) issues and grants research permits[208], including ethical evaluation. After granting the permission to use data, the service operator collects relevant data from different registers and edits, combines and anonymises the data before distributing it to the user. As also depicted in the figure 3, Findata ensures that data handling and transfer of data occurs in a secure environment and that the process meets all the requirements defined by Finnish law. To do this Findata uses a data management system including a secure remote user environment with associated tools. Additionally, a data description system serves as a centralised place for saving the metadata of available materials. The solution includes, among other things, a metadata editor for editing and updating data descriptions.

---

[207] Source: Implementation of the national Social and Health Data permit authority Findata. Johanna Seppänen, PhD, Director

[208] Findata's permit service; https://lupa.findata.csc.fi/
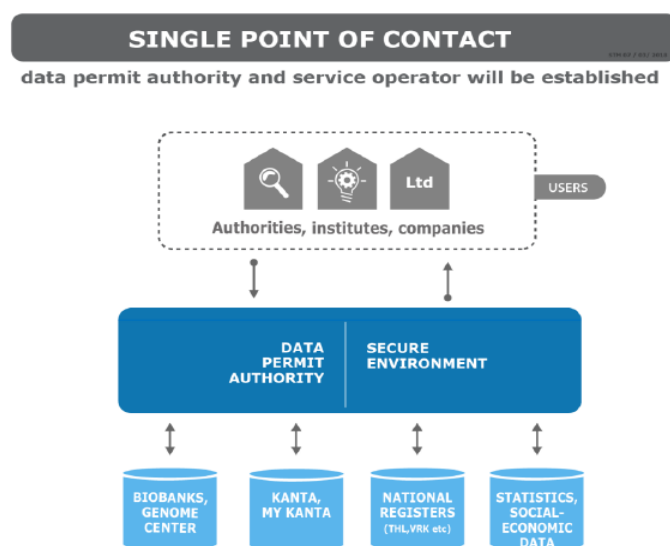
Figure 26 -Access to data[209]

There are two different levels of data and different ways to access related datasets:

- **Individual level data.** The data of this level can be used for scientific research, statistics, education, authorities' steering, supervision, planning and forecasting. This data is available in a remote access environment for a set period. The data has been anonymised or pseudonymised. A data utilisation plan is required for access to data sets.
- **Statistical level data.** The data of this level can be used for the aforementioned purposes and, in addition, for development and innovation and knowledge management. This kind of data are directly delivered to customers.

Health and social data are stored in various national and local databanks. There is a large variety of different kinds of patient record, well-being, social wellness and other data available. The usage of a unique national person ID-number makes it possible to combine personal records.

The following architecture represents the databases in scope and two types of data lakes:

- Local data lakes: County hospitals, local social and health care providers etc. have enormous data in various systems. In many places, the data is now gathered into data lakes.
- National data lakes: Social care data, Patient data and Prescription data, Personal Health Record and social data are stored in national data lakes.

---

[209] Source: Finland – Most advanced ecosystem for healthcare innovation. Nora Kaarela, Head of Industry, Health & Wellbeing, Invest in Finland, Business Finland.
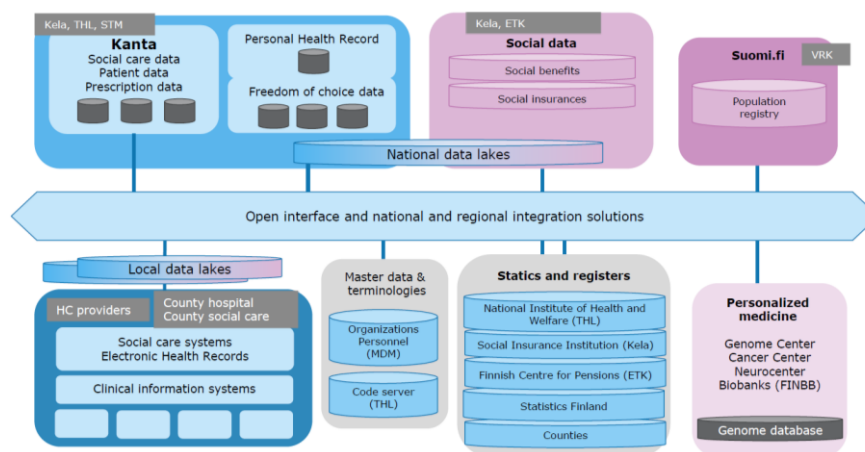
Figure 27 - Data sources and data lakes[210]

RatSWD

Introduction

The **German Data Forum (*Rat für Sozial- und WirtschaftsDaten*)** is a **public advisory council to the German federal government** and was founded in 2004. The RatSWD aims at sustainably **improving the research data infrastructure** that underlies empirical research and at contributing to the international competitiveness of said research.

It is made up of an **independent body of researchers and representatives of data holders**, and acts as an institution of exchange and of mediation between the interests of science and data producers. As such, it is an important platform for **communication and coordination**.

Although RatSWD itself does not make data available to re-users, it is an **intermediary responsible for the accreditation of Germany's Research Data Centres** (RDCs), which act as data holders and sometimes also as data re-users for research purposes. It coordinates these RDCs via a **Standing Committee Research Data Infrastructure (FDI Committee)** established in 2009.

Governance

The RatSWD was established by the **Federal Ministry of Education and Research** in 2004. It operates under rules of procedure determined by the Federal Ministry. These **rules of procedure** govern operations, tasks and competencies of the chair and the business office. The rules of procedure can be changed with a two-third majority of the members of the RatSWD and the consent of the Federal Ministry responsible for research.

The RatSWD is evaluated by the **German Council of Science and Humanities** – the last such evaluation revealed that the RatSWD succeeded in opening up and improving access to data and in creating synergies between scientific community and data holders.

The RatSWD consists of **16 members**. Of these, **8 are representatives from Germany's research community**, and are elected at the Conference for Social and Economic Data, held every 3 years. The remaining **8 members originate from data holders**, specifically from:

- The Federal Statistical Office;
- A (State) Statistical Office of the Länder;

---

[210] Source: Secondary Use of Health and Social Data in Finland, Joni Komulainen Ministrial adviser Master of Laws.

- The IAB Institute for Employment Research of the German Federal Employment Agency;
- An institution from the German social security system;
- An institution from the area of official health data;
- An institution from the area of official financial data; and
- An institution from the area of science-based data production.

The **members from data holders are appointed by the Federal Ministry for Education and Research** based on proposals made by their respective institutions in accordance with the Law for the Composition of Federal Committees (*Bundesgremienbesetzungsgesetz*) and on request from the Federal Ministry for Education and Research.

In addition to these members, up to two elected representatives from the Standing Committee Research Data Infrastructure have a permanent right to attend RatSWD meetings. Likewise, two representatives of the Federal Government, as well as two representatives of the Länder have a right to attend and are entitled to bring forward motions. Further, the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) has a permanent right to attend.

Operations

The RatSWD performs an advisory function, initiates new development and secures quality with regards to standardisation and data quality, and to development of RDCs and data service centres.

Its core tasks are the following:

- To issue **recommendations on further improving the data infrastructure**, specifically:
  - Recommendations on how to secure and further improve data access, particularly by establishing and evaluating research data centers and data service centers according to a set of clear standards;
  - Recommendations on how to improve data use through the provision of scientific and statistical data (research data portal; metadata) and appropriate documentation;
  - Recommendations on research topics and research tasks pertaining to the conceptual development of data infrastructures on the national, European and international level;
  - Recommendations on how to optimise the production and provision of research-relevant data;
- To **advise science and policy**, specifically:
  - Advising the Federal Ministry for Research and the Länder governments on the development of the research-based data infrastructure;
  - Advising public and private data producers;
  - Advising data producers that are institutionally unaffiliated with independent scientific research on how to receive certification as a scientific research institution (certification);
- To **monitor legal and technological developments**, specifically:
  - Monitoring national and legal developments in data provision;
  - Monitoring technological developments, e.g. virtual research environments; and
- To **organise and host the Conference for Social and Economic Data** every three years.

The Conference serves as a **platform for discussing** topics pertaining to empirical social, behavioural and economic data as well as process-produced or survey-based data production. Participation is open to all interested researchers. The German Data Forum furthermore hosts **colloquia, panel discussions and workshops** that foster an ongoing exchange between researchers and data producers.

The RatSWD currently boasts 11 working groups divided into 3 themes.[211] These are:
- **New data sources and data access for researchers**
  - Access to Big Data;
  - Tax and Wealth Data;
  - Further Development of Crime Statistics and Legal Data;
  - Archiving and Access to Qualitative Data;
  - Data Collection With New Information Technology; and
  - Remote Access to Data from Official Statistics Agencies.
- **Further development of the research data infrastructure**
  - Common guidelines in research data centres;
  - Improving access to existing data in research data centres;
  - Decentral archiving structure at research data centres; and
  - Skills development in research data centres (RDCs).
- **Advising of legislators and policy-makers**
  - The social sciences in roadmap processes.

FDI Committee

The RatSWD coordinates 38 accredited RDCs, which work together in a Standing Committee Research Data Infrastructure (FDI Committee) established in 2009.[212] The FDI committee produces recommendations for the RatSWD, and coordinates cooperation among RDCs with a view to continuously improve the research data infrastructure and to facilitating data access for researchers. The Committee also serves as a peer-review/peer-pressure mechanism to ensure the adherence by RDCs to the quality criteria these must uphold; as a complaints body to reach amicable settlements. It also attempts to standardise usage conditions such as charging practices among RDCs.

Figure 28 - The FDI Committee within the RatSWD
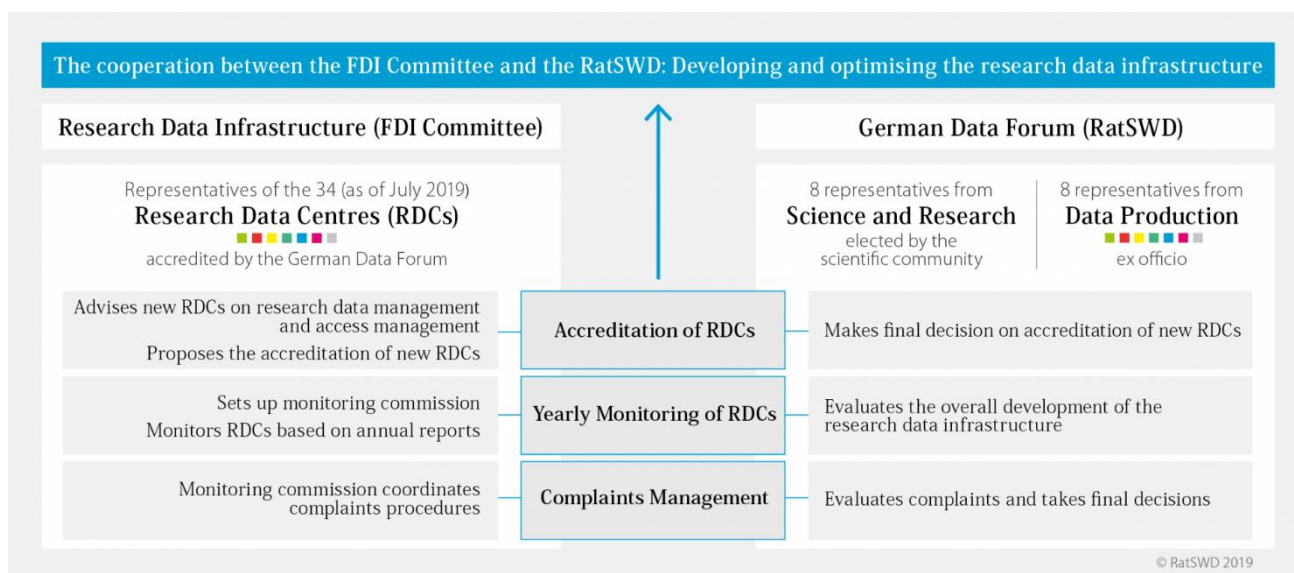
The FDI Committee lists a number of key activities for 2017-2020.[213] In particular, the Committee aims for:
- the harmonisation of processes in the RDCs (e.g. data use contracts);
- the expansion of access to research data from RDCs, for example, via guest researcher work stations of other RDCs;

---

[211] https://www.ratswd.de/en/activities/working-groups
[212] https://www.ratswd.de/en/data-infrastructure/fdi
[213] https://www.ratswd.de/en/data-infrastructure/fdi

- the opening up of the existing data infrastructure for scientific and official data producers; and
- the advancement of skill development training for the RDCs' employees and data users.

In order to be accredited by the RatSWD, the RDCs need to meet three core criteria.[214] They must:

- Provide at least one data access path;
- Provide sufficient data documentation; and
- Ensure the long-term availability of the data.

In addition, RDCs must be "fully operational", which is defined as having been in operation for at least six months, and having at least three external data re-users. Lastly, initial accreditation requires compliance with additional information criteria used to assess the scope and quality of the RDC's operation. These information criteria relate to the:

- Scope and development of the social, behavioural, and economic data provided
- Method for timely data provision
- Provision of tools
- Quality assurance of datasets
- Data protection safeguards in due consideration of the interests of researchers
- Service concept
- Single entity comprising institution and research data centre
- Provision of all datasets relevant to research
- Overlap and distinct features compared to existing RDCs
- Research activities
- Multiple provision of the same data (multiple hosting, not hosting at multiple sites)
- Time to process applications
- Staff
- Infrastructure development

All accredited RDCs contribute to annual reporting by completing a questionnaire. As with accreditation, the questionnaire is based on the mandatory and information criteria. The FDI Committee elects a monitoring commission for a three-year term concurrent with the German Data Forum's (RatSWD) appointment period. The main task of the monitoring commission is to collect and assess the research data centres' annual reports. Moreover, the commission handles complaints regarding RDC accreditation criteria and provisional accreditations.[215]

Financing

RatSWD is funded by a grant from the Federal Ministry of Education and Research.[216] In its early years, the annual cost of running RatSWD ranged between EUR 200,000 and 300,000. This has significantly increased however: RatSWD regularly holds meeting with representatives of the accredited RDCs. These have grown in number, and RatSWD is responsible for their travel.

Overall, RatSWD's current **annual budget amounts to EUR 900,000**, of which:

- Half (roughly EUR 450,000) corresponds to human resources;
- EUR 300,000 corresponds to expenses linked to RDCs meetings;
- EUR 75,000 go to renting RatSWD's premises and to technical support;

---

[214] https://www.ratswd.de/en/info/accreditation
[215] https://www.ratswd.de/en/info/monitoring-and-complaints-management
[216] Figures, particularly regarding the RatSWD's budget, will be collected via interviews.

- EUR 30,000 go to translation and ad-hoc legal advice on some publications; and
- EUR 25,000 cover travel expenses of RatSWD's staff.

Of the 30 RDCs accredited up until 2016, 21 report not charging any fees for providing access to data (see Fig. 6). The fees charged by the nine other RDCs are low (in the two-digit euro realm) and mostly used to cover the costs for media and contracting.[217] RatSWD itself is a free service.

Architecture

Partly due to data sensitivity and respective legal regulations, the data offered by the RDCs must remain at their respective data producing institutions. Therefore, the decentralised structure of the research data infrastructure is a tried and tested way to satisfy the demands of data producers, data users in science and research, and data protection.

RDCs' main responsibilities are:

- **Providing researchers with user-friendly, transparent, and high-quality access to data**. So far, this data has concerned mostly microdata that can be analysed statistically. The data are collected as part of official statistics, administrative operations, research projects, or scientific survey programmes. In their capacity as mediators, the RDCs help improve cooperation between data users and data producers.
- **Ensuring that data users comply with federal data protection policies and, if applicable, with policies specific to individual research areas, by taking appropriate technical and organisational measures**. Depending on the level of anonymisation (see info box 1), datasets are offered for off-site use (via download or mail order) in the form of Scientific Use Files (SUF), Public Use Files (PUF), or Campus Files (CF). Moreover, the generation of synthetic data can be an option to support research needs. To facilitate access to highly sensitive microdata, the RDCs offer the option of on-site use. In this case, users can access the data at a guest researcher workstation on the premises of the RDC.
- **Ensuring equal treatment of all data users by means of transparent and standardised application and access policies**. Incoming applications are not assessed with regard to the content of the proposed research; they are only reviewed in terms of their compliance with contractual or data protection policies.
- **Creating easy-to-analyse data products featuring quality-assured, standard-compliant metadata and comprehensive documentation**. The RDCs present information on their respective data services via their websites, in data and method reports, at scientific events, or in individual advising sessions.
- **Conducting independent research using the data they offer**. This helps ensure that each RDC has strong expertise regarding the data and their quality. At the same time, ongoing scientific discussions about methods and contents can inform the advising services provided to data users. Research activities by RDC staff do not involve any exclusive access to data products

Depending on the content or the unit of observation, making such data available is subject to various legal requirements, most notably by the EU General Data Protection Regulation and the German Federal Data Protection Act but also state data protection legislation, the German Social Code, and the Federal Statistics Act.

---

[217] https://www.ratswd.de/dl/RatSWD_Output1.6_QualityMgmt.pdf

In 2016, the RDCs offered a total of 3,214 datasets. The RDCs provide a wide range of access paths to their data. Generally, there are two distinct basic ways of accessing the data: on-site (i.e. on the physical premises of the RDC) and off-site (i.e. outside the RDC). Several RDCs offer multiple paths of access.[218]

Figure 29 – Data access paths in different RDCs
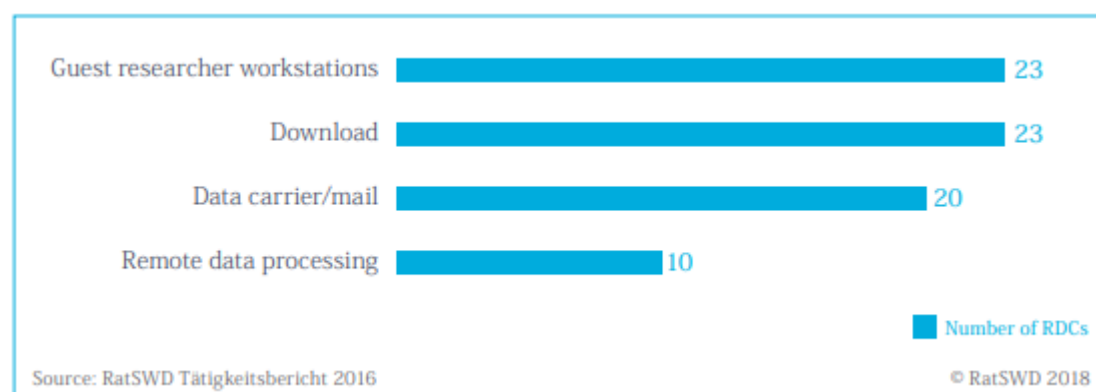
Note: These figures are from 2016 and only include 30 RDCs.

The RatSWD provides a search engine for data provided by most of the 38 RDCs.[219] This engine is da|ra, the registration agency for social and economic data that enables holders to register their data in order for it to be stored and easily identifiable on the long-term by researchers.[220] Still, one stakeholder pointed to the difficulty to ensure good quality metadata, especially with respect to historic data.

In addition to the Forschungsdatenzentrum (see section 2.1), the Research Data Centre of the German federal Employment Agency is one other accredited Research Data Centre.[221] It provides researchers with micro-data on social security and employment for research purposes. This data may be accessed using Scientific Use Files,[222] remotely, and on-site. In addition, it offers advice on data selection, access, handling, analysis potential, scope and validity; it regularly updates datasets; produces research; and organises workshops and conferences. Access to data requires an application and the conclusion of a user agreement.

Likewise, the Research Data Centre for Higher Education Research and Science Studies makes available quantitative and qualitative data from universities and research institutes to researchers, following an application process.[223] Information collected as part of this process includes the focus on the study, the data collection methods and types of data used, and the duration and status of the study. Access to data is done either via Scientific Use Files or Campus Use Files (anonymous data sets for teaching and exercise purposes at universities). Access can be on-site in a controlled environment, remotely via a virtual desktop, or through downloading highly anonymised datasets (see figure below).

---

[218] https://www.ratswd.de/dl/RatSWD_Output1.6_QualityMgmt.pdf
[219] https://www.ratswd.de/en/researchdata/search
[220] http://www.da-ra.de/en/home/
[221] https://fdz.iab.de/en.aspx
[222] These are de facto anonymized data records for scientific purposes.
[223] https://www.fdz.dzhw.eu/de

Figure 30 - Usage potential (*y* axis) and level of anonymisation (*x* axis) of microdata at the RDC for Higher Education Research and Science Studies
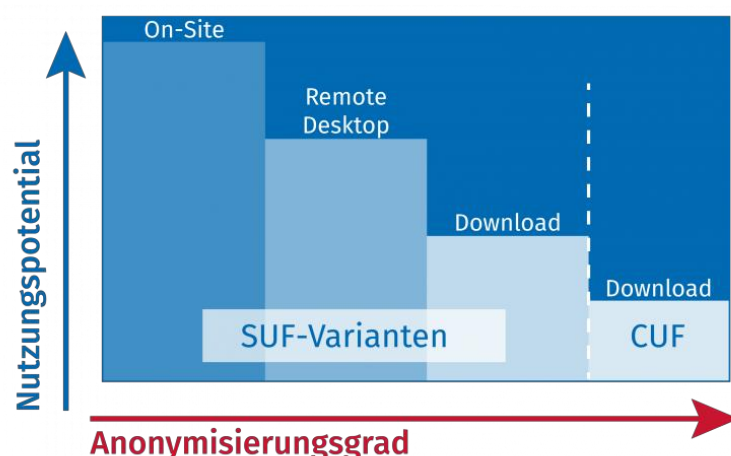
National Safe Haven (Scotland)

Introduction

The UK National Health Service (NHS) **National Services Scotland (NSS) national safe haven service** – allows data from electronic records to be used to support research when it is not practicable to obtain individual patient consent, while protecting patient identity and privacy. It provides **secure file transfer and submission services to data providers** and additional services (e.g. analytics platforms) to researchers.

Data safe havens provide a **secure and safe environment**, supported by trained staff and agreed processes to facilitate statistics and research work on **sensitive data**, including medical data (e.g. patient records and MRI images) and social data (e.g. census, government or police data). Health data can be processed and linked with other health data (and/or non-health related data) and made available in a de-identified form for analysis. Save havens serve as safeguards for confidential information which is being used for research purposes. Any researchers applying for access to health data must adhere to the Safe Haven principles,[224] namely:

- The ultimate aim of information sharing is to provide care;
- Citizens should have the choice about the use of their data; and
- Dialogue with the public should be maintained.

Safe Havens in Scotland were established as part of a national need for delivering research excellence and the need for rapid access to high-quality health data for research purposes. They were developed in line with the SHIP blueprint which outlined a programme for a Scotland-wide research platform for the collation, management, dissemination and analysis of anonymised Electronic Patient Records (EPRs). The agreed

---

[224] EPCC, NHS National Services Scotland (NSS) national safe haven. See: https://www.epcc.ed.ac.uk/projects-portfolio/nhs-national-services-scotland-nss-national-safe-haven. NHS Scotland, Data Safe Haven. See: https://www.nhsresearchscotland.org.uk/research-in-scotland/data/safe-havens. EUDAT Slides, Data safe havens: a future EOSC service? See: https://www.slideshare.net/EUDAT/data-safe-havens-a-future-eosc-service

principles and standards to which the Safe Havens are required to operate are set out in the Safe Haven Charter.[225]

NHS Scotland allows research using routinely collected, unconsented patient data. Additionally, these data can be linked to social data such as education. The research that this enables can have an enormous public benefit but the use of this data must be managed very carefully to safeguard privacy and maintain public trust and support.[226]

This case study details how Scotland's National Safe Haven operates in practice.

Governance

Safe havens are subject to strong national information governance policies.[227] The Scottish National Safe Haven within Information Services Division (ISD) is an expert unit within the Common Services Agency (known as NHS National Services Scotland).[228] The National Safe Haven is part of the Scottish Informatics Linkage Collaboration (SILC) that also includes the electronic Data Research and Innovation Service (eDRIS) and the National Records of Scotland indexing service. SILC facilitates linkage for research and statistical activities across many sectors including the NHS.

The NHS Research Scotland (NRS) nodes received funding, through NRS infrastructure allocations from the Chief Scientist Office, to help establish the Safe Havens.[229] Together, the National Safe Haven within SILC and the four NRS Safe Havens have formed a federated network of Safe Havens in order to work collaboratively to support health informatics research across Scotland. All the Safe Havens have the individual responsibility to operate at all times in full compliance with all relevant codes of practice, legislation, statutory orders and in accordance with current good professional practice. Each Safe Haven may also work independently to provide advice and assistance to researchers as well as secure environments to enable health informatics research on the pseudonymised research datasets they create. The governance of SILC involves representatives of the Safe Havens.[230]

EPCC at the University of Edinburgh, under contract with the NHS is now the operator of the new NHS National Services Scotland (NSS) national safe haven and responsible also for building, supporting, maintaining and hosting it, in collaboration with the Farr Institute of Health Informatics Research which provides the infrastructure. EPCC continues to develop the infrastructure and software to further enhance the service.[231]

[225] NHS Scotland, Data Safe Haven. See: https://www.nhsresearchscotland.org.uk/research-in-scotland/data/safe-havens.
[226] EPCC, The National Safe Haven for research using unconsented NHS data. See: http://www.epcc.ed.ac.uk/blog/2017/11/21/national-safe-haven-research-using-unconsented-nhs-data
[227] EUDAT Slides, Data safe havens: a future EOSC service? See: https://www.slideshare.net/EUDAT/data-safe-havens-a-future-eosc-service
[228] Scottish Government, A Charter for Safe Havens in Scotland. See: https://www.abdn.ac.uk/iahs/documents/00489000.pdf
[229] Scottish Government, Charter for Safe Havens in Scotland: Handling Unconsented Data from National Health Service Patient Records to Support Research and Statistics. See: https://www.gov.scot/publications/charter-safe-havens-scotland-handling-unconsented-data-national-health-service-patient-records-support-research-statistics/pages/4/
[230] *Ibid.*
[231] EPCC, NHS National Services Scotland (NSS) national safe haven. See: https://www.epcc.ed.ac.uk/projects-portfolio/nhs-national-services-scotland-nss-national-safe-haven.
EUDAT Slides, Data safe havens: a future EOSC service? See: https://www.slideshare.net/EUDAT/data-safe-havens-a-future-eosc-service
EPCC, The National Safe Haven for research using unconsented NHS data. See: http://www.epcc.ed.ac.uk/blog/2017/11/21/national-safe-haven-research-using-unconsented-nhs-data

Operations

The new NHS National Services Scotland (NSS) National Safe Haven service implementation work started in September 2015 with the live service rolled out during December and January 2016. Now fully operational, the safe haven is both physical and remote. It offers a secure file transfer and submission service for data providers and a range of access methods and analytics platforms and tools for researchers.[232] The Scottish national safe haven, as all Data Safe Havens, provides a platform for the use of NHS electronic data in research feasibility, delivery and pharmacovigilance offering:

- A secure environment, with trained staff operating to agreed principles and standards;
- A federated network across Scotland working collaboratively to deliver and develop health informatics capability;
- Access to a wide range of anonymised datasets (including national datasets through to specialised local datasets);
- Robust governance procedures to protect the confidentiality of the data;
- Single costing and contracting; and
- Daily updates.[233]

Working to agreed principles and standards, Safe Havens in Scotland provide access to health data and services to enable research while protecting the confidentiality of the data. Data remains under the control of the NHS and complies with legislation and NHS policies.[234]

The National Safe Haven has proven very successful and has supported more than 200 research projects over the two years it has been running. EPCC is currently working with the Health Informatics Centre at the University of Dundee to provide researchers with access to the Scottish NHS imaging data (X-ray, CT, MRI, ultrasound etc.) dating from 2010 onwards. Such data offers tremendous opportunities for a wide variety of research including examining early/preclinical diagnosis, disease progression, personalised medicine, genotype-phenotype associations, and the development of novel computer vision and machine learning algorithms. Additionally, the ability to link this image data with patient outcome data is unique to Scotland and offers even greater potential for world leading research that will have a major contribution to the future health of the nation.[235]

eDRIS was established as a specific ISD function within NSS and provides a single point of contact for advice on research project design and development as well as access via the National Safe Haven to a wide range of national datasets. Given the well-established close working relationship with ISD (the Data Controller of national NHS Scotland datasets), the National Safe Haven may be best placed to take the lead when research requires the processing and linkage of national datasets.[236]

Financing

The NRS nodes received funding through NRS infrastructure allocations from the Chief Scientist Office to help establish Safe Havens.[237] In particular, Chief Scientist Office (CSO) has invested significantly in Safe

---

[232] EPCC, NHS National Services Scotland (NSS) national safe haven. See: https://www.epcc.ed.ac.uk/projects-portfolio/nhs-national-services-scotland-nss-national-safe-haven.
[233] NHS Scotland, Data Safe Haven. See: https://www.nhsresearchscotland.org.uk/research-in-scotland/data/safe-havens
[234] *Ibid.*
[235] EPCC, The National Safe Haven for research using unconsented NHS data. See: http://www.epcc.ed.ac.uk/blog/2017/11/21/national-safe-haven-research-using-unconsented-nhs-data
[236] Scottish Government, Charter for Safe Havens in Scotland: Handling Unconsented Data from National Health Service Patient Records to Support Research and Statistics. See: https://www.gov.scot/publications/charter-safe-havens-scotland-handling-unconsented-data-national-health-service-patient-records-support-research-statistics/pages/4/
[237] *Ibid.*

Havens in each of the four NRS nodes, providing a platform for the use of NHS electronic data in research feasibility, delivery and pharmacovigilance.[238]

There is no particular formula used to allocate funding to the National Safe Haven (NHS NSS) but this is being done based on an assessment of need.[239] According to the Chief Scientist Office Outturn Summary for 2018-19, the amount of £150,000 was allocated to ISD Safe Haven [NHS National Services Scotland] in order to support a national data safe haven for the safe and secure provision and linkage of de-identified data from national health data sets for approved research. Additionally, CSO invests around GBP 40 million each year to support NHS Research Scotland to conduct research. The purpose of the principal funding streams is to support, among others, NRS infrastructure including NRS biorepositories and data safe havens.[240]

Architecture

The Scottish National Safe Haven is a secure environment where the project data is uploaded and accessed. It offers a high powered computing service, secure analytic environment, secure file transfer, and a range of analytic software including SPSS, STATA, SAS and R. The IT infrastructure is provided by the EPCC at Edinburgh University.

To use the unconsented NHS data, researchers must apply to the Public Benefit and Privacy Panel specifying the data that is required and the purpose of the research. If granted permission, the data will be selected, anonymised and linked (a process called pseudo-anonymisation) before being placed within the National Services Scotland (NSS) National Safe Haven. Researchers must process their data from within this National Safe Haven infrastructure.[241]

A data controller may require that their data is only accessed through a secure access point to ensure data security. A secure access point is a dedicated computer in a physically secure area where no external devices can be used or connected. The secure access point does not connect to the internet nor can it be accessed remotely. In case of work in a secure access point, a username and password for the linked data files will be given. The location of the secure access points may depend on the data to be accessed, being either at the offices of BioQuarter at the Royal Infirmary of Edinburgh or at selected universities across Scotland.

In some cases, data controllers may allow the researchers remote access to the data. This will be via a VPN (virtual private network). To remotely access the safe haven there is a 2 factor authentication process, the first part of which will be receipt of an access code via mobile phone. To access the National Safe Haven, researchers must use remote desktop software and log in using a high-security protocol. Researchers are then able to use the remote desktop session to access, process and analyse the data they have requested. The remote computer is installed with several statistics packages that the researcher can use. Crucially, the remote computer offers no access to the Internet neither to receive nor send data. Researchers can request files to be transferred into, or out of, the Safe Haven but such requests are subject to a manual verification process to ensure privacy is never breached.[242]

---

[238] Chief Scientist Office, Initiatives. See: https://www.cso.scot.nhs.uk/about/initiatives/
[239] Scottish Government, Formula used to allocate funding to NHS Boards: FOI release. See: https://www.gov.scot/publications/foi-18-02515/
[240] Chief Scientist office & Scottish Government, Chief Scientist Office Outturn Summary 2018-19. See: https://www.cso.scot.nhs.uk/wp-content/uploads/CSO1819OTsummary.pdf
[241] Public Health Scotland, Use of the National Safe Haven. See: https://www.isdscotland.org/products-and-services/edris/use-of-the-national-safe-haven/
EPCC, The National Safe Haven for research using unconsented NHS data. See: http://www.epcc.ed.ac.uk/blog/2017/11/21/national-safe-haven-research-using-unconsented-nhs-data
[242] Public Health Scotland, Use of the National Safe Haven. See: https://www.isdscotland.org/products-and-services/edris/use-of-the-national-safe-haven/

Outputs from the analyses are only released for the agreed purpose of the research. Data cannot be used in any other way. Outputs cannot be released until the Research Co-ordinator has assessed them for statistical disclosure control in line with the data controller's requirements specified for the study. The objective of this is to ensure that an output does not contain information which could be used either on its own or in conjunction with other data to breach an individual's privacy. Data is held in archive for a specific period set by the data controller, after which the study data are being deleted (upon prior confirmation). It is not possible to restore study data, as deletion is permanent.[243]

Statistics Denmark

Introduction

Statistics Denmark is the central authority on Danish statistics. Its mission is to collect, compile and publish impartial statistics on the Danish society, as a basis for democracy and the economy, covering a range of subjects, including:[244]

- Population and elections (including population and population projections; immigrants and their descendants; births; deaths and life expectancy; households, families and children; marriages and divorces; migrations; names; elections);
- Labour, income and wealth (including labour force participation; employment; unemployment; commuting; absence and work stoppages; earnings and labour costs; income; wealth and debt);
- Living conditions (including gender equality; quality of life; survey on living conditions; housing; health; childcare; persons receiving public benefits; social conditions; traffic accidents; criminal offences);
- Education and knowledge (including population by status of education; full-time education; courses and adult education; educational transitions; research, development and innovation; information society);
- Business Sector in general (including enterprises in general; enterprise development; accounts; globalisation; organic production and trade; tendency surveys); and
- Particular business sectors (including agriculture, horticulture and forestry; fishery and aquaculture; manufacturing industries; construction; transport; distributive trades; tourism; services sector).

This case study details how Statistics Denmark operates in practice.

Governance

Statistics Denmark is a state institution under the Danish Ministry of Social Affairs and the Interior. The executive board of Statistics Denmark is composed of the Director General and five Directors. The head of the management office also attends the weekly management meetings.

Statistics Denmark is managed by a supervisory board with the National Statistician as chair, and it includes seven other members. The supervisory board establishes its own rules of procedure and appoints one of its

---

EPCC, The National Safe Haven for research using unconsented NHS data. See:
http://www.epcc.ed.ac.uk/blog/2017/11/21/national-safe-haven-research-using-unconsented-nhs-data
[243] Public Health Scotland, Use of the National Safe Haven. See: https://www.isdscotland.org/products-and-services/edris/use-of-the-national-safe-haven/
[244] Statistics Denmark, About us. See: https://www.dst.dk/en/OmDS

members as vice-chair. By law, the supervisory board chaired by the National Statistician is responsible for the following:

- The overall strategy and financial management of Statistics Denmark, while the National Statistician has the sole responsibility for defining the professional criteria for development, collection, compilation and dissemination of Statistics Denmark's statistical output.
- The professional independence of the official statistics and of the institution of Statistics Denmark.
- Consideration and decisions in matters of interest to the strategic management of the institution, including work programme, statistical programme and budget. Decisions as to the extent and ways of collecting data from the business community, including for the purpose of implementing EU and national legislation. In this way, the supervisory board is responsible for the reporting task imposed by Statistics Denmark on the business community.
- Via the work programme and statistical programme, decisions about the data that public authorities and institutions must submit to Statistics Denmark.[245]

Operations

Authorisations to access de-identified microdata may be granted to researchers pre-approved by Statistics Denmark. These can be from public sector research organisations (such as ministries or universities) and from private sector non-profit foundations, NGOs and consultancies. Foreign researches cannot obtain access, unless via a Danish organisation.

The Division of Research Services is the one responsible for the provision of statistical microdata for research purposes. Annually, there are around 2,000 applications for access to de-identified data – of which 75% originate from the public sector, and 25% from the private sector.

The main tasks that Statistics Denmark perform with regards to microdata are:

- Examining and assessing applications to access data;
- De-identifying the data;
- Standardising past data (i.e. data dating back decades) to make it usable in combination with recent data;
- Fetching the requested data and merging different datasets; and
- Answering questions from reusers on specific datasets.

This is in addition to the back-office work, including research and development and coordination with other data holders.

Financing

Statistics Denmark's overall annual budget is DKK 35 million (i.e. EUR 4,7 million), of which DKK 7 million (EUR 940,000) is public funding from the Danish Ministry of Science. The remaining amount is fetched via user fees, and the share of budget covered by user fees is increasing.

Statistics Denmark applies different fees depending on the status of the applicant: public sector researchers pay a lower amount than researchers from the private sector, and this difference corresponds to the DKK 7 million of public funding. In other words, these public funds are used to subsidise public sector research.

---

[245] Statistics Denmark, Supervisory Board. See: https://www.dst.dk/en/OmDS/organisation/bestyrelsen

Of the total DKK 35 million budget, a tenth (DKK 3,5 million, or EUR 470,000) covers the cost of maintaining Statistics Denmark's secure data processing environment.

Architecture

Microdata is not handed over to researchers, but rather accessed via a research server at Statistics Denmark. It is separate from the other networks and contains exclusively de-identified microdata for research purposes. Remote (and encrypted) access is possible via the internet, following an agreement with Statistics Denmark.

This agreement states that all work on the microdata must take place on the server and prohibits attempts to remove microdata from the server or to identify individuals or businesses. However, aggregated data may be removed from the server.

All aggregated results from the researchers´ computer can be stored in a special file and such printouts are sent to the researchers by e-mail. This is a continuous process (every five minutes) and has shown to be quite effective. The advantage for Statistics Denmark is that all e-mails are logged at Statistics Denmark and checked by the Research Service Unit.

Several computer packets are available on the research server, such as SAS, SPSS, STATA, GAUSS and R. The programs are frequently updated with new versions.[246]

OpenSAFELY

Introduction

The OpenSAFELY project was developed in the United Kingdom in view of the global COVID-19 emergency, as a collaboration between the DataLab at the University of Oxford, the EHR group at London School of Hygiene and Tropical Medicine, TPP and other electronic health record software companies (who already manage NHS patients' records), working on behalf of NHS England and NHSX, with a growing list of broader collaborations including ICNARC. The team is composed of software developers, clinicians, and epidemiologists, all pooling diverse skills and knowledge to deliver high performance, highly secure and accurate health data analytics, using modern open software development techniques. The project has delivered because of its mixed skillset software developers, and "developer-epidemiologists", who can speak the same language as the technical teams within EHR system suppliers.[247]

This case study details how OpenSAFELY operates in practice.

Governance

The OpenSAFELY team works on behalf of NHS England, who is acting as Data Controller for the purposes of this project, while each EHR holder acts as Data Processor. The Secretary of State for Health issued NHS England/Improvement a notice under the Health Service (Control of Patient Information) Regulations 2002 3(4) which enabled NHS England to collect the data required from GP practices directly from their EHR vendor. All information governance for this urgent project is handled by NHS England.

The Data Protection Impact Assessments approving data flows and access approves linking GP data to outcomes data from the new NHS England and NHSX data store and other sources including COVID–19

---

[246] Statistics Denmark, The Danish System for Access to Micro Data. See:
https://www.dst.dk/ext/645846915/0/forskning/Access-to-micro-data-at-Statistics-Denmark_2014--pdf
[247] OpenSAFELY, Home. See: https://opensafely.org/#appendix-2-further-detail

Patient Notification System (CPNS) deaths data; Intensive Care National Audit & Research Centre (ICNARC) ITU admissions data; Second Generation Surveillance System (SGSS) PHE test data; Emergency Care Data Set (ECDS) Accident and Emergency (A&E) patient-level data; and Office for National Statistics (ONS) death data.[248]

Operations

OpenSAFELY deployed a new statistical analysis platform during the Covid-19 emergency to deliver urgent answers on key clinical and public health questions. It is successfully delivering analyses covering 40% of practices in the country and process data on over 24 million patients including their previous medical history, investigations, and current or past medications. Its first analysis identifies which patients are most at risk of death in hospital from COVID-19, with more accuracy than any previous analyses by an order of magnitude. The team has an extensive ongoing collaborations across the scientific community, running analyses to identify which patients are most at risk, and why, which treatments increase or decrease risk. The project is also supporting modellers to understand, evaluate and predict the spread of the disease and pressure on NHS services, using hyperlocal real-world data.[249]

Financing

OpenSAFELY team has developed and deployed a fully functional platform in five weeks with no funding. With modest financial resources, the team will sustain, accelerate, and expand its work. OpenSAFELY currently has funding applications under review with UK National Institute for Health Research (NIHR) and with UK Research and Innovation (UKRI).[250]

Architecture

OpenSAFELY is a secure analytics platform for electronic health records in the NHS England, delivering pseudonymised analyses. Its analytic software is open for security review, scientific review, and re-use. OpenSAFELY uses a new model for enhanced security and timely access to data: it does not remove large volumes of potentially disclosive (i.e. allowing for identification) pseudonymised patient data from the secure environments managed by the electronic health record software company; instead, trusted EHR analysts can run large scale computation across near real-time pseudonymised patient records inside the data centre of the electronic health records software company. This pragmatic and secure approach has allowed to deliver the first analyses in just five weeks from project start. [251]

All data that carries any privacy risk (even a theoretical risk, and even when pseudonymised) remains within the secure data centre of the electronic health record holder, where it already resides. This also means that all activity is logged for independent review. All processing takes place in the same secure data centre, where the patients' electronic records were already stored. The only information to ever leave the data centre is summary tables (with low numbers suppressed) from statistical models. Within the data centre, all pseudonymised data is stored in a tiered system of increasingly less disclosive data stores tailored to each analysis.

All underlying software and research code is open to review for security profiling, scientific evaluation, and to re-use as open source tools improving science across the community. Overall, this approach is therefore highly secure, and supports high quality science: in contrast to working on intermittent "data extracts",

---

[248] Ibid.
[249] *Ibid.*
[250] *Ibid.*
[251] *Ibid.*

OpenSAFELY's approach also ensures that the statistical models run across up-to-date records, which is vital during a global health emergency.

OpenSAFELY's approach to privacy and security exceeds standards for many other current EHR analysis projects. It severely restricts SQL query access to the "event-level" data, which would otherwise present the highest theoretical privacy risk. It then abstracts the key clinical features of each patient for each analysis into a "feature store" for statistical analysis: this summary data is perfectly matched to the needs of each project, but substantially less vulnerable to re-identification attacks; it is nonetheless still managed to the highest privacy standards, as if it were security-critical event-level data. All access to the secure platform is over highly secure VPN from specific IP addresses and MAC addresses for a very small number of highly trusted, named and experienced analysts whose activity is all fully logged. By building our analytics platform inside the originating EHR vendors' data centre, the team completely avoids transporting large raw primary care datasets which would otherwise present a substantial privacy risk, even when pseudonymised. [252]

The data linked include the full coded primary care record containing all previous medical history, test results, diagnoses, medications, treatments, and more; A&E attendance data; hospital death from Covid-19; ITU data; ONS death data including cause of death. The team is able to rapidly map and link new datasets where required. This big data approach with an unusually large volume of primary care data is necessary to get sufficient statistical power to detect associations with specific medications and medical conditions as early as possible during the pandemic and thereby save lives by modifying patient, clinician, and population behaviour. All code for the platform is compliant with open standards and designed to be portable, so that it can run against any platform produced by the NHS in the future to securely store rich and linked primary and secondary care patient data. [253]

Statbel

Introduction

Statbel, the Belgian statistical office, collects, produces and disseminates reliable and relevant figures on the Belgian economy, society and territory to frame complex issues and dilemmas and provide some sort of support for the society. As an official statistical institution, it offers a huge range of figures in terms of economy, population and demography, labour market, poverty, agriculture, industry, services, real estate, transport and traffic, environment, etc. These figures are available at national, regional, provincial, municipal and even more detailed level, as well as within a European context. [254] Statbel also makes microdata available for research to public institution and research centres. This case study details how Statbel operates in practice.

Governance

The Belgian statistical office, Statbel, is part of the general management of the **Federal Public Service of Economy**.[255] It is one of the partners of the Institution for the National Accounts.

Provision of statistical microdata for research is done in accordance with article 15 of the law of 4 July 1962 on public statistics.

---

[252] *Ibid.*
[253] *Ibid.*
[254] Statbel, Who we are. See: https://statbel.fgov.be/en/about-statbel/who-we-are
[255] Statbel, Mission, Vision and Objectives. See: https://statbel.fgov.be/en/about-statbel/who-we-are/mission-vision-and-objectives

Operations

The following microdata is available to researchers from public bodies and research institutes:

- Road accidents;
- 2011 Census;
- Causes of death;
- Household budgets;
- Infant deaths (under one year old);
- Divorces;
- Workforce;
- Marriage;
- Birth; and
- Income and living standards.

These are pseudonymised. To obtain access, researchers must contact Statbel or a specific statistician via email, and submit a formal request for data using a dedicated form. These must be submitted along with a preliminary contract via email. The requester is required to indicate which measures it has in place with regards to IT and physical security of the data, as well as the measures taken to ensure compliance with the GDPR when applicable. The requester must also clearly and in a detailed way indicate how it will use the data requested, as well as who within their organisation will access them

The request is then examined, and the final decision rests with Statbel's Director-General. Should the request be accepted, Statbel will draft and send a final contract that must be signed, and will deliver the data upon reception of the signed contract.

The result of the research, including analysis, studies and statistics produced, must be made available free of charge to Statbel, which may use them as it sees fit. Statbel may also forbid their publication.

Financing

The global budget of Statbel in 2020 was EUR 24 million, of which EUR 20.8 million or about 87% went to human resource costs. On December 1st 2019, Statbel employed 356 people, of which 40% were highly skilled employees, although there is a rising trend in this percentage. In addition to its own staff, there are another 350 surveyors conducting surveys among citizens for Statbel.[256] It must be noted that these numbers are for the whole of Statbel and thus not solely the budget and workforce needed for the handling of sensitive data: because the microdata for research service is not a service, organisationally speaking, the budget cannot be estimated. However, approximately 1 FTE works on providing microdata for researchers, according to Statbel.

The provision of microdata is fee-based: costs are intended to cover the cost of the production and delivery of the data (including the administrative work this entails), as well as to limit the number of requests to only those that cannot be fulfilled using the freely available aggregated datasets. This fixed fee of EUR 500 only applies to entities that are not government (national, regional, or local) entities, i.e. universities, research centres and others, but not ministries or cities.

---

[256] Statbel, 2020 Statistical programme. See:
https://statbel.fgov.be/sites/default/files/Over_Statbel_FR/ProgrammeStatistiqueStatbel2020_nl.pdf

Architecture

Due to historical reasons, Statbel's way of providing access to microdata for researchers differs from that of the other statistical offices examined in this study.

Indeed, after a request is approved and the necessary administrative steps taken, a Statbel statistician makes the requested anonymised microdata available in Statbel's data warehouse, and sends a copy of the file to the researcher's organisation (independent researchers may not receive microdata) via an SFTP connection. There are safeguards associated with this approach and specified in the data request form,[257] namely:

- The user may not obtain more copies of the data than are need for the objectives of the study;
- The user may not forward the data to third parties unless agreed by Statbel;
- The user may not use the data after the agreed timeline of the research conducted;
- Following the end of the research, the user must delete all copies of the data (or following the completion of the research objectives, whichever comes sooner);
- The user must ensure the data is used only by the personnel of the organisation they belong to;
- The data must remain on the servers of the organisation to which the user belongs; and
- That organisation must detail which security measures it has in place and demonstrate how it will securely handle the microdata.

According to Statbel, there have so far not been any issues or breaches with this approach, i.e. researchers always respect the contract.

In certain, very rare (under once a year) cases, a researcher may request access to microdata where they may be an identification risk (where anonymity of respondents cannot be guaranteed). In these cases, the researcher must use the microdata in Statbel's secure processing environment – i.e. the same approach taken by other statistical offices – and under the constant supervision of a statistician.

---

[257] Statbel, Demande de micro données pour finalités statistiques. See:
https://statbel.fgov.be/sites/default/files/files/documents/Formulaire%20de%20demande%20micro%20donn%C3%A9e s%20(FR).docx

Annex B – Subtask 1.4 One pagers

Digi.me

Intermediary Category	PIMS/PDS
Data Sharing Scenario	C2B
Type of data sharing	Personal Data *(however there is technical capacity also for industrial data)*
Year of Establishment	Founded in 2009 as a legal entity; operational in its current form since 2013
Stage	Growth
Country/-ies of Establishment	United Kingdom (HQ); France, Netherlands; Bosnia; United States; Australia
Profit/Non-Profit Driven	For profit
Number of Employees	50-60
Revenue/turnover	N/A
Funding	$30M Private Sources (including i.a. Swiss RE, Omidyar Network and many notable and other HNWs)
Business Model/Functionalities	Digi.me empowers the individual to share more & better data, to enable businesses to provide more & better value, with full privacy, security and consent. • Individuals benefit from the value in their data • Businesses get a complete view of their customer When using the digi.me app, the individual holds their own data and digi.me does not see, touch or hold user data. • Data encryption and normalization happens inside the app without digi.me ever being able to see or access user data. • Only the user has the credentials to access their digi.me library and must provide credentials directly to data sources. • Digi.me stores no user data. The user chooses their own location (e.g. dropbox, google drive, onedrive) where encrypted data is stored. When businesses/services request data from an individual they use the digi.me consent system and certificate for explicit and informed consent as per GDPR (certificate number goes into the API and the certificate is displayed to the individual). Upon individual's consent to data being transferred to the business/service, then digi.me charges the business/service $0.10 with a cap of $3 per individual per business/service per year to be paid by 'postal fee' on data transfer. Use of the digi.me app by the individual is free as the individual's cost to digi.me is negligible.
Data volume	N/A
Architecture	Decentralized
Client Base/Use Cases	• Horizontal client base (including i.a. health and wellbeing, finance, retail, insurance banking, government and IoT, research) • More than 700,000 users of digi.me app over time in 140 countries and 1,000s of sources of data

DAWEX

Intermediary Category	Data Marketplace & Data Exchange Solution Provider
Data Sharing Scenario	B2B
Type of data sharing	All types of data (industrial data, anonymized data, personal data)
Year of Establishment	2015
Stage	Late
Country/-ies of Establishment	France; US; Canada
Profit/Non-Profit Driven	For profit
Number of Employees	Approximately 50
Revenue/turnover	N/A
Funding	Approximately global funding of €18M from private sources and loans
Business Model/Functionalities	**Data Exchange Platform** technology that facilitates data sharing, data sourcing, data commercialization, data orchestrating, providing in particular the following functionalities: • Facilitate the connectivity between the parties: place where data providers and data users meet and allow cross-border and cross-sector data exchange; automatic matching of supply and demand; • Allow the parties to interact with each other and keep track of these interactions; • For data providers/holders: packaging and describing data offering which can be very diverse (e.g API based, combination of several files, one-time transaction or subscription, description of product and definition of terms and conditions for making data available, licencing terms and pricing) • For data users/buyers: easy search, discovery and filtering certain data products and offering; assessment of offering, negotiation of transaction terms and offering • For Data Exchange Platform orchestrator: fully featured administration console including various automation features for managing participants and stimulate activity, at scale. **Global data marketplace**, built upon Data Exchange Platform technology and providing, among others, the following features and services: • Data monetisation or free-of-charge data exchange • Full control given to data providers and data users over the terms of the exchange • Volume, variety and speed of data exchange • Configurable contract license as well as open data license supported • Data visualization algorithms for quick evaluation and promotion of data quality • Visualization of representative data samples • Recurring transaction services • Integrated payment processing • Administration and traceability services for the data providers an data users • Three subscription plans to access the service (Free, Business, Enterprise) meeting various usage levels and support requirements
Data volume	N/A
Architecture	Flexible architecture allowing users to choose centralized and/or decentralized models for exchanging data depending on the use case
Client Base/Use Cases	• Broad horizontal client base that includes more than 10,000 organizations in more than more than 50 countries and more than 20 industries (incl. i.a. automotive; agriculture; bank, air and space; insurance & financial services; energy; environment; health; banking; telecommunications; retail and consumer goods; public sector; tourism and sports; shipping and logistics) • Customized implementations of Data Exchange Platform technology (white label) target corporates, consortiums and governments (sectorial or regional scope)

International Data Spaces Association

Intermediary Category	Trusted third party providing a reference architecture
Data Sharing Scenario	B2B
Type of data sharing	Industrial data
Year of Establishment	2016
Stage	Growth
Country/-ies of Establishment	Germany
Profit/Non-Profit Driven	Non-profit
Number of Employees	20
Revenue/turnover	N/A
Members	117
Funding	Sources: public and private: Industrial Data Space was created in a research project funded by the German Federal Ministry of Education and Research (BMBF) involving multiple Fraunhofer institutes. It currently works with membership fees.
Business Model/Functionalities	The IDSA reference architecture forms the basis for a variety of certifiable software solutions, smart services and business models. The business model of the IDSA itself is based on an annual membership fee that depends on the size and type of the organisation[258]
Data volume	Not applicable because the International Data Spaces positions itself as an architecture to link different cloud platforms through secure exchange and trusted sharing of data, i.e. through data sovereignty
Architecture	Decentralized
Client Base/Use Cases	Horizontal Client base with 50 use cases

---

[258] https://www.internationaldataspaces.org/wp-content/uploads/2020/06/IDSA-MembershipFeeRegulations-2020.pdf

Inrupt (Solid)

Intermediary Category	Personal Data Store
Data Sharing Scenario	C2B
Type of data sharing	Personal Data
Year of Establishment	2017
Stage	Growth
Country/-ies of Establishment	USA (Boston, Massachusetts); UK (London)
Profit/Non-Profit Driven	Non-profit
Number of Employees	Around 25
Revenue/turnover	N/A
Funding	US$ 6.4M from private sources (including Octopus Ventures; Glasswing Ventures)
Business Model/Functionalities	Business model is currently evolving. Solid is a set of technical agreements that enable an ecosystem of data. Aim of Inrupt is to grow the Solid ecosystem for multiple companies. Based on Solid open-source software built to decentralize the web by organizing data, applications, and identities, while focusing on universality by building on existing web/open standards, Inrupt offers software and services for: a. *Organizations and Developers:* • Design a new breed of applications and get better value from data • Inrupt maintains an open source SDK and other tools, hosts a Pod server (Node Solid Server), an open-source software on inrupt.net in order to support development work, and is working on a suite of tools to help enterprise developers. Inrupt.net is a cloud-hosted instance of the open source software Node Solid Server. This software allows users to establish Pods — virtual places where they can store all kinds of data and choose how to share it with applications or other users. The data on all Solid Pods is organized in a common, machine-readable format called Linked Data. This means any Solid app can read and write data to any Solid Pod, decoupling applications from backends. b. *End users through Solid Pods, can:* • store personal data from multiple sources and choose how to share or use personal data. • control permissions on their data and draw value from an ecosystem of beneficent applications On top of the core Solid specification functionality, the Inrupt Enterprise Solid Server (ESS) also provides a number of additional features, including: • **Enhanced Security** — Better protect your data with advanced Auditing, end-to-end TLS encryption, and OIDC/OAuth Access Control features and support. • **Operational Tooling** — Operate a production system with confidence via native monitoring, distributed logging, backup/restore, and simple integration with industry leading ops platforms. • **SLAs** — ESS's microservices architecture enables simple scaling, high performance, and support for highly available deployment configurations. • **Support** — Inrupt offers 24/7 high SLA support for operators and developers with a commercial license for ESS..
Data volume	N/A
Architecture	Decentralized
Client Base/Use Cases	• Horizontal client base (including health, government, media, finance) • Aim of Solid concept is to make data available to be used in every possible domain • Currently around 20-30 companies having tried and considering the open Solid technology

Meeco

Intermediary Category	N/A *(it could fall under different categories depending on the use case)*
Data Sharing Scenario	Both C2B *(Me2B)* and B2B *(+B2B2C)*
Type of data sharing	Both personal and industrial data (mainly personal)
Year of Establishment	2012
Stage	Growth
Country/-ies of Establishment	Belgium; UK; Australia
Profit/Non-Profit Driven	For profit
Number of Employees	Around 20
Revenue/turnover	N/A *(790% increase within the last 12 months)*
Funding	A$17.2M from Private Sources (including i.a SVX Group, Present Group Developments, A$11.5M Cash, $750K in kind, $5M Assets)
Business Model/Functionalities	Meeco platform provides various functionalities for: A. *Enterprises. Meeco's Privacy by Design tools:* • Enable customers to control their personal data. Deploy our personal data APIs to develop customer centric Privacy by Design applications. Deliver B2B, B2B2C, C2C & Me2B use cases, always with audit and consent. • Meeco API and Consent Engine a) allows 2-way access to data and verified attributes directly with the customer; b) offers full data and attribute control to your customers aligned with new data regulations; c) introduce new services for customers to delegate authority and manage data access; d) Gain customer consent in all data exchange journeys supported by a fully auditable event log • Meeco Labs program provides organisations an opportunity to test hypotheses and prove business value prior to making substantial investments. It is a custom designed process and pathway to new products, services, experiences and business models. B. *Individuals. Manage and share personal data:* • Gain access and store your personal data across your digital life, encrypted and securely available from any device. • Control whom to share data with. If your information changes, update it once and it will be distributed to your connections. • Delegate permission to enable the people and organisations you trust to act on your behalf. • Share data on your terms and maintain a permanent record of your explicit consent. You can change or revoke access at any time. C. *Developers. Developer Portal utilise Meeco's APIs to generate:* • Privacy by Design data store/wallet/vault include SDKs and extensive support documentation • Consent Engine for permissioned access by duration and date • Key Encryption Store to support Zero Value Knowledge D. *Blockchain & distributed ledger. Implement Standards based solutions for:* • Decentralised and Self Sovereign Identity SSID • Verified Credentials/Claims • DID generation and resolution E. *Data Integration partners.* Meeco's product range includes integrations with a broad range of data sources including social, financial health and IoT, including industry leading partners such as: Xero – accounting data; Class Super – superannuation data; Core Logic (RP Data) — property data; Yodlee — banking, transaction and financial data; XPLAN — financial and estate planning; TopDocs — legal data; Suitebox — compliance data; Citrix RightSignature — consent data.
Data volume	N/A
Architecture	It varies to centralized or decentralized model depending on the use case
Client Base/Use Cases	Horizontal client base with use cases in, among others, banking, government, health, retail, airline, financial services, accounting, education, children with special needs, early children development

MIDATA

Intermediary Category	Data Cooperative
Data Sharing Scenario	C2B
Type of data sharing	Personal Data
Year of Establishment	2015
Stage	Growth
Country/-ies of Establishment	Switzerland
Profit/Non-Profit Driven	Non-profit
Number of Employees	3
Revenue/turnover	N/A
Funding	Funding is limited and comes from: • Foundations • Research projects (research grants) • Limited revenues so far, coming from partner organizations (universities, pharmaceutical companies)
Business Model/Functionalities	• MIDATA operates a data platform, acts as a trustee for data collection and guarantees the sovereignty of citizens over the use of their data, showing how data can be used for the common good, while at the same time ensuring the citizens' control over their personal data. • The MIDATA model is designed for international application: MIDATA Switzerland supports the foundation of regional or national MIDATA cooperatives that share the data platform infrastructure. • At present, MIDATA focuses on health data and smartphone app based services. Startups, IT providers and research groups can connect mobile apps to the platform. The apps may offer data-based services and collect data for analysis. • Owners of a data account at MIDATA may actively contribute to medical research and clinical studies by granting selective access to their personal data. Members own and control the cooperative by governing it at the general assembly. Members write the statutes and decide how the profits will be allocated. Financial benefits/incentives for the members are excluded. • Personal data are stored on the MIDATA platform. Data account holders can participate in app-based research projects and benefit from app-based services. • All datasets are encrypted; only data account holders have access to their individual data. Each access to data is logged. To enable global research and clinical studies, secure data access via individual national cooperatives will be implemented, while at the same time maintaining account holders' full control over their personal data.
Data volume	N/A
Architecture	Cloud infrastructure; All datasets of a member are encrypted into one account under the same password
Client Base/Use Cases	• MIDATA currently has 20.000 users of the platform and 60 members • Horizontal approach of use cases, currently focusing on health and education sector • Around 5-10 partners including, among others, Zurich University hospital and Bern University hospital, ETH Zurich and Bern University of Applied Science in Bern, Leiden University Medical Centre

MindSphere (Siemens)

Intermediary Category	Industrial Data Platform
Data Sharing Scenario	B2B
Type of data sharing	Industrial Data
Year of Establishment	2016 (since 2017 in its current version)
Stage	Growth
Country/-ies of Establishment	Germany; available worldwide
Profit/Non-Profit Driven	For profit
Number of Employees	N/A
Revenue/turnover	N/A
Funding	Developed and funded by Siemens
Business Model/Functionalities	Industrial IOT Operating System, an open platform running on a infrastructure as a service solution: • Connect assets and upload data to the cloud; data may come from all kind of devices, other platforms and databases • Collect, monitor, and analyze data in real-time • Gain insights that improve efficiency and profitability • Add apps that increase the business value of the data An Ecosystem for Developers and Makers: • Open environment for development and operations • Ready-to-use APIs and services • Operating on AWS, Azure & Alibaba infrastructures as public cloud solution; also available as private cloud solution • Thriving community of developers and corporate partners (incl. strategic partners, technology partners, connectivity partners either offering new ways of connecting things or help implementing connectivity, partners creating their own applications) In particular, MindSphere platform offers the following functionalities: • For users: Connecting and monitoring assets and systems and performing advanced analytics; • For developers: Developing and delivering industry IoT applications; • For operators: Deploying and monitoring running applications and see what customers are doing; • For sellers: Marketing solutions to a growing, worldwide MindSphere user base.
Data volume	• 1.4 M # of Connected Devices • 262% increase in Connected Assets at Siemens within the last 1 year
Architecture	Cloud based centralized architecture
Client Base/Use Cases	• More than 6.1000 customers • More than 500 ecosystem partners • Horizontal client base covering broad range of domains (e.g.. manufacturing, machine builders, campuses and cities)

Nallian

Intermediary Category	Industrial Data Platform *(not developed by any industry dominant player)*
Data Sharing Scenario	B2B
Type of data sharing	Industrial data *(sometimes personal data might also be involved)*
Year of Establishment	2012
Stage	Growth
Country/-ies of Establishment	Belgium *(+ Asia establishment to be opened soon)*
Profit/Non-Profit Driven	For profit
Number of Employees	23
Revenue/turnover	N/A *(recurring revenue which comes from the usage of the platform has doubled over the past two years)*
Funding	€1.3M from both public (subsidies by Belgian government) and private sources (incl. i.a. Newion)
Business Model/Functionalities	Nallian platform empowers the different stakeholders at logistic hubs to efficiently align, coordinate their cross-company processes and operate as one (one representation of logistics worldwide). The Open Data Sharing Platform underpins a rich ecosystem of collaborative applications that are developed with, for and by air cargo communities. Tailored to the reality of the logistic hub, they: • enable efficient landside management: Empowering ground handlers, freight forwarder and trucking companies to streamline freight pick-up and delivery from A-Z • provide granular insights and analytics: Providing granular levels of insights that help take informed decisions and fuel user's strategy • facilitate regulatory processes: Streamlining planning, communication and data-exchange with regulatory and governmental instances, such as customs or federal food agencies • enable end-to-end track & trace: Gain control and adopt a pro-active approach with end-to-end traceability and visibility on your shipment's journey • allow seamless data sharing across processes, easy access to innovation, without vendor lock-in and short time to market, as well as solidintegration with existing systems and processes. The platform also presents the following features: • Open approach: Working with legacy systems, easy to add (3rd party) apps as the user grows • Data owner in control: The source always stays in control of who sees which part of his data in which context • Flexible & configurable: Starting with the functionality needed, adding another as the user grows • Single version of truth: Avoiding duplicate data entry and manual errors by sharing a single version of truth (data processing and correlation) • Community-led: Community members decide which use case to participate in, allowing gradual adoption
Data volume	N/A
Architecture	It varies to centralized or decentralized model depending on the use case
Client Base/Use Cases	• Horizontal client base [including i.a. logistics hubs: air cargo (airport authorities, ground handling agents, freight forwarders, trucking companies, regulatory institutions); maritime and shippers; consignees] • More than 100% increase of client base number within the last two years • Examples of clients include among others airports in Brussels, Luxembourg, Vienna, London, Dallas (Texas), Asia, ground handlers such as Swissport, WFS and dnata, forwarders such as DHL global forwarding, and recently also airlines as well as actors in the chemical supply chain such as BASF.

Ocean Protocol

Intermediary Category	N/A (Protocol that allows decentralized exchange of data and digital assets)
Data Sharing Scenario	Both B2B and C2B
Type of data sharing	Generic protocol that can be used for both personal and industrial data
Year of Establishment	2017
Stage	Growth
Country/-ies of Establishment	Germany, Berlin (BigChainDB GmbH); Singapore (Ocean Protocol Foundation)
Profit/Non-Profit Driven	Non-profit
Number of Employees	20-25 core team members	Around 40 advisors
Revenue/turnover	N/A
Funding	$28.1M Private sources/individuals (including Kosmos Capital, Fabric Ventures, Outlier Ventures, IOSG Ventures, Zeroth.AI, Julian Sarokin, Synapse Capital)
Business Model/Functionalities	- Ocean protocol is an open source software that acts as an enabler, helping its partners to build and provide data sharing services (i.a. computational services, storage) - Protocol designing technology that allows decentralized exchange of data and digital assets and helps developers build **marketplaces** and other apps to privately and securely publish, exchange, and consume data. Using Ocean software components, connected to the decentralized Ocean data sharing network: - Data providers can monetize data while preserving privacy and control; - Data consumers can access private data that they could not get before. - As data assets are exchanged, **blockchain technology** provides the security, privacy, and control benefits of Ocean Protocol and makes sellers and buyers benefit from the auditability of purchase transactions. (Tokenized environment) **Web3 Access Control** (blockchain-enabled access control) allows maintaining control and granting access over the data set, without a centralized intermediary. Each dataset registered in an Ocean Protocol marketplace has a Web3 account as owner attached to it, based on the account used to publish the data set. Only that owner can modify or transfer ownership for the data set. Likewise, only Web3 accounts can consume your data, once they have been granted access to it. **Compute-to-data** resolves the tradeoff between the benefits of using private data, and the risks of exposing it. It lets the data stay on-premise, yet allows 3rd parties to run specific compute jobs on it to get useful compute results like averaging or building an AI model. - Ocean Protocol's marketplaces and Compute-to-Data help data scientists & AI practitioners get more data, including private data. The non-profit Ocean Protocol Foundation commissioned BigChainDB GmbH to design and develop the core architecture and components of the existing Ocean Protocol software. Both Ocean Protocol Foundation and BigChainDB are actively looking for partnerships to help drive adoption and grow the open source ecosystem engaged in implementing the software for real life data exchange use cases.
Data transaction volume	N/A
Architecture	Decentralized
Client Base/Use Cases	- Horizontal client base mainly in the industries of automotive, logistics and healthcare, ranging from small SMEs to large MNCs - Relatively small client base number at the moment

Polypoly

Intermediary Category	PIMS; Data cooperative (Foundation of three companies including one cooperative)
Data Sharing Scenario	C2B
Type of data sharing	Personal Data
Year of Establishment	2019
Stage	Early
Country/-ies of Establishment	Liechtenstein; Germany (Berlin); South Africa
Profit/Non-Profit Driven	Polypoly Gmbh and Polypoly Cooperative SCE for profit / Polypoly Foundation non-profit
Number of Employees	Around 45
Revenue/turnover	Recently closed deals with three blue chips with an estimated 250.000€ MRR
Funding	Around 20M EUR, from both public and private sources with three pillars: • entrepreneurs/industrial crowdfunding • public sources (including both national and EU funding) • citizens' donations/memberships

Business Model/Functionalities	Polypoly software is not acting in the name of a 3rd party/user, as an intermediate would do – the software is running on the device of the user, providing them the capacity to act on their own. Aim of this software is to make GDPR rights executable for its users. Hosting of data is free of charge for both companies and individuals. Application called polyPod, offering services for:
	   a. Individuals
	• Automation of the process of retrieving data from data providers (i.e Facebook, Amazon, WhatsApp, Google) based on GDPR provisions; finding and managing personal user data available on the Internet and storing it securely on their own electronic devices;
	• Gain financial benefit from data in an anonymous way;
	• No collection nor selling of the users's data by Polypoly.
	   b. Organisations (companies, governmental organizations and NGOs)
	• personalised services using the data stored in individual polyPods (by giving data back to the users), all with the user's consent and without it leaving the user's device; All computations are being done on the end-users device and nobody is needed to enforce end-users' rights;
	• access to better and cheaper data sets, while upholding GDPR compliance, and reducing costs related to storing data and keeping them up to date, IT and cybersecurity, research, product development and improvement, GDPR compliance costs;
	• Aim for auditing, contract manufacturing, training, and big data services in the future.
	A European Cooperative (Polypoly SCE) has recently been set up:
	• Owned, driven and controlled by the users/members, by installing the software with an optional charge of 5 EUR; No legal body can buy a share of the cooperative
	• Changes in the fundamental element of the software require the agreement of the users; 1 vote corresponds to 1 user no matter how many shares they own..

Data volume	N/A
Architecture	Decentralized
Client Base/Use Cases	Polypoly GmbH: • Horizontal Use Cases (including i.a banks, insurance companies, mobility providers) Cooperative Polypoly SCE: • No users yet (planned go live date in early 2021) • Aim for horizontal approach of use cases (including i.a.neutral payment infrastructure, banks, messaging infrastructure, federated AI) Polypoly Foundation • Building Data Coops outside of Europe (Franchise model), running negations with India, South Africa, Canada, Switzerland, UK and the USA • Running polyPedia: open source database about the data behavior of several thousand companies.

Streamr

Intermediary Category	Data Union; Data Marketplace
Data Sharing Approach	Both C2B and B2B (mainly C2B)
Type of data sharing	Both personal and industrial (mainly personal)
Year of Establishment	Founded in 2015; operational since 2017
Stage	Growth
Country/-ies of Establishment	Finland; Switzerland
Profit/Non-Profit Driven	For profit
Number of Employees	Approximately 35
Revenue/turnover	N/A
Funding	27.7M EUR in 2017 Crowdfunded; both private and public sources (incl. Firestartr (Lead Investor); EIT Digital Accelerator; Fabric Ventures ; Andreas Schwartz)

Business Model/Functionalities	Provision of three sets of (neutral) functions rolled into one:
	• the ability to transport data from one platform to the end buyer of the data
	• the marketplace with two functions:
	  o Discovery of the data
	  o Aggregation of individuals' data within the data union framework into one product
	• Micropayments (for the payment of data union members)

Data volume	• 1000s messages/second through the Streamr Network • Financial value of data transactions: "small" transactions at the moment; much larger ones expected when unique data sets (e.g Swash) are viable and sold within a six-month to a year timeframe
Architecture	Decentralized peer to peer network
Client Base/Use Cases	Horizontal client base (incl. i.a. environment, IoT, transportation, energy, health, retail) • Swash (1,00+ members) • Tracey app (WWF Philippines - UnionBank - TX partnership) • 15 business partners (incl. Hewlett Packard Enterprise, WWF and DXC Technology, Bosch, Fastems, and Capita)

Smart Connected Network (SCSN)

Intermediary Category	Trusted third party that enables a fast, secure and interoperable exchange of information across company borders based on an open standard (called the SCSN standard) that builds on the IDSA (International Data Spaces Association) standard.
Data Sharing Scenario	B2B
Type of data sharing	Industrial data
Year of Establishment	The SCSN initiative started in 2015 as a project called "Connections in the chain" and partners cooperate in a field lab to develop the standard. The foundation SCSN is established in 2020.
Stage	Growth
Country/-ies of Establishment	The Netherlands
Profit/Non-Profit Driven	Non-profit
Number of Employees	Various partners contribute with FTE, so about 15 FTE are involved. However, the foundation does not have employees only a board and a supervisory board
Revenue/turnover	The foundation is expected to cost 200k (e.g. to maintain the standard). Potentially part of these costs will be covered by in-kind contributionsl via branche associations
Funding	Currently public-private funding
Business Model/Functionalities	• They will work with a participation fee in the future to maintain the data exchange standard. Some within SCSN involved service providers will also use a pay-per use fee
Data volume	• Not applicable because SCSN enables the data exchange based on a standard that is building on the IDSA. SCSN does not store the data of the users.
Architecture	Decentralized
Client Base/Use Cases	Horizontal client base (but a main focus on Manufacturing) The SCSN network involve the following partners: • OEMs first-tier suppliers, smaller second and third-tier suppliers, but also (steel) wholesalers and steel manufacturers are connected via the SCSN standard. • Service providers are IT partners who facilitate the connection to the SCSN network for manufacturing companies. Service Providers have set up various standard connections with a diverse portfolio of IT systems, so that they can easily connect manufacturing companies to the SCSN network. The idea is that a manufacturer only needs to be connected once to get access the whole supply network. • 9 service providers are currently involved in the SCSN network. These 9 service providers can be defined as data intermediaries. Manufacturing companies can chose the service provider they prefer. These are: *Fujitsu Glovia*, *Supplydrive*, *Tradecloud*, *ISAH Business Software*, *Trivest Connect*, *Ketenlink*, *Exact, Easy2Trade (INAD)* , *Attributes* • Almost 300 manufactures are currently connected via SCSN. The goal is to grow to 1000 connected partners in the coming year. • There is an onboarding program for both new manufacturers as well as new service providers.)

Annex C – Cost-Benefit Analysis

Subtask 1.1:

The figure below presents the input summary for the cost-benefit analysis for sub-task 1.1

Input & Summary

Input	Unit	Value	Source/estimate
Data authorities			
CAPEX			
PO3 (central a.b. Findata)	EUR	10 500 000	Findata
PO2 (one-stop) %-of Findata	% of total	50%	Estimate based on Findata costs
PO1			
OPEX			
PO3			
Running costs (est. for after 2023)	EUR p.a.	5 037 000	Findata
Budget (est. for after 2023)	EUR p.a.	1 000 000	Findata
Training	EUR p.a.	8 395	Findata
PO2			
Running costs (excluding data processing environment	EUR p.a.	900 000	RatSWD (Germany)
Secure data processing environment	EUR p.a.	610 000	Estimate based on Statistics Denmark and Epiconcept
Revenues/fees PO2	EUR/application	250	Assumption
Revenues/fees PO3	EUR/application	2 546	Estimate based on Findata costs
Data holders			
OPEX			
PO2 - Coordinating and liaising with the one-stop shop	EUR p.a.	5 400	Assumption
Benefits/cost savings			
PO3 - Time/resources resulting from not processing data access applications	EUR p.a.	405 000,0	Statistics Denmark
PO2 & PO3 - Time/resouces resulting from not pre-processing & providing data	EUR p.a.	1 215 000,0	Statistics Denmark
PO2 & PO3 - Average amount of pre-processing/providing data work saved by data holders	% of total	30%	Assumption
PO2 & PO3 - Secure data processing environment	EUR p.a.	610 000	Estimate based on Stat. Denmark and Epiconcept
PO2 & PO3 - Average no. of data holders abolishing infrastructure	% of total	20%	Assumption
Data re-users			
Costs PO3 - Estimates based on Findata 600 applications	EUR/application	2 546	Estimate
Benefits PO3 - Time/resources saved not having to submit separate applications	EUR/application	2 000	Stakeholder information
Benefits PO2 - Time/resources saved not having to search for data holder	EUR/application	590	Assumption
General			
Social Discount Rate	%	3%	

Results

Benefits/Costs PO3 - Total	PO3	Benefits	Costs	NPV	B/C-ratio
Data authorisation body	PO3	945,8	(2 022,0)	(1 076,2)	0,5
Data holders	PO3	5 573,2	-	5 573,2	n/a
Data re-users	PO3	742,8	(945,8)	(203,0)	0,8
Total	PO3	7 261,9	(2 967,8)	4 294,1	2,4

Benefits/Costs PO2	PO2	Benefits	Costs	NPV	B/C-ratio
Data authorisation body	PO2	185,7	(351,2)	(165,5)	0,5
Data holders	PO2	3 041,4	(33,8)	3 007,6	90,1
Data re-users	PO2	219,1	(185,7)	33,4	1,2
Total	PO2	3 446,2	(570,7)	2 875,5	6,0

The figure below presents the cost-benefit analysis for sub-task 1.1.

Cost-Benefit Analysis

Total	M€ (constant prices)	PO	MS	Stakeholder	Category	Subcategory	NPV @ 3%	Total	2023	2024	2025	2026	2027	2028
Total	**Data authorisation body**	PO3	Total	Data authorisation body	Costs	CAPEX	(556.0)	(572.7)	(572.7)	-	-	-	-	-
		PO3	Total	Data authorisation body	Costs	OPEX	(1,466.0)	(1,648.6)	-	(329.7)	(329.7)	(329.7)	(329.7)	(329.7)
		PO3	Total	Data authorisation body	Benefits	REVENUES	945.8	1,063.6	-	212.7	212.7	212.7	212.7	212.7
	Data holders	PO3	Total	Data holders	Costs	OPEX	-	-	-	-	-	-	-	-
		PO3	Total	Data holders	Benefits	OPEX savings	5,573.2	6,267.2	-	1,253.4	1,253.4	1,253.4	1,253.4	1,253.4
	Data re-users	PO3	Total	Data re-users	Costs	OPEX	(945.8)	(1,063.6)	-	(212.7)	(212.7)	(212.7)	(212.7)	(212.7)
		PO3	Total	Data re-users	Benefits	OPEX savings	742.8	835.3	-	167.1	167.1	167.1	167.1	167.1
	Costs total	PO3	Total	Total	Costs	Costs total	**(2,967.8)**	**(3,284.8)**	**(572.7)**	**(542.4)**	**(542.4)**	**(542.4)**	**(542.4)**	**(542.4)**
	Benefits total	PO3	Total	Total	Benefits	Benefits total	**7,261.9**	**8,166.2**	**-**	**1,633.2**	**1,633.2**	**1,633.2**	**1,633.2**	**1,633.2**
	Net Cashflow NPV	PO3	Total	Net Cashflow NPV	NPV	NPV	**4,294.1**	**4,881.4**	**(572.7)**	**1,090.8**	**1,090.8**	**1,090.8**	**1,090.8**	**1,090.8**
	Benefit/Cost-ratio	PO3	Total	Benefit/Cost-ratio	BCR	BCR	**2.4**							
Total	**One-stop shop**	PO2	Total	Data authorisation body	Costs	CAPEX	(278.0)	(286.3)	(286.3)	-	-	-	-	-
		PO2	Total	Data authorisation body	Costs	OPEX	(73.2)	(82.4)	-	(16.5)	(16.5)	(16.5)	(16.5)	(16.5)
		PO2	Total	Data authorisation body	Benefits	REVENUES	185.7	208.8	-	41.8	41.8	41.8	41.8	41.8
	Data holders	PO2	Total	Data holders	Costs	OPEX	(33.8)	(38.0)	-	(7.6)	(7.6)	(7.6)	(7.6)	(7.6)
		PO2	Total	Data holders	Benefits	OPEX savings	3,041.4	3,420.1	-	684.0	684.0	684.0	684.0	684.0
	Data re-users	PO2	Total	Data re-users	Costs	OPEX	(185.7)	(208.8)	-	(41.8)	(41.8)	(41.8)	(41.8)	(41.8)
		PO2	Total	Data re-users	Benefits	OPEX savings	219.1	246.4	-	49.3	49.3	49.3	49.3	49.3
	Costs total	PO2	Total	Total	Costs total	Costs total	**(351.2)**	**(368.7)**	**(286.3)**	**(16.5)**	**(16.5)**	**(16.5)**	**(16.5)**	**(16.5)**
	Benefits total	PO2	Total	Total	Benefits total	Benefits total	**3,226.7**	**3,628.6**	**-**	**725.7**	**725.7**	**725.7**	**725.7**	**725.7**
	Net Cashflow NPV	PO2	Total	Net Cashflow NPV	NPV	NPV	**2,875.5**	**3,259.9**	**(286.3)**	**709.2**	**709.2**	**709.2**	**709.2**	**709.2**
	Benefit/Cost-ratio	PO2	Total	Benefit/Cost-ratio	BCR	BCR	**9.2**							

Subtask 1.2:

The figure below presents the input summary for the cost-benefit analysis for sub-task 1.2.

Input & Summary

Input	Unit	Value	Source/estimate
Number of stakeholders			
data holders (citizens)	total no. EU27 in 2023	5 000 000	Estimate based on expert interviews and desk research
data holders (businesses), Ass.: 10% of 5000 businesses, task 1.4	total no. EU27 in 2023	500	Estimate based on expert interviews and desk research
data intermediaries (assumption: 1 per MS)	total no. EU27 in 2023	27	Estimate based on expert interviews and desk research
data re-users	total no. EU27 in 2023	15 000	Estimate based on expert interviews and desk research
Stakeholders affected (total)			
PO1	total no. EU27 in 2023	5 015 515	Assumption based on expert research and expert interviews
PO2	total no. EU27 in 2023	5 015 515	Assumption based on expert research and expert interviews
PO3	total no. EU27 in 2023	5 015 515	Assumption based on expert research and expert interviews
Benefits affected stakeholders (cost savings/efficiency gains)			
PO1	% of OPEX p.a.	-	Assumption based on expert interviews and expert research
PO2	% of OPEX p.a.	5-50% and qualitative	Assumption based on expert interviews and expert research
PO3	% of OPEX p.a.	merely qualitative	Assumption based on expert interviews and expert research
OPEX per company on average for 5yrs			
PO1	OPEX total 2024-2028 EUR	-	Assumption based on expert research and qualitative survey
PO2	OPEX total 2024-2028 EUR	100 000	Assumption based on expert research and qualitative survey
PO3	OPEX total 2024-2028 EUR	25 000	Assumption based on expert research and qualitative survey
Costs (implementation of PO for data intermediaries)			
PO1	Implementation (2023) in EUR	-	Assumption based on expert research and qualitative survey
PO2	Implementation (2023) in EUR	30 000	Assumption based on expert research and qualitative survey
PO3 - Companies	Implementation (2023) in EUR	10 500	Assumption based on expert research and qualitative survey
PO3 - NGOs (10% discount)		9 450	
Social Discount Rate	%	3%	CBA Guide

Results M€

Benefits/Costs PO1 - Tot PO1	Benefits	Costs	NPV	BCR

Benefits/Costs PO2 -	PO2	Benefits	Costs	NPV	BCR
Data intermediaries	PO2	28,4	(10,4)	18,1	2,7
Data holders	PO2	-	-	-	n/a
Total	PO2	28,4	(10,4)	18,1	2,7

Benefits/Costs PO3 - Da PO3	PO3	Benefits	Costs	NPV	BCR
Data intermediaries	PO3	-	(29)	(29)	-
Data holders	PO3	-	-	-	n/a
Total	PO3	265	(42)	224	6,3

The figure below presents the cost-benefit analysis for sub-task 1.2.

Total	€ (constant prices)	PO	MS	Stakeholder	Category	Subcategory	NPV @ 3%	Total	2023	2024	2025	2026	2027	2028
Total	**Benefits**	PO2	Total	Data intermediaries	no.	total intermediaries	n/a	n/a	1 255	1 255	1 255	1 255	1 255	1 255
		PO2	Total	Data intermediaries	no.	participating in the voluntary scheme	125		11	125	125	125	125	125
		PO2	Total	Data intermediaries	efficiency gains %	OPEX	n/a	n/a	-	10%	10%	10%	10%	10%
		PO2	Total	Data intermediaries	revenues	total additional revenues	8 966 675	10 220 000	-	220 000	2 500 000	2 500 000	2 500 000	2 500 000
		PO2	Total	Data intermediaries	add.revenues	aver. revenues per intermediary	88 926	100 000	-	20 000	20 000	20 000	20 000	20 000
		PO2	Total	Data re-users	no.	no	n/a	n/a	15 000	15 000	15 000	15 000	15 000	15 000
		PO2	Total	Data re-users	Benefits	total value of data**	19 460 297	22 002 250	-	3 500 350	4 000 400	4 500 450	5 000 500	5 000 550
		PO2	Total	Data re-users	Benefits	average value per unit of data	46	50		10	10	10	10	10
		PO2	Total	Data holders	no.	total holders (Citizens)	n/a	n/a	5 000 000	5 000 000	5 000 000	5 000 000	5 000 000	5 000 000
		PO2	Total	Data holders	no.	participating in the voluntary scheme	n/a	n/a	300 000	350 000	400 000	450 000	500 000	500 000
		PO2	Total	Data holders	no.	total holders (Companies)	n/a	n/a	500	500	500	500	500	500
			Total	Data holders	no.	participating in the voluntary scheme	n/a	n/a	30	35	40	45	50	55
		PO2	Total	Data intermediaries	Client base	number of clients %			-	40%	10%	10%	10%	10%
		PO2	Total	Data intermediaries	Client base	number of use cases			-	35%	10%	10%	10%	10%
		PO2	Total	Data intermediaries	Increased competition	B2B Market			-	-	3%	3%	3%	3%
		PO2	Total	Data intermediaries	Increased competition	C2B Market			-	10%	5%	5%	5%	5%
		PO2	Total	Data intermediaries	Benefits	Benefits	-							
	Costs	PO2	Total	Data re-users	Costs	Maintenance costs*	n/a	n/a	-	-	-	-	-	-
		PO2	Total	Data intermediaries	Costs	Compliance one-off for obtaining the certification for the first time	(3 640 777)	(3 750 000)	(3 750 000)	-	-	-	-	-
		PO2	Total	Data intermediaries	Costs	Compliance recurrent costs of renewing certification	(6 725 007)	(7 665 000)	-	(165 000)	(1 875 000)	(1 875 000)	(1 875 000)	(1 875 000)
		PO2	Total	Data intermediaries	Reduced Competition	B2B Market	-25%							
	Costs total	PO2	Total	Total	Costs	Costs total	(10 365 783)	(11 415 000)	(3 750 000)	(165 000)	(1 875 000)	(1 875 000)	(1 875 000)	(1 875 000)
	Benefits total	PO2	Total	Total	Benefits	Benefits total	28 426 972	10 220 000	-	220 000	2 500 000	2 500 000	2 500 000	2 500 000
	Net Cashflow NPV	PO2	Total	Net Cashflow NPV	NPV	NPV	(1 399 108)	(1 195 000)	(3 750 000)	55 000	625 000	625 000	625 000	625 000
	Benefit/Cost-ratio	PO2	**Total**	**Benefit/Cost-ratio**	**BCR**	**BCR**	**2,7**							

Total	€ (constant prices)	PO	MS	Stakeholder	Category	Subcategory	NPV @ 3%	Total	2023	2024	2025	2026	2027	2028
Total	**Benefits**	PO3	Total	Data intermediaries	no.	Companies	n/a	n/a	1 255	1 255	1 255	1 255	1 320	1 350
		PO3	Total	Data intermediaries	no.	NGOs			55	55	55	60	65	75
		PO3	Total	Data intermediaries	efficiency gains %	OPEX	n/a	n/a	10%	10%	10%	10%	10%	10%
		PO3	Total	Data re-users	no.	no.	n/a	n/a	15 000	15 000	15 500	15 750	16 000	16 500
		PO3	Total	Data re-users	Benefits	total value of data**	265 491 825	300 030 000	-	50 005 000	55 005 500	60 006 000	65 006 500	70 007 000
		PO3	Total	Data re-users	Benefits	average value per unit of data	44	50	-	10,00	10,00	10,00	10,00	10,00
		PO3	Total	Data holders*	no.	Citizens	n/a	n/a	5 000 000	5 000 000	5 500 000	6 000 000	6 500 000	7 000 000
		PO3	Total	Data holders*	no.	Companies	n/a	n/a	500	500	550	600	650	700
		PO3	Total	Data holders	Benefits	Qualitative Benefit: increased trust	n/a	n/a						
	Costs	PO3	Total	Data re-users	Costs	Maintenance costs*	n/a	n/a	[already included in task 1.1]					
		PO3	Total	Data intermediaries	Companies	One-off authorisation costs per company	(10 194)	(10 500)	(10 500)	-	-	-	-	-
		PO3	Total	Data intermediaries	Companies	total one-off authorisation costs	(12 793 689)	(13 177 500)	(13 177 500)					
		PO3	Total	Data intermediaries	NGOs	One-off authorisation costs per NGO	(9 175)	(9 450)	(9 450)	-	-	-	-	-
		PO3	Total	Data intermediaries	NGOs	total one-off authorisation costs	(504 612)	(519 750)	(519 750)					
		PO3	Total	Data intermediaries	Costs	recurrent costs/mainentance costs per intermediary	(22 899)	(25 000)		(5 000)	(5 000)	(5 000)	(5 000)	(5 000)
		PO3	Total	Data intermediaries	Costs	total recurrent costs/mainentance costs	(28 578 796)	(32 175 000)	-	(6 275 000)	(6 275 000)	(6 275 000)	(6 600 000)	(6 750 000)
	Costs total	PO3	Total	Total	Costs	Costs total	(41 877 097)	(45 872 250)	(13 697 250)	(6 275 000)	(6 275 000)	(6 275 000)	(6 600 000)	(6 750 000)
	Benefits total	PO3	Total	Total	Benefits	Benefits total	265 491 825	300 030 000	-	50 005 000	55 005 500	60 006 000	65 006 500	70 007 000
	Net Cashflow NPV	PO3	Total	Net Cashflow NPV	NPV	NPV	223 614 728	254 157 750	(13 697 250)	43 730 000	48 730 500	53 731 000	58 406 500	63 257 000
	Benefit/Cost-ratio	PO3	**Total**	**Benefit/Cost-ratio**	**BCR**	**BCR**	**6,3**							

Subtask 1.3:

The figure below presents the input summary for the cost-benefit analysis for sub-task 1.3.

Input & Summary

Input	Unit	Value	Source/estimate
Number of stakeholders			
data re-users + data holders	total no. EU27 in 2023	700 000	Estimate based in EU Data Monitoring Tool
data intermediaries	total no. EU27 in 2023	100	Estimate
other (data companies)	total no. EU27 in 2023	280 000	Estimate based in EU Data Monitoring Tool
Stakeholders affected (re-user&holders)			
PO1	total no. EU27 in 2023	700	Assumption based on expert research
PO2	total no. EU27 in 2024	800	Assumption based on expert research
PO3	total no. EU27 in 2025	900	Assumption based on expert research
Benefits affected stakeholders (cost savings/efficiency gains)			
PO1	% of OPEX p.a.	15%	Assumption based on IDS and expert research
PO2	% of OPEX p.a.	15%	Assumption based on IDS and expert research
PO3	% of OPEX p.a.	15%	Assumption based on IDS and expert research
OPEX per company on average for 5yrs			
PO1	OPEX total 2024-2028 EUR	50 000 000	Assumption based on expert research
PO2	OPEX total 2024-2028 EUR	45 000 000	Assumption based on expert research
PO3	OPEX total 2024-2028 EUR	40 000 000	Assumption based on expert research
Costs (implementation of PO)			
PO1	Implementation (2023) in EUR	24 000	Assumption based on expert research
PO2	Implementation (2023) in EUR	280 000	Assumption based on expert research
PO3	Implementation (2023) in EUR	3 500 000	Assumption based on expert research
General			
Social Discount Rate	%	3%	

Results

Benefits/Costs PO1 - Total	PO1	Benefits	Costs	NPV	BCR
European Commission	PO1	-	(0,0)	(0,0)	-
Data re-user/holders	PO1	4 668,6	-	4 668,6	n/a
Total	PO1	4 668,6	(0,0)	4 668,6	200 362,2

Benefits/Costs PO2 - Data re	PO2	Benefits	Costs	NPV	BCR
European Commission	PO2	-	(0,3)	(0,3)	-
Data re-user/holders	PO2	5 335,6	-	5 335,6	n/a
Total	PO2	5 335,6	(0,3)	5 335,3	19 627,3

Benefits/Costs PO3 - Data re	PO3	Benefits	Costs	NPV	BCR
European Commission	PO3	-	(3,4)	(3,4)	-
Data re-user/holders	PO3	6 002,5	-	6 002,5	n/a
Total	PO3	6 002,5	(3,4)	5 999,1	1 766,5

The figure below presents the cost-benefit analysis for sub-task 1.3.

Cost-Benefit Analysis

Total	M€ (constant prices)	PO	MS	Stakeholder	Category	Subcategory	NPV @ 3%	Total	2023	2024	2025	2026	2027	2028
Total	**Benefits**	PO1	Total	Data re-user/holders	no.	no.	n/a	n/a	-	700	700	700	700	700
		PO1	Total	Data re-user/holders	efficiency gains %	efficiency gains %	n/a	n/a	-	*15%*	*15%*	*15%*	*15%*	*15%*
		PO1	Total	Data re-user/holders	OPEX	OEPX	44	50	-	10	10	10	10	10
		PO1	Total	Data re-user/holders	Benefits	Benefits	4 668,6	5 250,0	-	1 050,0	1 050,0	1 050,0	1 050,0	1 050,0
	Costs	PO1	Total	European Commission	Costs	Implementation	(0,0)	(0,0)	(0,0)	-	-	-	-	-
	Costs total	PO1	Total	Total	Costs	Costs total	(0,0)	(0,0)	(0,0)	-	-	-	-	-
	Benefits total	PO1	Total	Total	Benefits	Benefits total	4 668,6	5 250,0	-	1 050,0	1 050,0	1 050,0	1 050,0	1 050,0
	Net Cashflow NPV	PO1	Total	Net Cashflow NPV	NPV	NPV	4 668,6	5 250,0	(0,0)	1 050,0	1 050,0	1 050,0	1 050,0	1 050,0
	Benefit/Cost-ratio	PO1	Total	Benefit/Cost-ratio	BCR	BCR	200 362,2							
Total	**Benefits**	PO2	Total	Data re-user/holders	no.	no.	n/a	n/a	-	800	800	800	800	800
		PO2	Total	Data re-user/holders	efficiency gains %	efficiency gains %	n/a	n/a	-	*15%*	*15%*	*15%*	*15%*	*15%*
		PO2	Total	Data re-user/holders	OPEX	OEPX	44	50	-	10	10	10	10	10
		PO2	Total	Data re-user/holders	Benefits	Benefits	5 335,6	6 000,0	-	1 200,0	1 200,0	1 200,0	1 200,0	1 200,0
	Costs	PO2	Total	European Commission	Costs	Implementation	(0,3)	(0,3)	(0,3)	-	-	-	-	-
	Costs total	PO2	Total	Total	Costs	Costs total	(0,3)	(0,3)	(0,3)	-	-	-	-	-
	Benefits total	PO2	Total	Total	Benefits	Benefits total	5 335,6	6 000,0	-	1 200,0	1 200,0	1 200,0	1 200,0	1 200,0
	Net Cashflow NPV	PO2	Total	Net Cashflow NPV	NPV	NPV	5 335,3	5 999,7	(0,3)	1 200,0	1 200,0	1 200,0	1 200,0	1 200,0
	Benefit/Cost-ratio	PO2	Total	Benefit/Cost-ratio	BCR	BCR	19 627,3							
Total	**Benefits**	PO3	Total	Data re-user/holders	no.	no.	n/a	n/a	-	900	900	900	900	900
		PO3	Total	Data re-user/holders	efficiency gains %	efficiency gains %	n/a	n/a	-	*15%*	*15%*	*15%*	*15%*	*15%*
		PO3	Total	Data re-user/holders	OPEX	OEPX	44	50	-	10	10	10	10	10
		PO3	Total	Data re-user/holders	Benefits	Benefits	6 002,5	6 750,0	-	1 350,0	1 350,0	1 350,0	1 350,0	1 350,0
	Costs	PO3	Total	European Commission	Costs	Implementation	(3,4)	(3,5)	(3,5)	-	-	-	-	-
	Costs total	PO3	Total	Total	Costs	Costs total	(3,4)	(3,5)	(3,5)	-	-	-	-	-
	Benefits total	PO3	Total	Total	Benefits	Benefits total	6 002,5	6 750,0	-	1 350,0	1 350,0	1 350,0	1 350,0	1 350,0
	Net Cashflow NPV	PO3	Total	Net Cashflow NPV	NPV	NPV	5 999,1	6 746,5	(3,5)	1 350,0	1 350,0	1 350,0	1 350,0	1 350,0
	Benefit/Cost-ratio	PO3	Total	Benefit/Cost-ratio	BCR	BCR	1 766,5							

Subtask 1.4:

The figure below presents the input summary for the cost-benefit analysis for sub-task 1.4.

Input & Summary

Input	Unit	Value	Source/estimate
Number of stakeholders			
data holders (citizens)	total no. EU27 in 2023	10000-5000000	Estimate based on expert interviews and desk research
data holders (businesses)	total no. EU27 in 2023	500-250000	Estimate based on expert interviews and desk research
data intermediaries (100 in the C2B, 50 in the B2B market)	total no. EU27 in 2023	150	Estimate based on expert interviews and desk research
data re-users	total no. EU27 in 2023	############	Estimate based on expert interviews and desk research
Stakeholders affected (intermediaries)			
PO1	total no. EU27 in 2023 and in 2028	150 and 165	Assumption based on expert research and expert interviews
PO2	total no. EU27 in 2023 and in 2028	150 and 180	Assumption based on expert research and expert interviews
PO3	total no. EU27 in 2023 and in 2028	150 and 210	Assumption based on expert research and expert interviews
Benefits affected stakeholders (cost savings/efficiency gains)			
PO1	% of OPEX p.a.	5%	Assumption based on expert interviews and expert research
PO2	% of OPEX p.a.	7,5%	Assumption based on expert interviews and expert research
PO3	% of OPEX p.a.	10%	Assumption based on expert interviews and expert research
OPEX per company on average for 5yrs			
PO1	OPEX total 2024-2028 EUR	62 500	Assumption based on expert research and qualitative survey
PO2	OPEX total 2024-2028 EUR	125 000	Assumption based on expert research and qualitative survey
PO3	OPEX total 2024-2028 EUR	200 000	Assumption based on expert research and qualitative survey
Costs (implementation of PO)			
PO1	Implementation (2023) in EUR	15 000	Assumption based on expert research and qualitative survey
PO2	Implementation (2023) in EUR	35 000	Assumption based on expert research and qualitative survey
PO3	Implementation (2023) in EUR	40 000	Assumption based on expert research and qualitative survey
Social Discount Rate	%	3%	CBA Guide

Results M€

Benefits/Costs PO1 - Total	PO1	Benefits	Costs	NPV	BCR
Data intermediaries	PO1	46,8	(9,0)	37,8	5,2

Benefits/Costs PO2 -	PO2	Benefits	Costs	NPV	BCR
Data intermediaries	PO2	65,6	(24,5)	41,1	2,7

Benefits/Costs PO3 -	PO3	Benefits	Costs	NPV	BCR
Data intermediaries	PO3	65,6	(24,5)	41,1	2,7

The figures below presents the cost-benefit analysis for sub-task 1.4.

Cost-Benefit Analysis

€ (constant prices)	PO	MS	Stakeholder	Category	Subcategory	NPV @ 3%	Total	2023	2024	2025	2026	2027	2028
Benefits	PO1	Total	Data intermediaries	no.	no.	n/a	n/a	150	113	113	113	113	165
	PO1	Total	Data intermediaries	efficiency gains %	OPEX	n/a	n/a	-	5%	5%	5%	5%	5%
	PO1	Total	Data intermediaries	revenues	total revenues	46 827 053	50 000 000	-	25 000 000	6 250 000	6 250 000	6 250 000	6 250 000
	PO1	Total	Data intermediaries	revenues	revenues per intermediary	400 992	426 768	-	222 222	55 556	55 556	55 556	37 879
	PO1	Total	Data intermediaries					-	25%	30%	35%	35%	25%
	PO1	Total	Data intermediaries	Client base	number of clients			-	25%	15%	8%	8%	8%
	PO1	Total	Data intermediaries	Client base	number of use cases			-	25%	15%	8%	8%	8%
	PO1	Total	Data intermediaries	Benefits	Benefits				1,4	1,7	2,0	2,0	2,1
	PO1	total	Data intermediaries	Competitiveness	Increased Competition							20,0%	
Costs	PO1	Total	Data intermediaries	Costs	Compliance one-off for	(2 184 466)	(2 250 000)	(2 250 000)	-	-	-	-	-
	PO1	Total	Data intermediaries	Costs	Compliance recurrent	(6 802 233)	(7 687 500)	-	(1 406 250)	(1 406 250)	(1 406 250)	(1 406 250)	(2 062 500)
	PO1	Total	Data intermediaries	Costs	Reduced Competition %		-25,00%						
Costs total	PO1	Total	Total	Costs	Costs total	(8 986 699)		(2 250 000,0)	(1 406 250,0)	(1 406 250,0)	(1 406 250,0)	(1 406 250,0)	(2 062 500,0)
Benefits total	PO1	Total	Total	Benefits	Benefits total	46 827 053		-	25 000 000,0	6 250 000,0	6 250 000,0	6 250 000,0	6 250 000,0
Net Cashflow NPV	PO1	Total	Net Cashflow NPV	NPV	NPV	36 476 459,4		(2 250 000,0)	23 593 750,0	4 843 750,0	4 843 750,0	4 843 750,0	4 187 500,0
Benefit/Cost-ratio	PO1	Total	Benefit/Cost-ratio	BCR	BCR	5,2							

€ (constant prices)	PO	MS	Stakeholder	Category	Subcategory	NPV @ 3%	Total	2023	2024	2025	2026	2027	2028
Benefits	PO2	Total	Data intermediaries	no.	no.	n/a	n/a	150	138	138	138	138	180
	PO2	Total	Data intermediaries	efficiency gains %	OPEX	n/a	n/a	-	8%	8%	10%	10%	10%
	PO2	Total	Data intermediaries	revenues	total additional revenues	65 557 875	70 000 000	-	35 000 000	8 750 000	8 750 000	8 750 000	8 750 000
	PO2	Total	Data intermediaries	add.revenues	revenues per intermediary	463 824	494 066	-	254 545	63 636	63 636	63 636	48 611
	PO2	Total	Data intermediaries	Client base	number of clients %			-	40%	10%	10%	10%	10%
	PO2	Total	Data intermediaries	Client base	number of use cases			-	35%	10%	10%	10%	10%
	PO2	Total	Data intermediaries	Increased competitio	B2B Market			-	-	3%	3%	3%	3%
	PO2	Total	Data intermediaries	Increased competitio	C2B Market			-	0,1	0,1	0,1	0,1	0,1
	PO2	Total	Data intermediaries	Benefits	Benefits			-					
Costs	PO2	Total	Data intermediaries	Costs	Compliance one-off for obtaining the certification	(5 097 087)	(5 250 000)	(5 250 000)	-	-	-	-	-
	PO2	Total	Data intermediaries	Costs	Compliance recurrent costs of renewing certification	(19 408 853)	(21 900 000)	-	(4 125 000)	(4 125 000)	(4 125 000)	(4 125 000)	(5 400 000)
	PO2	Total	Data intermediaries	Reduced Competition	B2B Market		-25,0%		-	-	-	-	-
	PO2	Total											
Costs total	PO2	Total	Total	Costs	Costs total	(24 505 940)		(5 250 000,0)	(4 125 000,0)	(4 125 000,0)	(4 125 000,0)	(4 125 000,0)	(5 400 000,0)
Benefits total	PO2	Total	Total	Benefits	Benefits total	65 557 875		-	35 000 000,0	8 750 000,0	8 750 000,0	8 750 000,0	8 750 000,0
Net Cashflow NPV	PO2	Total	Net Cashflow NPV	NPV	NPV	39 142 482,0		(5 250 000,0)	30 875 000,0	4 625 000,0	4 625 000,0	4 625 000,0	3 350 000,0
Benefit/Cost-ratio	PO2	Total	Benefit/Cost-ratio	BCR	BCR	2,7							

€ (constant prices)	PO	MS	Stakeholder	Category	Subcategory	NPV @ 3%	Total	2023	2024	2025	2026	2027	2028
Benefits	PO3	Total	Data intermediaries	no.	no.	n/a	n/a	150	110	121	132	155	210
	PO3	Total	Data intermediaries	efficiency gains %	OPEX	n/a	n/a	- 7,5%	7,5%		10%	10%	10%
	PO3	Total	Data intermediaries	revenues	total additional revenues	46 611 499	49 000 000	-	35 000 000	3 500 000	3 500 000	3 500 000	3 500 000
	PO3	Total	Data intermediaries	revenues	revenues per intermediary	394 884	412 870	-	318 182	28 926	26 515	22 581	16 667
	PO3	Total	Data intermediaries	Client base	number of clients %			-	40%	10% 10,0%	10,0%	10%
	PO3	Total	Data intermediaries	Client base	number of use cases			-	35%	10% 10,0%	10,0%	10%
	PO3	Total	Data intermediaries	Increased competitio	B2B Market			-	-	3% 3,0%	3,0%	3%
	PO3	Total	Data intermediaries	Increased competitio	C2B Market			-	10%	5% 5,0%	5,0%	5%
	PO3	Total	Data intermediaries	Benefits	Benefits								
Costs	PO3	Total	Data intermediaries	Costs	Compliance one-off for obtaining the certification	(5 825 243)	(6 000 000)	(6 000 000,0)	-	-	-	-	-
	PO3	Total	Data intermediaries	Costs	Compliance recurrent	(25 650 961)	(29 120 000)	-	(4 400 000)	(4 840 000)	(5 280 000)	(6 200 000)	(8 400 000)
	PO3	Total	Data intermediaries	Reduced Competition	B2B Market		-25,0%		-	-	-	-	-
	PO3	Total	Data intermediaries		C2BMarket		-20,0%						
Costs total	PO3	Total	Total	Costs	Costs total	(31 476 204)		(6 000 000,0)	(4 400 000,0)	(4 840 000,0)	(5 280 000,0)	(6 200 000,0)	(8 400 000,0)
Benefits total	PO3	Total	Total	Benefits	Benefits total	46 611 499		-	35 000 000,0	3 500 000,0	3 500 000,0	3 500 000,0	3 500 000,0
Net Cashflow NPV	PO3	Total	Net Cashflow NPV	NPV	NPV	13 777 678,8		(6 000 000,0)	30 600 000,0	(1 340 000,0)	(1 780 000,0)	(2 700 000,0)	(4 900 000,0)
Benefit/Cost-ratio	PO3	Total	Benefit/Cost-ratio	BCR	BCR	1,5			-7,95	-0,72	-0,66	-0,56	-0,42

Annex D – Macro economic analysis | Top/down

Data sharing | Economic Impact

M€	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028
Real GDP (% change p.a. EIU/OECD)		(7,8%)	5,3%	2,8%	2,1%	1,8%	1,5%	1,5%	1,6%	1,6%
EU Data Monitoring Tool 2020 - baseline										
Data revenues	64 262	71 050	75 866	81 008	86 499	92 362	98 623	100 144	101 711	103 321
Data market value	58 214	62 244	65 795	69 584	73 628	77 948	82 564	83 837	85 149	86 497
Value of Data Economy										
Direct Impact	58 214	54 081	58 481	63 239	68 385	73 948	79 965	81 198	82 469	83 775
Indirect Backward Impact	3 197	3 105	3 324	3 559	3 811	4 081	4 369	4 436	4 506	4 577
Indirect Forward Impact	155 389	150 887	161 556	172 979	185 209	198 305	212 326	215 600	218 975	222 441
Induced Impact	108 058	98 853	115 213	134 280	156 502	182 402	212 589	215 867	219 246	222 717
Total Impact	**324 858**	**306 926**	**338 574**	**374 057**	**413 907**	**458 736**	**509 249**	**517 101**	**525 197**	**533 510**
EU Data Monitoring Tool 2020 - high growth										
Data revenues	64 262	71 050	80 943	92 215	105 055	119 684	136 350	138 453	140 620	142 846
Data market value	58 214	62 244	69 320	77 236	86 097	96 020	107 139	108 791	110 494	112 243
Value of Data Economy										
Direct Impact	58 214	54 081	62 005	71 090	81 505	93 447	107 139	108 791	110 494	112 243
Indirect Backward Impact	3 197	3 105	3 622	4 224	4 928	5 748	6 704	6 808	6 914	7 024
Indirect Forward Impact	155 389	150 887	176 002	205 296	239 467	279 324	325 817	330 840	336 020	341 339
Induced Impact	108 058	98 853	129 651	170 044	223 023	292 506	383 638	389 553	395 652	401 915
Total Impact	**324 858**	**306 926**	**371 279**	**450 655**	**548 922**	**671 026**	**823 298**	**835 992**	**849 081**	**862 521**
Data sharing [% of total Data Economy]		80,0%	80,0%	80,0%	80,0%	80,0%	80,0%	80,0%	80,0%	80,0%
- share linked to trust [% of total Data Economy]		50,0%	50,0%	50,0%	50,0%	50,0%	50,0%	50,0%	50,0%	50,0%
Data sharing [% linked to trust]		40,0%	40,0%	40,0%	40,0%	40,0%	40,0%	40,0%	40,0%	40,0%
Data sharing linked to trust - potential gap	-	-	13 082	30 639	54 006	84 916	125 620	127 556	129 553	131 604

Data sharing │ Economic Impact

M€	2023	2024	2025	2026	2027	2028
Policy impact - top-down estimation						
Policy Option 1 [% realisation of gap]						
1.1		0.15%	0.15%	0.15%	0.15%	0.15%
1.2		0.3%	0.3%	0.3%	0.3%	0.3%
1.3		1.0%	1.0%	1.0%	1.0%	1.0%
1.4		1.0%	1.0%	1.0%	1.0%	1.0%
Policy Option 2 [% realisation of gap]						
1.1		0.8%	0.8%	0.8%	0.8%	0.8%
1.2		1.2%	1.2%	1.2%	1.2%	1.2%
1.3		2.0%	2.0%	2.0%	2.0%	2.0%
1.4		4.0%	4.0%	4.0%	4.0%	4.0%
Policy Option 3 [% realisation of gap]						
1.1		1.0%	1.0%	1.0%	1.0%	1.0%
1.2		1.5%	1.5%	1.5%	1.5%	1.5%
1.3		3.0%	3.0%	3.0%	3.0%	3.0%
1.4		5.0%	5.0%	5.0%	5.0%	5.0%
Policy Option 1 [realisation of gap]						
1.1		127	188	191	194	197
1.2		255	377	383	389	395
1.3		849	1 256	1 276	1 296	1 316
1.4		849	1 256	1 276	1 296	1 316
Policy Option 2 [realisation of gap]						
1.1		679	1 005	1 020	1 036	1 053
1.2		1 019	1 507	1 531	1 555	1 579
1.3		1 698	2 512	2 551	2 591	2 632
1.4		3 397	5 025	5 102	5 182	5 264
Policy Option 3 [realisation of gap]						
1.1		849	1 256	1 276	1 296	1 316
1.2		1 274	1 884	1 913	1 943	1 974
1.3		2 547	3 769	3 827	3 887	3 948
1.4		4 246	6 281	6 378	6 478	6 580
Policy Package 1 (low intensity)		6 793	10 050	10 205	10 364	10 528
Policy Package 2 (high intensity)		8 916	13 190	13 393	13 603	13 818
Policy Package 3 (mixed option)		7 048	10 426	10 587	10 753	10 923

Annex E – Macro economic analysis | Bottom/up

Data sharing │ Economic Impact

M€	2023	2024	2025	2026	2027	2028
Policy impact - bottom up (based on CBA result						
Policy Option 1 - direct						
1.1	-	-	-	-	-	-
1.2	-	-	-	-	-	-
1.3	(0.0)	1 050.0	1 050.0	1 050.0	1 050.0	1 050.0
1.4	(2.3)	23.6	4.8	4.8	4.8	4.2
Policy Option 2 - direct						
1.1	(286.3)	709.2	709.2	709.2	709.2	709.2
1.2	(3.8)	0.1	0.6	0.6	0.6	0.6
1.3	(0.3)	1 200.0	1 200.0	1 200.0	1 200.0	1 200.0
1.4	(5.3)	30.9	4.6	4.6	4.6	3.4
Policy Option 3 - direct						
1.1	(572.7)	1 090.8	1 090.8	1 090.8	1 090.8	1 090.8
1.2	(13.7)	43.7	48.7	53.7	58.4	63.3
1.3	(3.5)	1 350.0	1 350.0	1 350.0	1 350.0	1 350.0
1.4	(6.0)	30.6	(1.3)	(1.8)	(2.7)	(4.9)
Policy Package 1 (low intensity) - direct		1 940	1 914	1 914	1 914	1 913
Policy Package 2 (high intensity) - direct		2 515	2 488	2 493	2 497	2 499
Policy Package 3 (mixed option) - direct		1 984	1 963	1 968	1 972	1 976
EU Data Monitoring Tool Multipliers (% of direct)						
Baseline						
Direct Impact	1.00	1.00	1.00	1.00	1.00	1.00
Indirect Backward Impact	0.06	0.06	0.05	0.05	0.05	0.05
Indirect Forward Impact	2.71	2.68	2.66	2.66	2.66	2.66
Induced Impact	2.29	2.47	2.66	2.66	2.66	2.66
Total Impact	6.05	6.20	6.37	6.37	6.37	6.37
High Growth						
Direct Impact	1.00	1.00	1.00	1.00	1.00	1.00
Indirect Backward Impact	0.06	0.06	0.06	0.06	0.06	0.06
Indirect Forward Impact	2.94	2.99	3.04	3.04	3.04	3.04
Induced Impact	2.74	3.13	3.58	3.58	3.58	3.58
Total Impact	6.73	7.18	7.68	7.68	7.68	7.68

Data sharing │ Economic Impact

M€	2023	2024	2025	2026	2027	2028
Multiplier (indirect forward) applied						
Policy Option 1 - indirect forward		2.68	2.66	2.66	2.66	2.66
1.1		-	-	-	-	-
1.2		-	-	-	-	-
1.3		2 815.7	2 788.0	2 788.0	2 788.0	2 788.0
1.4		63.3	12.9	12.9	12.9	11.1
Policy Option 2 - indirect forward						
1.1		1 901.9	1 883.2	1 883.2	1 883.2	1 883.2
1.2		0.1	1.7	1.7	1.7	1.7
1.3		3 218.0	3 186.3	3 186.3	3 186.3	3 186.3
1.4		82.8	12.3	12.3	12.3	8.9
Policy Option 3 - indirect forward						
1.1		2 925.2	2 896.3	2 896.3	2 896.3	2 896.3
1.2		117.3	129.4	142.7	155.1	168.0
1.3		3 620.2	3 584.6	3 584.6	3 584.6	3 584.6
1.4		82.1	(3.6)	(4.7)	(7.2)	(13.0)
Policy Package 1 (low intensity) - indirect		5 203	5 083	5 083	5 083	5 080
Policy Package 2 (high intensity) - indirect		6 745	6 607	6 619	6 629	6 636
Policy Package 3 (mixed option) - indirect		5 320	5 211	5 224	5 237	5 246
Summary - results						
Top-down						
Policy Package 1 (low intensity) - total		6 793	10 050	10 205	10 364	10 528
Policy Package 2 (high intensity) - total		8 916	13 190	13 393	13 603	13 818
Policy Package 3 (mixed option) - total		7 048	10 426	10 587	10 753	10 923
Bottom-up						
Policy Package 1 (low intensity) - total		7 143	6 998	6 998	6 998	6 993
Policy Package 2 (high intensity) - total		9 260	9 095	9 112	9 125	9 135
Policy Package 3 (mixed option) - total		7 304	7 174	7 192	7 209	7 222
Average						
Policy Package 1 (low intensity) - total		6 968	8 524	8 601	8 681	8 761
Policy Package 2 (high intensity) - total		9 088	11 142	11 253	11 364	11 477
Policy Package 3 (mixed option) - total		7 176	8 800	8 890	8 981	9 073
GDP - EDM pre Covid	13 287 687	13 487 002	13 690 074	13 901 151	14 118 796	14 342 287
GDP - EDM post Covid (EIU/OECD correction)	12 742 133	12 972 879	13 168 121	13 371 151	13 580 498	13 795 468

Data sharing | Economic Impact

M€	2020	2021	2022	2023	2024	2025	2026	2027	2028
Impact on the Economic Value of the Data Economy compared to GDP [m€]									
Baseline	306 926	338 574	374 057	413 907	458 736	509 249	517 101	525 197	533 510
% Baseline to GDP	*2.66%*	*2.79%*	*3.00%*	*3.25%*	*3.54%*	*3.87%*	*3.87%*	*3.87%*	*3.87%*
Policy Package 1 (top-down)	306 926	338 574	374 057	413 907	465 529	519 299	527 305	535 561	544 039
% Policy Package 1 to GDP	*2.66%*	*2.79%*	*3.00%*	*3.25%*	*3.59%*	*3.94%*	*3.94%*	*3.94%*	*3.94%*
Policy Package 2 (top-down)	306 926	338 574	374 057	413 907	467 652	522 439	530 494	538 800	547 329
% Policy Package 2 to GDP	*2.66%*	*2.79%*	*3.00%*	*3.25%*	*3.60%*	*3.97%*	*3.97%*	*3.97%*	*3.97%*
Policy Package 3 (top-down)	306 926	338 574	374 057	413 907	465 784	519 675	527 688	535 950	544 433
% Policy Package 3 to GDP	*2.66%*	*2.79%*	*3.00%*	*3.25%*	*3.59%*	*3.95%*	*3.95%*	*3.95%*	*3.95%*
Baseline	306 926	338 574	374 057	413 907	458 736	509 249	517 101	525 197	533 510
% Baseline to GDP	*2.66%*	*2.79%*	*3.00%*	*3.25%*	*3.54%*	*3.87%*	*3.87%*	*3.87%*	*3.87%*
Policy Package 1 (bottom-up)	306 926	338 574	374 057	413 907	465 879	516 247	524 099	532 195	540 504
% Policy Package 1 to GDP	*2.66%*	*2.79%*	*3.00%*	*3.25%*	*3.59%*	*3.92%*	*3.92%*	*3.92%*	*3.92%*
Policy Package 2 (bottom-up)	306 926	338 574	374 057	413 907	467 996	518 344	526 212	534 322	542 645
% Policy Package 2 to GDP	*2.66%*	*2.79%*	*3.00%*	*3.25%*	*3.61%*	*3.94%*	*3.94%*	*3.93%*	*3.93%*
Policy Package 3 (bottom-up)	306 926	338 574	374 057	413 907	466 040	516 423	524 293	532 406	540 732
% Policy Package 3 to GDP	*2.66%*	*2.79%*	*3.00%*	*3.25%*	*3.59%*	*3.92%*	*3.92%*	*3.92%*	*3.92%*

References

Borgman, Christine L. 2015. *Big Data, Little Data, No Data: Scholarship in the Networked World*. Cambridge, UNITED STATES: MIT Press. http://ebookcentral.proquest.com/lib/georgetown/detail.action?docID=3339930.

Edwards, Paul N., Matthew S. Mayernik, Archer L. Batcheller, Geoffrey C. Bowker, and Christine L. Borgman. 2011. "Science Friction: Data, Metadata, and Collaboration." *Social Studies of Science* 41 (5): 667–90. https://doi.org/10.1177/0306312711413314.

European Commission. 2019a. "Facts and Figures of Open Research Data." Text. European Commission - European Commission. 2019. https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science/open-science-monitor/facts-and-figures-open-research-data_en.

———. 2019b. "Cost-Benefit Analysis for FAIR Research Data : Cost of Not Having FAIR Research Data." Website. https://op.europa.eu:443/en/publication-detail/-/publication/d375368c-1a0a-11e9-8d04-01aa75ed71a1.

Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford. 2020. "Datasheets for Datasets." *ArXiv:1803.09010 [Cs]*, March. http://arxiv.org/abs/1803.09010.

Holzner, Andre, Peter Igo-Kemenes, and Salvatore Mele. 2009. "First Results from the PARSE.Insight Project: HEP Survey on Data Preservation, Re-Use and (Open) Access." *ArXiv:0906.0485 [Hep-Ex, Physics:Physics]*, June. http://arxiv.org/abs/0906.0485.

OECD. 2015. "Making Open Science a Reality." 25. OECD Science, Technology and Industry Policy Papers. Paris: OECD Publishing. https://wiki.lib.sun.ac.za/images/0/02/Open-science-oecd.pdf.

OpenAire, Super. 2019. "RDM Costs." OpenAIRE. 2019. https://www.openaire.eu/how-to-comply-to-h2020-mandates-rdm-costs.

Tenopir, Carol, Elizabeth D. Dalton, Suzie Allard, Mike Frame, Ivanka Pjesivac, Ben Birch, Danielle Pollock, and Kristina Dorsett. 2015. "Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide." Edited by Peter van den Besselaar. *PLOS ONE* 10 (8): e0134826. https://doi.org/10.1371/journal.pone.0134826.