

Empowering society by reusing privately held data for official statistics — A European approach

FINAL REPORT PREPARED BY THE
HIGH-LEVEL EXPERT GROUP ON FACILITATING THE
USE OF NEW DATA SOURCES FOR OFFICIAL STATISTICS

2022 edition



**Empowering society by
reusing privately held data
for official statistics —
A European approach**

FINAL REPORT PREPARED BY THE
HIGH-LEVEL EXPERT GROUP ON FACILITATING THE
USE OF NEW DATA SOURCES FOR OFFICIAL STATISTICS

2022 edition

Manuscript completed in June 2022

This document should not be considered as representative of the European Commission's official position.

Luxembourg: Publications Office of the European Union, 2022

© European Union, 2022



The reuse policy of European Commission documents is implemented based on Commission Decision 2011/833/EU of 12 December 2011 on the reuse of Commission documents (OJ L 330, 14.12.2011, p. 39). Except otherwise noted, the reuse of this document is authorised under a Creative Commons Attribution 4.0 International (CC-BY 4.0) licence (<https://creativecommons.org/licenses/by/4.0/>). This means that reuse is allowed provided appropriate credit is given and any changes are indicated.

For any use or reproduction of elements that are not owned by the European Union, permission may need to be sought directly from the respective rightholders. The European Union does not own the copyright in relation to the following elements:

Copyright for the photograph: Cover © Andrii Yalanskyi/Shutterstock

For more information, please consult: <https://ec.europa.eu/eurostat/about/policies/copyright>

The information and views set out in this publication are those of the authors and do not necessarily reflect the official opinion of the European Union. Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use which may be made of the information contained therein.

Theme: General and regional statistics

Collection: Statistical reports

ISBN 978-92-76-53600-0 ISSN 2529-3222 doi: 10.2785/948477 KS-FT-22-004-EN-N

Official statistics provide an indispensable element in the information system of a democratic society, serving the government, the economy and the public with data about the economic, demographic, social and environmental situation. To this end, official statistics that meet the test of practical utility are to be compiled and made available on an impartial basis by official statistical agencies to honour citizens' entitlement to public information.

Fundamental Principles of Official Statistics, Article 1

<https://unece.org/statistics/fundamental-principles-official-statistics>

We acknowledge the importance of developing sound statistical infrastructures, including through dedicated statistical surveys, appropriate domestic, national and international legal and technical frameworks for data access and use, while protecting personal data and privacy, strengthening of NSOs' capabilities in using linked data, increased availability of open data, and enhanced collaboration with the private sector and relevant stakeholders, including in exploring alternative sources of data and data collection practices.

Declaration of G20 Digital Ministers

https://www.g20.org/wp-content/uploads/2021/08/DECLARATION-OF-G20-DIGITAL-MINISTERS-2021_FINAL.pdf

Foreword

Digital transformation is changing our societies and economies profoundly at unprecedented speed. The abundance of data created as by-products of digital services is opening enormous opportunities to use them to improve government services and foster economic growth and job creation. It also allows for the provision of timelier and more granular official statistics and insights to support evidence-based decision-making and public debate.

To meet increasing demands and challenges in a digital society, European official statistics must innovate, respond faster to the data needs of emerging users, and adapt its professional culture to a digital world. European official statistics must take full advantage of the opportunities provided by the digital transformation in terms of the sustainable use of new data sources and digital technologies.

Moreover, Eurostat and National Statistical Institutes in the EU should invest in innovation, collaborate and establish partnerships with data providers and research communities.

We therefore invited a group of independent experts to reflect on how to enhance the use of new data sources for the compilation of official statistics under the European Data Strategy. I would like to thank the 20 experts who supported Eurostat and the European Commission in this endeavour. This report summarises their work and contains key recommendations that will contribute to ensuring the systematic, responsible and sustainable use of new data for compilation of European official statistics.

The report calls on the Commission, the Member States and all stakeholders to prioritise actions in four strands: fair and effective partnerships between businesses and statistical authorities, modernising the means of producing official statistics, engaging the general public and involving other stakeholders in reusing privately-held data for the compilation of official statistics.

Mariana Kotzeva, Director-General of Eurostat



Executive summary

Official statistics are statistics provided by national and international statistical bodies as a public good, to everybody and for free. They are objective and impartial and comply with internationally agreed quality standards. They aim at providing high-quality information for policymakers and the citizens. Official statistics are essential for effective decision-making over a wide range of areas, from economic development to environmental measures, from social policies to emergency response and from public health to public services. They support informed decision-making and the general knowledge of wider society, evidence-based public debate and a democratic process that allows all stakeholders to participate on an equal footing. And they are essential for businesses' day-to-day operations and investment decisions. In short, **official statistics are indispensable to well-functioning economies and democracies.**

Recent crises – from the financial crisis to the COVID-19 pandemic, and now the war in Ukraine – have shown that there is no room for complacency. In our rapidly changing and increasingly data-driven society, there is a growing demand for better, more up-to-date and more detailed statistics. Providers of official statistics must therefore **innovate and equip open societies with the tools they need to respond to an increasingly uncertain environment** and the many associated challenges and opportunities. For instance, serious data gaps emerged during the 2008-2009 financial crisis and the COVID-19 pandemic. NextGenerationEU, the European Green Deal and other flagship EU policy programmes involve ambitious changes such as the green transition, the digital revolution and environmental challenges. Radical advances must be made in data supply and use, together with major investment in data production, data quality and metadata. The widening gap between what is expected from official statistics and what is currently delivered in practice requires vision, energy and urgent action. The world of official statistics is therefore undergoing profound transformation at national, EU and global level.

The EU is playing a leading role in this context. It is a global leader in the promotion of personal data protection standards. It has undertaken a wide range of measures to upgrade its technological infrastructure. It has set ambitious, evidence-based and data-driven goals and targets in the areas of climate change and sustainable development, economic growth and stability, employment and social inclusion, and energy, food and health security. The European strategy for data has already produced a number of legislative acts including the Open Data Directive, the Data Governance Act, the Digital Markets Act and the Digital Services Act. These will soon be joined by other legislation such as the Data Act.

The challenges and opportunities considered in this report lie at the core of the EU's approach for developing official statistics. Statistics are still in practice largely based on data collected by surveys and administrative records rather than on the wealth of data which is held by private businesses and civil society organisations and which is powering the digital economy. The current data revolution provides unprecedented opportunities to collect, analyse and make use of new sources of data that can empower our economies and societies, improve decision-making and bring benefits for all members of society.

Eurostat has set up an expert group on 'Facilitating the use of new data sources for official statistics' to reflect on these new opportunities and make recommendations to enhance the reuse of private sector data in official statistics under the European strategy for data. In this way, statistical authorities will strengthen society by providing more and better statistics that respond to evolving needs. The term 'reuse' reflects the fact that measures enabling statistical authorities to extract statistical information from data do not necessarily involve any actual transfer of data and can include a wide range of other technical solutions. The term is also a reminder that, while data are primarily collected and used for one purpose (business), reusing them for a different purpose (official statistics) entails additional challenges in a variety of areas ranging from methodological and technical issues to legal and ethical questions.

A well-established legal framework for official statistics, at both the national and EU level, already mandates data collection by means of surveys and the reuse of administrative records. However, **the current framework does not cover the reuse of privately held data** for statistical purposes. This report highlights regulatory gaps, fragmentation of practices and a lack of clarity regarding businesses' rights and obligations. Voluntary partnerships between businesses and statistical authorities have recently been set up and rolled out. These have often led to one-off projects, innovation and cutting-edge exploratory research. However, they have rarely led to **sustainable** data reuse in the regular production of official statistics.

To obtain the maximum benefit for society, the Expert Group proposes to prioritise action in **four areas**.

First, **fair and effective partnerships between businesses and statistical authorities** must be promoted on a systematic and regular basis. Such partnerships have to be based on a mutual recognition that it is legitimate for the different parties to have different roles and interests; they also need to be based on trust, social responsibility, and the premise that the reuse of privately held data for official statistics will benefit the whole of society – including the partners themselves. This presupposes a balanced regulatory framework that makes it possible to reuse privately held data for statistical purposes in a sustainable way.

- Statistical authorities and private data holders should **develop a partnership approach** to maximise business incentives and minimise risk, based on mutually agreed operational modalities of data reuse. Statistical authorities should provide incentives and make it easier for private data holders to set up successful partnerships with them. They could, for example, help private data holders to enhance their public reputation and provide feedback advice on ways to improve and better use their data.
- The legal framework should set out a clear set of requirements and safeguards for private data holders. Where it is not possible to establish voluntary partnerships based on incentives, there should be a **legal requirement to ensure the sustainable production of official statistics – subject to a clear set of limitations, conditions and safeguards** (proportionate costs and benefits; mechanisms to deal in a fair way with costs incurred by data holders; protection of business interests and limitation of business liability; and adequate protection for data confidentiality). While statistical authorities should not pay for data, viable cost compensation schemes should be developed, including for services offered by data holders that the statistical authorities decide to use. Dispute resolution mechanisms should also be put in place.

Second, the **approach to dealing with data and producing statistics should be modernised** to make the best possible use of emerging opportunities and to make the system of official statistics fit for the future.

- A transparent **methodological and quality framework** for the reuse of privately held data should be developed for official statistics. The framework should take into account the fact that the future production of official statistics will be based on the integration of different data sources and incorporate new statistical products. Statistical authorities

should consider new analytic and modelling methods, apply open science principles, and be ready to learn from the private sector's ways of processing data.

- **Technical and organisational measures** should safeguard the security and confidentiality of privately held data that is being reused for statistical purposes. The adoption in this context of advanced privacy enhancing techniques should be promoted. For certain kinds of privately held data, access and processing for statistical purposes should be arranged at EU level. Eurostat can organise this on behalf of statistical authorities at all levels whenever this would reduce costs and the administrative burden for statistical authorities and data holders.
- Statistical products and processes involving the reuse of privately held data should be significantly **harmonised across the EU and globally**. This would promote efficiency and effectiveness, and involve wide-ranging standardisation. It would also help to limit and reduce the administrative burden on private businesses as much as possible.

Third, given the lack of direct involvement of the general public at the data collection stage, the reuse of privately held data for statistical purposes requires **social validation** based on engaging with the general public and involving all stakeholders.

- **Inclusive and open participatory mechanisms** should be promoted to identify the key questions raised by the reuse of privately held data for statistical purposes. It is essential to engage directly with the public – who are often the subjects of the data – with a view to identifying information gaps and enhancing the legitimacy and public acceptance of data reuse.
- Formal consultation mechanisms with **balanced representation of all stakeholders** (the general public, businesses, users of official statistics) should be taken into account when selecting statistical products based on the reuse of privately held data.
- The definition of official statistics based on privately held data requires an **agile decision process**. The statistical system must be able to identify and respond in a timely way to the needs of a changing society as expressed by the participatory and consultation mechanisms.

Fourth, the reuse of privately held data should **involve the wider data ecosystem** in the innovation of official statistics. It is important to actively strengthen the collaboration with research communities and to develop data stewardship functions for all organisations that are concerned.

- Statistical authorities should **establish robust collaborative research networks** with researchers from academic institutions, private companies and non-profit organisations across a diverse range of scientific disciplines. Trust in statistical methods and results should be based on shared and open science principles, and should be supported by a transparent vetting process for which statistical authorities should take responsibility. Statistical authorities and private data holders should be offered incentives and support from dedicated funding schemes, encouraging them to extend data reuse agreements to include collaborative research activities that involve external researchers.
- Both statistical authorities and the private sector should create and **implement the function of data steward** and put in place organisational structures dedicated to efficient data reuse. It is important to develop a community of data stewards guided by professional principles. Statistical authorities should be encouraged to serve as data stewards for the public sector.
- Statistical authorities should strengthen their **educational role and improve public data literacy** in connection with the reuse of privately held data for public purposes. Such literacy measures should target both the general public and businesses.

Finally, the **future system of official statistics requires considerable, coordinated and long-term funding and investment**. Public authorities at all levels as well as the private sector need to make significant investments in statistical infrastructure with a view to enhancing data research and innovation, data reuse for statistical purposes, product and process quality, technology, skills and capacities. It will be necessary to use public funding instruments and to leverage the contribution of private investors. Statistical authorities will have to increase their investment in public goods such as common services and shared infrastructures that support the reuse of privately held data.

The Expert Group believes that **the reuse of privately held data can be of great value to the economy and society** and that urgent action needs to be taken to use this opportunity to the full. The implementation of the recommendations set out in this report should lead to a thriving data culture and ecosystem in the EU, with balanced and innovative partnerships and mechanisms to ensure clear legitimacy and social validation. Statistical authorities will be able to produce more, better and timelier statistics in innovative, agile and responsive ways that will allow the general public, businesses and policymakers to form evidence-based opinions and take informed decisions. An innovative system of official statistics that is fit for the challenges of the future is essential if we are to sustain a healthy, prosperous and resilient democratic society in the information era.

Table of contents

Foreword	4
Executive summary	5
Table of contents	9
Table of figures	11
Abbreviations	12
Introduction	13
1. Background and objectives	15
1.1. The state of official statistics	15
1.1.1. Historical perspective	15
1.1.2. Rules underpinning the statistical system	16
1.2. Future of official statistics	17
1.2.1. European data strategy and the Data Act	17
1.2.2. Need and drivers for change	18
1.2.3. Limits on voluntary collaboration	19
1.2.4. Possible outlook for the future statistical system	21
2. Considerations regarding the reuse of data for official statistics	24
2.1. Overview	24
2.2. Citizen participation()	25
2.2.1. The case for citizen participation	25
2.2.2. Scope of citizen participation	26
2.2.3. Stages of citizen participation	27
2.2.4. Mechanisms of citizen participation	27
2.3. Operational modalities of data reuse	29
2.3.1. Dimensions of operational modalities	29
2.3.2. Data access	30
2.3.3. Process quality	30
2.3.4. Roles and responsibilities()	31
2.3.5. Cost of reuse	32
2.4. Business interest	32
2.4.1. Risks perceived by business	33
2.4.2. Opportunities perceived by business	36
2.5. Data stewardship roles	37

3. Principles for the reuse of privately held data for official statistics	40
3.1. General principles	40
3.2. Principles concerning the interaction between private data holders and statistical authorities	44
3.3. Enablers.....	46
4. Recommendations on the reuse of privately held data for official statistics ...	48
4.1. Engaging with citizens and businesses	48
4.2. The approach to data	49
4.3. Balanced partnerships.....	50
4.4. Engaging with researchers.....	52
4.5. Improving the institutional environment	52
4.6. Next steps.....	54
Annex I	57
Glossary of Terms.....	57
Annex II.....	59
Members of the High-Level Expert Group	59

Table of figures

Figure 1: The cycle of citizen participation.....	27
Figure 2: The DIKW pyramid	34
Figure 3: Solove’s taxonomy of privacy	35
Figure 4: The 9Rs framework: the business case for data reuse in the public interest	36
Figure 5: Representation of the decision process leading to determination of the official statistics portfolio and selection of priority areas.....	42

Abbreviations

B2G	Business to Government data sharing
B2G4S	B2G for official statistics
D4D	Data for Development
D4R	Data for Refugees
DaaS	Data as a Service
DGA	Data Governance Act
EC	European Commission
ESAC	European Statistical Advisory Committee
ESGAB	European Statistical Governance Advisory Board
ESS	European Statistical System
ESTP	European Statistical Training Programme
FAIR	Findable, Accessible, Interoperable, Reusable
NUTS	Nomenclature des unités territoriales statistiques (the commonly used French abbreviation for the Nomenclature of territorial units for statistics)
PET	Privacy Enhancing Technologies
SPC	Secure Private Computing

Introduction

The High-Level Expert Group on facilitating the use of new data sources for official statistics has been created in the context of the data and digital strategy of the European Commission (EC). The task of the Expert Group is to provide recommendations aimed at enhancing data sharing between businesses and government (B2G) for the purpose of producing official statistics (B2G4S). The Expert Group consists of high-level experts with various backgrounds that are particularly relevant to B2G4S.

Businesses generate and use data primarily for business-related purposes. The motivation for B2G4S stems from the high societal value that such privately held data can potentially generate when transformed into reliable, relevant and timely official statistics that are made available to everybody, for free. Transforming data into statistical information requires cooperation between private data holders and statistical authorities. On a voluntary basis there have been many collaborative efforts by businesses and statistical authorities to produce statistics based on privately held data, but for various reasons the use of such data for official statistics is still far below the level required to provide society with the high-quality and timely official statistics it needs in the increasingly data-driven world.

The challenge is to arrange for sustainable reuse of privately held data for official statistics in a way that is respectful of the interests of the businesses holding the data. To this end, proper safeguards must be provided, based on a genuine dialogue with businesses and other stakeholders. Proportionality and mutual recognition of interests are key. Achieving this common goal involves efforts and contributions from both sides. On one hand, the system of official statistics will need to adapt, to respond to society's increasing information needs, and part of the challenge relates to the fact that dealing with privately held data requires a leap forward in innovation of statistical processes and products. On the other hand, companies must enable the reuse of the data, provided due process is applied and appropriate safeguards are in place. In this context, it is important to realise that reusing data for official statistics does not necessarily entail actual transfer of raw data to the statistical authorities, since in certain situations aggregate data may suffice. Also, data processing can be done on the premises of (and possibly by) the data holder, and novel privacy-enhancing technologies (PET) are available to allow extraction of statistical information without accessing the raw data directly. In other words, direct access to (and sharing of) the raw data must be seen as *one possible means*, not a goal per se. Alternative approaches that allow statistical information to be extracted without relying on traditional data access or data transfer schemes must be considered legitimate options. For this reason, the Expert Group has chosen to refer to the broadly general notion of “data reuse”. The goal is therefore to ensure a *sustainable reuse* of privately held data for official statistics, with or without direct access by statistical authorities to the data.

Some of these observations do not only apply to data held by businesses, but also to data held by civil society organisations such as non-governmental organisations, think tanks, research

centres, labour unions, social movements and philanthropic foundations. Where this report refers to privately held data, this is meant to include data held by civil society organisations.

This report aims to contribute to the development of the statistical system of today and make it fit for the society of tomorrow. That is, a statistical system that is dynamic and responsive and that will provide more, better and timelier statistics that are relevant and valuable for citizens, businesses, policymakers, the research community and society as a whole. In this process, some kinds of privately held data sources have an irreplaceable role to play.

This report proposes a set of principles and concrete recommendations that are based on these principles. The principles and recommendations apply to members of the European Statistical System (ESS) at national and international level, companies, other public authorities and civil society organisations, and even to the wider data ecosystem and society, since all play a role as stakeholders in the future statistical system. The main goal of this report is to provide advice for the decision-making process on data sharing and data infrastructure at European level that will frame the data ecosystem of the future. The considerations, principles and recommendations in this report include updates to the legislative framework but also interventions and incentives that go beyond legislation.

This report starts with an overview of the current state of official statistics and challenges for the future (Chapter 1). It then develops a number of aspects and issues related to reusing privately held data for official statistics that deserve consideration when drafting principles and recommendations. These include the need for increased citizen participation in the process and ways to take into account business interests (Chapter 2). The proposed set of principles is then presented (Chapter 3), followed by the proposed recommendations (Chapter 4). The main terms used in this report are listed in Annex I, and the abbreviations are explained at the beginning of the document before the Introduction.

1

Background and objectives

1.1. The state of official statistics

1.1.1. Historical perspective

The evolution of statistical systems in Europe during the past 200 years reflects the increasing recognition of the crucial role that statistics play in democratic societies. The development, production and dissemination of “official statistics” are delegated to statistical authorities with, in principle, guaranteed independence from economic interests and political powers.

Legal provisions and deontological rules, at national level (and, for the EU, at international level), are based on the principles of impartiality and compliance with high scientific standards. All official statistics are based on the fundamental principles of official statistics established by the UN⁽¹⁾.

Compliance with such rules is underpinned by agreed ethical standards and, for the EU, supported by an elaborate and binding quality framework that includes peer reviews. The statistical governance structures involve committees and advisory boards composed of scientists and external experts from different European countries, and stakeholders are involved in the process. Official statistics are delivered to everybody at the same time and for free.

The role of the statistical authorities is to deliver an impartial and trustworthy quantitative picture of society, including the economy, to provide knowledge and allow policymakers, businesses and citizens to form evidence-based opinions and take informed decisions. In this sense, independent and impartial official statistics are essential to sustaining a healthy democratic society. For the EU, this complex and elaborate set of rules and practices has taken the form of the European Statistical System (ESS), which is the partnership between the statistical authorities of all European countries and Eurostat, the statistical authority of the European Commission. Together they are responsible for developing, producing and disseminating European statistics.

This statistical system has evolved in a world where data were a scarce resource and had to be either collected directly by the statistical authorities themselves, through censuses and surveys, or accessed from other public bodies (administrative records). However, the statistical system has struggled to keep up with the new data-rich world, although it has very seriously tried to access new data sources and reuse privately held data in partnership with businesses.

(1) UN General Assembly, Fundamental Principles of Official Statistics, 2014, <https://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx>

A number of statistical authorities in Member States have succeeded in producing experimental statistical results based on privately held data⁽²⁾, and the same is true for Eurostat⁽³⁾. These efforts have been going on for quite some time now, but unfortunately, the results so far have been seriously limited in terms of (i) statistical domains and statistical output covered, and (ii) integration of new data sources in regular statistics.

Data sharing by private data holders for official statistics has been limited, and where the data they hold is reused, sustainability is not ensured. The reasons for this are discussed later in this chapter. In fact, the ESS still has to cross the threshold of systematically using and integrating new data sources in European statistics. It has arrived at a historic crossroads and has the ambition to make a leap forward and innovate to adapt to present and future opportunities and needs. This is where the Expert Group aspires to make a significant contribution.

The challenges posed to official statistics can be seen in the context of the wider challenges posed to society if it is to engender optimal data sharing between the private and public sector. This had led the European Commission to launch the High-Level Expert Group on Business-to-Government Data Sharing, which released its final report⁽⁴⁾ in 2020. The current Expert Group on B2G4S builds on the results of the B2G Expert Group, of which detailed account was taken.

1.1.2. Rules underpinning the statistical system

The ESS is based on an elaborate legislative framework, with the Regulation on European statistics⁽⁵⁾ at its centre. The products included in the portfolio of official statistics established by the ESS governance bodies constitute public goods of high interest for society and the economy. The relevant statistical authorities at national and European level are given the legal mandate to produce such statistical products in a way that complies with the fundamental principles of official statistics mentioned above and ensures the highest possible standards in terms of quality, trustworthiness and transparency.

Linked to the legislation is another set of rules that are essential for the statistical system – the quality framework. They are based on the European Statistics Code of Practice⁽⁶⁾, which was first adopted in 2005 and revised in 2011 and 2017. It sets the current standards for developing, producing and disseminating European statistics:

- 16 core principles under which the EU and the national statistical authorities operate
- 84 indicators of best practice and standards.

The core principles cover a wide range, such as on professional independence, statistical confidentiality and data protection, impartiality and objectivity, methodological standards, minimisation of response burden on businesses and citizens, cost effectiveness, accuracy and reliability, coherence and comparability, and accessibility. Thus, the framework ensures quality in its widest sense.

Next to the Code of Practice, there is the Quality Assurance Framework for the ESS. This is a guideline on how to implement the Code of Practice. Furthermore, Eurostat has adopted the Protocol on Impartial Access to Eurostat data⁽⁷⁾ for users, which also supports the Code of

(2) The ESS has a hub for such statistics: <https://ec.europa.eu/eurostat/web/experimental-statistics/overview/ess>

(3) <https://ec.europa.eu/eurostat/web/experimental-statistics>

(4) https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=64954

(5) Regulation (EC) No 223/2009, amended by Regulation (EP, Council) No 2015/759, <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:02009R0223-20150608&from=EN>

(6) European statistics Code of Practice, revised edition 2017, <https://ec.europa.eu/eurostat/web/quality/european-quality-standards/european-statistics-code-of-practice>

(7) https://ec.europa.eu/eurostat/documents/4187653/5798057/IMPARTIAL_ACCESS_2014_JAN-EN.PDF/1f1ffb9b-046d-4c47-94fd-f5ff001d7381?t=1421404518000

Practice and is based on the fundamental principle of impartiality and objectivity. It addresses Eurostat users, Eurostat staff and all partners in the production of European statistics. It contains Eurostat's dissemination policy, including pre-release access arrangements.

The current monitoring system for the quality framework is through ESS peer reviews. There have already been two rounds of peer reviews in 2006-08 and 2013-15. The third round of peer reviews is currently taking place (from 2021 to 2023).

Peer reviews cover the ESS statistical authorities engaged in developing, producing and disseminating European statistics. The reviews are followed by a period of annual monitoring of how the improvement measures developed to address the recommendations in the review reports are implemented.

1.2. Future of official statistics

1.2.1. European data strategy and the Data Act

Promoting the reuse of privately held data for statistical purposes is part of the European data strategy⁽⁸⁾, which aims to make the EU a world leader in terms of data-driven society. Creating a single market for data will allow it to flow freely within the EU and across sectors, for the benefit of businesses, civil society, researchers and public administrations⁽⁹⁾. In the context of the European data strategy a number of legislative initiatives have been taken and will be taken, some of which are relevant to official statistics and data reuse.

The availability of open data in the single market⁽¹⁰⁾ is promoted by the Open Data Directive⁽¹¹⁾. This act considers 'open data' to mean: available free of charge, machine readable, provided via APIs, and where relevant provided as bulk download. The Directive, adopted in 2019, identifies certain "high-value datasets", including for statistics, since many official statistics qualify already or are intended to qualify in the future as open data. A number of implementing acts are planned under the Directive.

Also worth mentioning is the Data Governance Act (DGA), which has gone through much of the legislative process already and is expected to be adopted during the first half of 2022⁽¹²⁾. The DGA contains governance provisions for voluntary data sharing, including on data altruism. For different domains 'European data spaces' are planned, such as for health, the Green Deal, and energy. They will facilitate data pooling and sharing by creating infrastructure and governance frameworks to promote data-driven innovation. They are currently being developed and implemented. Finally, the Act provides for a European Data Innovation Board, which would promote things like data interoperability between data-sharing organisations.

Of particular importance to official statistics is the planned Data Act, the proposal for which was published in February 2022⁽¹³⁾. This Act will set rules on who can use and access data generated in the EU across all economic sectors. It will ensure fairness in the digital environment, stimulate a competitive data market, open opportunities for data-driven innovation and make data more accessible for all. Among many other inputs, it builds on the recommendations of the B2G Expert Group mentioned in previous section. The Data Act will

⁽⁸⁾ Shaping Europe's digital future:

https://ec.europa.eu/info/sites/default/files/communication-shaping-europes-digital-future-feb2020_en_4.pdf

⁽⁹⁾ https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en

⁽¹⁰⁾ <https://digital-strategy.ec.europa.eu/en/policies/legislation-open-data>

⁽¹¹⁾ <https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1561563110433&uri=CELEX:32019L1024>

⁽¹²⁾ The proposal can be found here:

<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020PC0767>

⁽¹³⁾ https://ec.europa.eu/commission/presscorner/detail/en/ip_22_1113

include a chapter on access and use of data by public sector bodies. However, that chapter is limited to use for exceptional needs⁽¹⁴⁾. An impact assessment was carried out for the Act⁽¹⁵⁾, which also covered official statistics. The assessment and its support studies showed that the potential benefits to society of reusing privately held data for official statistics are huge.

Since the Data Act limits its provisions on access and use of data by public sector bodies to exceptional needs, official statistics are not generally covered by this regulation. The Data Act allows for implementing acts to be adopted later, and in this context, it is already planned to draft a proposal to revise the Regulation on European statistics. In February 2022, an initiative aimed at that revision was launched⁽¹⁶⁾. The recommendations of the current Expert Group on B2G4S are meant to be considered in the proposal for the revision. However, as stated earlier, their relevance goes beyond mere legislation.

1.2.2. Need and drivers for change

The value of official statistics as a source of trustworthy information is not less important in the new data-rich and data-driven society than it was in the past world of scarce data. The abundance of “data” does not eliminate society’s need for trustworthy statistics. On the contrary: in a world awash in data (which some scholars have argued have “killed facts”), there is a need for a factual basis on which democratic deliberations and decisions can be made and evaluated.

Accordingly, the importance of respecting the fundamental principles of official statistics – above all, impartiality and independence from other private and public interests – has not diminished. However, while the general principles remain fully valid, the way in which statistical authorities operationalise those principles needs to be profoundly changed.

There are, of course, challenges as well as opportunities. The core challenge is to meet modern needs. In the more complex and faster changing world of today, societies need more, better and timelier official statistics than in the past. These needs are expected to be growing in the future. Due to the extensive changes data availability and the needs of society at large (in particular decision-makers), the statistical system of today is confronted – and will continue to be confronted – with gaps and barriers to using new data sources. That is, to reuse data that is there already. These must be addressed if statistical authorities are to continue to play their key role in delivering essential, impartial, high-quality public information.

The challenges can also be viewed as opportunities. A major opportunity addressed in this report is enabling the reuse of privately held data by statistical authorities. A second one is the possibility for statistical authorities to intensify their dialogue with key stakeholders and society at large, by using more and new channels, so as to identify information gaps and new emerging needs, but also to strengthen their knowledge-building role in a world that is more and more driven by data.

By doing so, statistical authorities may enhance their role in the wider data ecosystem. Furthermore, they need to improve their capacities and capabilities to deal with new, more complex sources of data, including those generated by the private sector and civil society. In so doing, they need to be supported by a legislative framework that enables predictable, ethical and sustainable reuse of external data from different sources for official statistics.

An important driver for change is the fact that trusted data has become a critical resource for decision-making across sectors and domains. Facing multiple, structural uncertainties, there is a growing awareness that data gaps, both in terms of quantity and quality, have affected society’s ability to analyse and draw lessons from the past, to evaluate and manage the

⁽¹⁴⁾ See Chapter V of the proposal: <https://ec.europa.eu/newsroom/dae/redirection/document/83521>

⁽¹⁵⁾ <https://digital-strategy.ec.europa.eu/en/library/impact-assessment-report-and-support-studies-accompanying-proposal-data-act>

⁽¹⁶⁾ https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13332-European-Statistical-System-making-it-fit-for-the-future_en

challenges of the present, and to anticipate and prepare for the risks and opportunities of the future.

Moreover, unequal access to information has deepened asymmetries in society, potentially affecting the proper functioning of democracy. In this context, the role, responsibilities and importance of official statistics have increased and need permanent adaptation to present and future needs.

A forceful and swift response to these challenges is necessary to enable more informed public and private decision-making, strengthen public trust in the information society and promote a data-driven and more cohesive vision of the future. This requires an assessment of what a statistical system is and should be, of what the responsibilities of statistical offices are and how they should operate.

This assessment would guide the investment in the capabilities needed to harness the power of new data. This would not only lead to an acceleration of data reuse for statistical purposes, but, importantly, to a substantial improvement in the ability of the statistical system to deliver more, better and more timely statistics, targeting the principal questions identified by society at large.

It requires also the full mobilisation of the energies and resources of all stakeholders in the statistical systems, both public and private – that is, governments and institutions, businesses and non-governmental or community organisations, local and global players, technology providers and researchers, etc.

We must build an information society that is open, inclusive, advanced, respectful of rights and freedoms and capable of harnessing the new opportunities provided by globalisation, technology and education.

1.2.3. Limits on voluntary collaboration

Preferably, the reuse of privately held data for official statistics will take place on a voluntary basis. There are and have been, in fact, a number of successful cases of voluntary partnerships. One of the notable success stories is the reuse of scanner data from retail businesses for statistics on prices. However, there are limits (of various types) to what can be realised in official statistics on the basis of purely voluntary collaboration.

Obviously, working on a voluntary basis leaves it open to the business to engage in the partnership or not. Although in many cases businesses are responsive, statistical authorities have experienced cases where businesses are not prepared to even discuss a possible partnership. Businesses that are prepared to engage in a discussion may decide that a partnership is not sufficiently beneficial to them. The fact that so many businesses allow the reuse of scanner data for official statistics has to do with the lower response burden that is the direct effect of collaborating with the statistical authorities. However, in other contexts and situations such benefits are not present.

Businesses may also set conditions – for instance financially – that make it hard for statistical authorities to reach an agreement. Moreover, even if both partners try hard, with the business recognising its social responsibility and the statistical authorities trying to provide incentives, there is no guarantee that agreement can be reached and maintained, for all kinds of legitimate reasons. Conditions that are considered essential by one of the partners may not be acceptable for the other, for instance if costs or efforts are likely to be too high, if business interests are judged to be too much at risk, or if basic conditions for official statistics cannot be met.

Competitive dynamics are also at play, and the willingness of one business to cooperate with the statistical authority may dissuade its competitors from doing the same, making it impossible for the statistical authority to establish partnerships with multiple firms in the same sector.

Thus, part or many of the businesses whose collaboration on the reuse of the data held by them is needed for an official statistic may not be convinced to engage in a partnership. Especially in

quasi-monopolistic situations, the collaboration of a specific business may be essential for official statistics. Some statistics cannot be made without reusing privately held data, such as statistics on the gig economy.

Where partnerships are feasible, in many cases the experiences have been positive. There are many examples where the possibility of reusing privately held data for official statistics has been successfully explored. But the conditions for *exploration* are much softer than for the actual reuse of the data to *produce regular* official statistics. To explore what would be potentially feasible, quality standards and guarantees are not as strict as for regular official statistics. Representativeness tends to be less of an issue in the first exploratory stages. This is one of the reasons why there are many cases of successful partnerships aimed at exploration, but very few cases where the actual step to reuse the data for regular official statistics has been taken.

Sustainability of partnerships is another issue. Businesses may terminate partnerships at any time and for any reason, including changes in senior management and business restructuring. In some cases partnerships have been ended by businesses without their motives being stated or otherwise becoming clear. There may be public image issues or changes in the public mood that make businesses withdraw from collaborative partnerships. From the point of view of statistical authorities, there is no guarantee that businesses, whose collaboration is essential for a given statistic, do not drop out, or that the required quality of a statistic can be maintained through time. If sustainability cannot be ensured, there may not be enough basis to invest at all in statistics relying on the reuse of privately held data, and this is true for statistical authorities as well as for the businesses themselves.

There are even more limitations. The fragmented nature of voluntary partnerships makes the introduction of standards difficult. Comparing the results of experiments across different businesses is very difficult. Building partnerships on a voluntary basis often takes much time. In many cases the process takes more than a year – sometimes years – before the first data test set can actually be explored for statistical purposes.

Based on these limitations of the voluntary approach, the ESS has reached the conclusion⁽¹⁷⁾ that, in addition to ongoing efforts to make voluntary collaboration as attractive to businesses as possible, there is a need to include enforceable obligations in the legislative framework, obliging businesses to enable the reuse of privately held data. It is not only the ESS that has reached this conclusion. In preparing the Data Act, a public consultation⁽¹⁸⁾ was held in 2021. The consultation showed that a clear majority of stakeholders (in particular citizens and public administrations) expressed the opinion that B2G data sharing should be compulsory – albeit with clear safeguards – for specific use-cases that have a clear public interest, including in the domain of official statistics⁽¹⁹⁾. This necessity has also been signalled by the European Statistical Governance Advisory Board (ESGAB) in its 2021 annual report⁽²⁰⁾.

Creating the legal possibility to require businesses, under specific conditions and subject to a set of safeguards, to enable the reuse of data held by them for official statistical purposes would have a number of advantages. It would give businesses clarity about their rights and obligations. There would be clear rules and procedures. Grey areas and uncertainties in legislation would be eliminated. This would underpin a level playing field, equally treating all businesses in the same sector with regard to costs and benefits, duties and rights. Asymmetries could be avoided. Negotiations to enter partnerships could become considerably shorter if both partners know in advance that the main issues to be solved concern specific implementation

⁽¹⁷⁾ ESS position paper:

<https://ec.europa.eu/eurostat/documents/13019146/13405116/main+ESS+position+paper+on+future+Data+Act+proposal.pdf/37f3b5c7-abfd-5a05-6be2-fdc4b87ee7d2?t=1631695372906>

⁽¹⁸⁾ <https://digital-strategy.ec.europa.eu/en/public-consultation-data-act-summary-report>

⁽¹⁹⁾ See page 10 of the Explanatory Memorandum to the proposal for the Data Act, <https://ec.europa.eu/newsroom/dae/redirection/document/83521>

⁽²⁰⁾ <https://ec.europa.eu/eurostat/documents/34693/14172844/2021+ESGAB+Annual+Report.pdf/db14138c-59ae-ec6e-5c2c-8e203cb58d9d?t=1642539150270>

aspects of the partnership, not the question of whether or not to engage in a partnership. Furthermore, if the legislation is agreed at European level, there would be less fragmentation in national reuse practices, with advantages in terms of standardisation and comparability, gains in efficiency and effectiveness (especially for businesses operating internationally), and fewer competitive imbalances across the EU.

Adding a baseline requirement for businesses to enable data reuse by statistical authorities would be in line with the approach to official statistics from other sources. Official statistics are a public good, which is the justification for existing statistical reporting obligations for businesses. By the same logic, these obligations would be extended to privately held data, for which the public benefit may arguably be much larger. It is, however, recognised that this approach requires addressing a number of issues that are specific to this data source. The Expert Group has the right composition to provide advice on these issues.

1.2.4. Possible outlook for the future statistical system

If all efforts bear fruit, what could the envisaged future statistical system look like when statistical authorities take stock of privately held data on a regular basis? In terms of its statistical outputs, the products in the **official statistics portfolio** would continue to be specified and publicly documented, along with a detailed description of variables, spatial and temporal granularity, frequency and timeliness of release, etc.

However, the official statistics portfolio would be considerably enriched and continuously improved. Among other advances, new regularly produced statistics would be more detailed (e.g. higher degrees of variable disaggregation, spatial or temporal resolution, higher frequency, etc.), would be released at more frequent intervals, and be more uniform at European level. This would enable decision-makers at all levels to better prepare for, detect and respond to new emergencies. Some illustrative examples of statistical products are given in the box below.

Possible official statistics based on new data sources

These examples show what might be included in the future portfolio of statistics, depending on what is deemed useful by the competent bodies. They are solely for illustrative purposes.

Products based on mobile network operator data.

- Present population: spatial density of people and patterns of temporal variations for them.
- Inbound tourism flows, averaged on agreed temporal basis and spatial granularity (e.g. municipality, county or using NUTS⁽²¹⁾ classification) and specified origin categories.
- Share of population commuting daily.
- Indices of commuting patterns across countries: trans-national workers (daily cross-border commuters), weekly commuters and seasonal workers; labour migration.

Products based on financial transactions data.

- Final consumption expenditure by detailed product categories on a monthly basis, also as a contribution to national accounts.
- Detailed information on cross-border payment transactions by country of origin/destination, both by businesses and by consumers and tourists.
- Quantification of the volume of online shopping in a timely and granular manner.
- Indices derived from network structures to assess economic and systemic vulnerabilities.
- In times of crisis or economic shock: early detection of changes in consumption patterns and financial transactions flows.

Products based on scanner data.

- Sales of consumer goods at retail outlets, aggregated by product category, weekly and regional breakdown (volume and price).
- In times of crisis (such as the COVID pandemic): immediate data on changing consumption patterns, including hoarding behaviour.

What could **future collaboration** between businesses and statistical authorities look like? Even more than is already the case, it would be a true partnership. Statistical authorities would systematically take into account business interests with appropriate safeguards. And businesses would be incentivised to leverage such partnerships to improve their corporate social responsibility and public reputation.

The collaboration arrangements would be based on agreements reached through dialogue for all aspects related to the operational modalities of data reuse. The reuse of privately held data would be implemented systematically and in a considerably broader scope than is currently the case. This increased reuse would require stronger confidentiality and data protection measures, which include business-sensitive data. There may also be collaboration that goes beyond the portfolio of official statistics, for the benefit of all partners concerned.

What could the future **regulatory framework** look like? The European Commission would promote the creation of data partnerships between statistical authorities and data holders, and possibly external researchers, by identifying and removing the main barriers (e.g. legal uncertainties, unequal opportunities) and putting in place a solid set of enablers and policy

⁽²¹⁾ Nomenclature des unités territoriales statistiques (the commonly used French abbreviation for the Nomenclature of territorial units for statistics)

incentives. The regulatory framework would include rules designed to make effective and efficient use of privately held data, with commensurate roles and responsibilities for the partners in the ESS. It would be designed in a way that makes the ESS agile and responsive to new needs.

Above all, the regulatory framework would provide the answer to the conundrum of combining the need to base reuse on true partnerships and agreements with the need to provide a legal baseline for systematic and sustainable reuse where satisfactory voluntary collaboration cannot be established. The regulatory framework would be expected to have elements of enforceable requirements for collaboration, combined with strict and equally enforceable conditions and safeguards for business interests.

What could be the **position of official statistics** in society? First of all, given the increased use of privately held data, which are mostly *about* citizens but produced without their direct involvement (compared e.g. to surveys and administrative data), the need for statistical authorities to acquire a “social licence” is higher than ever. To this end, statistical authorities would further strengthen their dialogue with citizens as well as users, policymakers, and other stakeholders. In this process, the centrality of official statistics in a democratic society would be further strengthened.

It may also be expected that the need to cope with a richer and more complex set of data sources and data stakeholders leads the ESS to promote stable research collaborations with a wider range of external partners. It may expand collaborative innovation structures engaging external researchers (academic institutions, research and technology organisations, private sector companies, non-profit organisations) in methodological research and development activities aimed at continuously improving and extending the products and processes of official statistics.

This would involve multi-disciplinary approaches and collaborations between professional statisticians and experts from a wide range of other disciplines. The role of such a collaborative approach would be to:

- i. provide input to the ESS governance bodies, for identifying new candidate products;
- ii. support the ESS experts in the methodological research and development for innovative candidate products that are already included in the official statistics portfolio;
- iii. contribute to the methodological review and further improvement of established products, in the interests of continuously improving and updating the whole official statistics portfolio.

Furthermore, statistical authorities may aspire to have a more prominent role in the wider data ecosystem, in particular being at the forefront of data-driven transformation in the public sector.

What **other characteristics** could the future statistical system have? It may be anticipated that the development of the future statistical system would involve significant public investment and capacity building, in infrastructure, procedures, services, etc.. This would be needed at European as well as national level. The ESS would have to diversify and enrich its staff's skills so as to be better equipped to face the new challenges stemming from the richer and more complex data ecosystem. Testing environments (“platforms”) in support of methodological development will presumably be reinforced at both European and national levels.

2

Considerations regarding the reuse of data for official statistics

2.1. Overview

The previous chapter presented the case for sustainably incorporating the reuse of privately held data in official statistics development, production and dissemination in the ESS. This chapter provides an initial exploration of the organisational aspects, building on the diverse experiences of the members of the Expert Group. It provides the background for the principles to be developed in the following Chapter 3.

Implementing the reuse of privately held data for official statistics requires us to address several questions, such as:

- what are the different roles and possible responsibilities?
- what can be done with the data?
- how to take into account business interests?
- how to ensure the resulting statistics meet the quality requirements?
- how to cover the additional costs?
- how to build an effective partnership between data holders and statistical authorities?

To address such issues, it is important to understand the operational modalities of data reuse and the context in which the reuse would actually take place. This chapter will make an attempt to capture the different operational dimensions of privately held data reuse.

Operational modalities of data reuse – examples

- Raw data are periodically sent to the statistical authority. This model is currently implemented for scanner or electricity data in some countries.
- The data are processed applying standard and documented methods (developed or co-developed by statistical authorities) on the premises of each individual data holder. The resulting aggregate data are passed to the statistical authority for further processing and possibly integration with other data sources.
- The data of multiple data holders is processed by Secure Private Computing (SPC) infrastructure, preventing the disclosure of input data and intermediate data to other, possibly competing, parties, with only the final results passed on to the statistical authority.

The Expert Group has identified four broad dimensions to take into account when designing and implementing the reuse of data for official statistics:

1. **Citizen participation** in the process of establishing what official statistics should measure using privately held data. As such data *about citizens* are not generated

explicitly *by the citizens*, as is the case with survey and administrative records (both of which are based on declarations by the data subjects), strengthening participatory mechanisms at other stages of the decision process is important for reasserting the social legitimacy of statistics produced from such data.

2. **Operational modalities** of data reuse, which in turn determine a number of key factors such as roles, responsibilities, risks and costs.
3. The **business perspective** of data reuse needs to be properly taken into account, to build genuine and effective partnerships between private data holders and statistical authorities.
4. As the wider data ecosystem is currently undergoing a fundamental reshaping, with the value of data increasingly depending on the ability of different (public and private) organisations to cooperate and join forces (and data), the definition and implementation of **data stewardship functions** across private companies and statistical authorities was seen by the Expert Group as an important enabling factor.

2.2. Citizen participation⁽²²⁾

2.2.1. The case for citizen participation

Shaping the framework of reusing privately held data for official statistics and applying that framework to actually generate official statistics will need increased attention on the involvement of citizens, because some (groups of) citizens are not as digitally literate as others, leading to possible imbalances in representing their interests.

Going further, it is useful to distinguish the supply side and the demand side of this citizen participation issue.

On the **supply side**, there is growing availability of private sector data that is inherently generated about citizens as data subjects. Intrinsically, this makes the case for these data subjects having their say in defining how data by and about themselves are to be reused and what it can measure.

On the **demand side**, citizen participation is essentially about addressing the predicament of representative democracy and concerns about the future of social contracts that can be put at risk by the wide availability of tech platforms and digital services which gather data, and the issues of disinformation, polarisation and animosity expressed in public discussions. In this context, if official statistics are to remain trusted and authoritative vis-à-vis alternative statistics that support “alternative facts”, the official statistical process should more than ever involve a dialogue with citizens in the first place. It would ensure human control and oversight, and would make them feel that they count, instead of merely being counted.

More generally, this calls for increased awareness of what official statistics are and do, and better abilities to understand and use them on a daily basis. It remains essential that democratic

⁽²²⁾ This chapter had various sources of inspiration, such as <https://www.peopleledinnovation.org/#/>, <https://the100questions.org/>, <https://thedataassembly.org/>, <https://idsd.network/> (on digital self-determination), <https://youreducationyourvoice.org/>, <https://arxiv.org/pdf/1707.07232.pdf>, (on algorithmic social contracts), https://resources.dial.community/resources/md4d_handbook, (on mobile data for development), <https://blogs.worldbank.org/opendata/promoting-trust-data-through-multistakeholder-data-governance>, and <https://dspace.mit.edu/bitstream/handle/1721.1/123471/Beyond%20Data%20Literacy%202015.pdf?sequence=1&isAllowed=y> (on data literacy).

discussion is rooted in facts and figures that can be rationally debated. The trust and respect people have for official statistics can contribute to this. Promoting channels and incentives for citizens in the official statistical process, especially when it comes to using their data, and certainly any data about them, will counteract some of the opposite tendencies that are emerging today.

2.2.2. Scope of citizen participation

The scope of citizen engagement in the context of reusing privately held data is, in principle, all data collected by private companies and civil society organisations. Although all citizens are covered, it may be useful to categorise them, since there are quite a few differences in the extent of their social inclusion and natural involvement.

For instance, the following (not mutually exclusive) groups can receive special attention :

- non-expert adult citizens;
- young people, including students;
- children of school age;
- representatives of interest groups, including civil society organisations, unions, etc.
- elected officials;
- more marginalised or less connected adults and families, such as migrants, people with disabilities, homeless people, people living in poverty, people in less connected rural areas, etc.

Establishing the groups that are covered (and why) will influence how they ought to be involved and engaged. Their mutual relationships, their relationships with other stakeholders, and possible overlaps (and the way these are captured in privately held datasets) must be well understood. In this context, a case can be made that special attention should be paid to students, young people and even children. Marginalised groups probably deserve greater attention. In terms of priority, representatives (elected people, civil society organisations) should be given high consideration.

2.2.3. Stages of citizen participation

Schematically, citizens may be involved with different degrees of intensity at the following stages, which can be depicted as a cycle (see Figure 1):

- a. Determining what is worth measuring using privately held data, to reflect new demands, expectations, etc., and fully leveraging the new possibilities offered by modern analytical capacities.
- b. Giving input on conditions under which private data is managed, stored and shared, in terms of granularity, frequency, conditions and duration of storage, business models and perhaps remuneration for data subjects.
- c. Giving input on how indicators are computed and made available, including any biases and privacy risks; auditing the adopted pseudonymisation routines and computation algorithms.
- d. Overseeing, monitoring and evaluating the dissemination of indicators and their use by policymakers and other stakeholders.

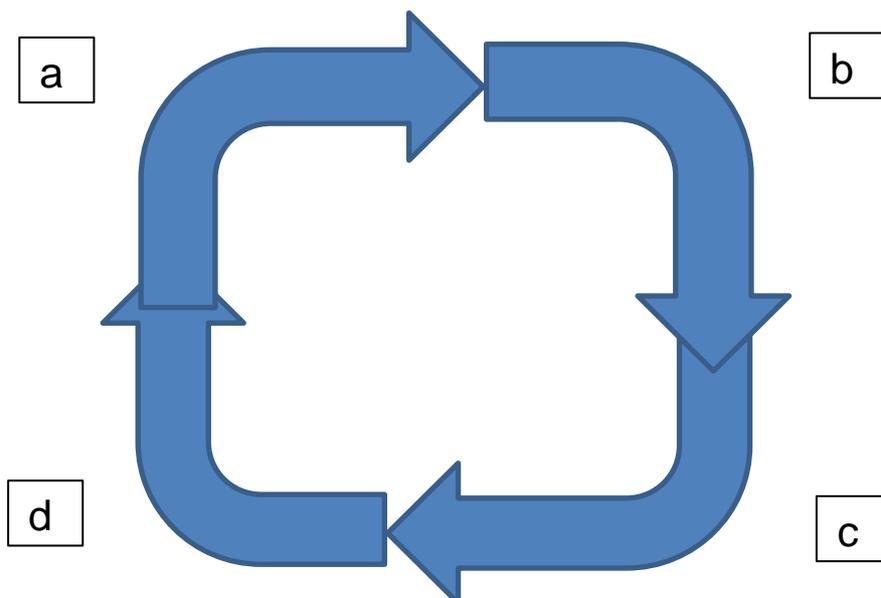


Figure 1: The cycle of citizen participation

2.2.4. Mechanisms of citizen participation

Several mechanisms may be considered to promote citizen participation at the different stages of the cycle, taking into account underlying conditions such as awareness, ability, desire to participate, etc. The mechanisms include targeting specific groups and tailoring to them. Possible examples are:

- Data cooperatives where citizens and/or civil society organisations pool their data and govern it together, including possibly by combining them with private and public sector data in ways that give them more weight and influence⁽²³⁾.
- Data challenges such as D4D (data for development) or D4R (data for refugees) where citizens could be part of teams to better understand the opportunities and requirements of using private sector data.
- Statistical authorities develop communication and advocacy campaigns about their work and the use of private data, to avoid unnecessary or excessive backlash.
- Strengthen data literacy skills and awareness among students and youths.
- Using surveys or adding questions in surveys to ask respondents about new indicators they would like to see measured and their possible concerns and expectations regarding official statistics.
- Crowdsourcing or surveying substantive priorities and concerns of citizens regarding the use of private data⁽²⁴⁾.
- Obligation or possibility to allow groups of citizens to audit or request audits of algorithms, or to be involved in audits.
- Having citizen groups being systematically represented or involved in all innovative data projects developed by statistical authorities as part of CODEs⁽²⁵⁾ or equivalent.
- Support and communicate initiatives leveraging different data sources and parts of societies – as in the case of Alicante⁽²⁶⁾.
- Create a “citizen commission” involving citizens to discuss these topics and make recommendations at EU level, including on business models – as in the case of climate in France⁽²⁷⁾.
- Raise the awareness of elected officials about the potential and requirements of modern official statistics.

It is also important to recognise that citizen participation is also improved by actually reaching out to citizens, not just by offering more ways for them to reach out and be involved.

Although the focus is on citizen participation, this is in fact based on the “whole society” paradigm. Therefore, synergies may be promoted between all parts of society, making invisible and voiceless groups more visible and audible. This also means that, as multiple different mechanisms may be used, ensuring the combination is both complementary and consistent should be an important consideration. It also means that communication channels between different groups should be strengthened, to allow these synergies to be identified. For example, a mechanism designed to allow access to researchers may be repurposed or adapted to allow other stakeholders, for example students, to participate in more controlled research or data challenges, to increase their awareness and interest.

⁽²³⁾ <https://knowledgecentre.euclidnetwork.eu/european-social-enterprise-monitor/>

⁽²⁴⁾ See also Society in the loop: Programming the algorithmic social contract, by Iyad Rahwan, <https://dl.acm.org/doi/10.1007/s10676-017-9430-8>

⁽²⁵⁾ The CODE for building participatory and ethical data projects, see <https://www.data4sdgs.org/news/code-building-participatory-and-ethical-data-projects>

⁽²⁶⁾ <https://www.wired.co.uk/article/valencia-ai-covid-data>

⁽²⁷⁾ <https://www.conventioncitoyennepourleclimat.fr/>

2.3. Operational modalities of data reuse

2.3.1. Dimensions of operational modalities

There are multiple different sources of privately held data, and for each kind of data, a number of possible ways to access and reuse them. Selecting the most suitable operational modality depends on conditions inherent to the specific project, use case or constellation of stakeholders working together for the common goal. There is no single most desirable approach for all data sources; it would be difficult even to make a selection of modalities that would cover most of the existing success stories and would not be refuted by those left out. Key criteria for prioritising them can be: burden minimisation, fit-for-purpose, flexibility and agility.

The operational modalities have to comply with data protection rules and therefore include technological and organisational measures that are proportional to the privacy risk. For certain kinds of privately held data, the degree of granularity and detail of personal information may be higher than more traditional micro-data collections and therefore require higher levels of protection. The adopted privacy protection framework should cover access rights, data format and storage, data location and data movement between stakeholders, use of the data and non-disclosure conditions for the publication of the results. On these aspects, the solutions and protection measures adopted in the context of privately held data might be different from those in place for more traditional data sources, due to the different characteristics and associated levels of risk.

The operational modalities of data reuse also determine the costs and burden on the data holder, but the optimum scenario depends on a multitude of factors and must be assessed case by case. For instance, some data holders may have already invested in building a flexible IT infrastructure and proprietary workflow for data processing. Reusing these for official statistics would require lower marginal costs than setting up a completely new workflow or infrastructure, but could involve fair compensation for the initial investment (possibly including intellectual property rights for proprietary methods).

On the other hand, the need to also apply uniform methodologies and open workflows in the (pre-)processing stage may call for the implementation of additional separate workflows, or components of these, dedicated to statistical production. This would lead to additional development and implementation costs. In such cases, minimising the marginal costs involves a trade-off between reusing existing (possibly proprietary) workflow components and deploying new ones. In other cases, the data holder may not have any legacy infrastructure or processing pipeline in place and would prefer instead to export raw data with minimal or zero pre-processing.

These considerations illustrate how the most suitable operational modality depends on a number of specific elements (technical and business) that vary widely from one data holder to the other, even within the same business sector. Due to such heterogeneity and the complexity of the interdependencies between technical, business and organisational aspects, the task of establishing fit-for-purpose operational modalities that minimise the total costs, risks and burden should be left to **mutual agreement** between the statistical authorities and the data holder concerned.

The next sections discuss four of these dimensions that underpin this ecosystem:

- access
- quality
- roles and responsibilities
- cost.

It is possible to distinguish more aspects of operational modalities, such as:

- accountability
- the level and constraints on collaboration models that are required to be able to extract value (e.g. the need for technical expertise; domain expertise; security concerns arising from broad data sharing and dissemination)
- the extent to which datasets from different partners have to be combined⁽²⁸⁾.

Some of these aspects will be mentioned in the section on business interest considerations.

2.3.2. Data access

Data access underpins several aspects:

- **Time extension** – whether the project is intended for continuous production of statistical products or a one-time effort, for instance for exploring and designing potential new statistical products.
- **Purpose** – conducting research, exploring and designing potential statistical products or producing regular official statistical products will lead to different operational modalities.
- **Data (pre-)processing** – data can be made accessible in raw/granular or (pre-) processed/aggregated form. Access to granular data for statistical authorities would leave more flexibility for experiments and methodological validations and adjustments. Raw/granular data would also give statistical authorities a bigger role in data processing, reducing the burden on the private data holders. Aggregated data, on the other hand, increases the importance of transparency and documentation in the process, to assess the fitness for purpose and quality of the data produced. It would place more responsibilities on the private data holder.
- **Location** – several options exist for data location and storage. Most broadly, these can be divided into on-site and off-site, where off-site can also mean on the premises of the statistical authorities, trusted intermediary or other government body, including research/academic bodies or a government cloud. Advanced privacy-enhancing technologies may be adopted for storing and processing data that requires higher levels of confidentiality. Location consideration has also a cost attribute, meaning that reusing already established infrastructure can decrease costs significantly.

2.3.3. Process quality

To ensure the quality of the process (and thereby also the output), there are four aspects to be considered:

- **Methods** – selecting the optimum methods and identifying best practices is essential. Is there a standard approach or is it fit-for-purpose? Does the data form allow the chosen methods to be implemented, what limitations can occur and what is the impact on output quality? Methodological decisions should be always considered against fit-for-purpose.
- **Input** – input data quality has a significant effect on output quality and should therefore be subject to quality assurance procedures, as set out by the quality assurance framework.

⁽²⁸⁾ See for example page 42 of this report: <https://datacollaboratives.org/existing-practices.html>

- **Throughput** – throughput quality checking has the potential to discover methodological inconsistencies and inherent changes in the data that are revealed after some processing. Similarly to input, throughput is subject to quality checks and quality monitoring and requires a quality assurance framework.
- **Output** – output quality checks are also part of the quality assurance framework and its procedures. Output quality also entails comparability over time and across different processes. It also involves consistency checks with other types of statistical products/indicators that would allow the output to be validated.

While these issues are also in place for traditional data sources (survey, census and administrative records), for which frameworks and practices are already well established in the statistical system, statistical authorities should be aware that the solutions to be adopted for privately held data may be different from legacy ones, considering the specific aspects of such data sources. Therefore, the existing quality framework needs to be reassessed and possibly adapted to deal with privately held data.

2.3.4. Roles and responsibilities⁽²⁹⁾

Roles and responsibilities should be set out for all functions:

- **Privacy protection** – this responsibility can be assigned to the involved stakeholders (statistical authorities, the private data holder or other trusted intermediaries), depending on their other roles in the process and in compliance with existing data protection regulations.
- **Methodology development** – the process of transforming raw granular data into final statistics involves a chain (or pipeline) of data processing operations, from initial stages of data preparation (e.g. selection, cleaning, classification, combination of individual data points) up to more sophisticated statistical methodologies (e.g. aggregation, inference).

Because the input data were not originally produced for statistical purposes, coupled with their very high level of granularity and semantic complexity, the methods and algorithmic solutions to be adopted at the initial data preparation stages may be more complex for privately held data than for other, more traditional data sources. Developing the different stages of an end-to-end processing pipeline may involve different stakeholders with varying levels of intensity. For instance, the detailed definition of the initial data preparation methods may benefit from the technology-specific knowledge of the respective data holder, or other specialised third parties, while the statistical authorities must maintain a predominant if not exclusive role in establishing the final statistical processing stages, and in the overall methodological framework.

In this sense, setting out an end-to-end methodology for reusing privately held data in official statistics may involve some degree of methodological co-development between statistical authorities and data holders, and possibly specialised third parties, with different intensities of involvement by the different stakeholders at different stages. For some components along the methodological chain, methods and algorithmic solutions may already exist and be reused, possibly with some adjustment, or may be developed from scratch. Some methodological components may be directly co-developed, which would allow each partner to access knowledge from other partners that they might not have in-house.

- **Methodology implementation and execution** – once the methodological chain has been fully developed (i.e. specified), it can be implemented and executed in one of three ways, depending on the availability and costs of IT infrastructure and agreement

⁽²⁹⁾ This section does not include data stewardship roles, which are described in section 2.5.

on data reuse arrangements. Implementation and execution can be performed directly by the statistical authorities, the private data holder can be required to do this or it can be outsourced to a third party.

- **Methods auditing and approval** – the statistical authorities should retain responsibility for auditing and approving the methods and solutions for the methodological components for which development and implementation was delegated to the data holder or a third party. In this way the authorities can guarantee compliance with the overall methodological framework and end-to-end quality, even where methodological components are used that were not developed directly by them (e.g. data preparation stages).
- **Integration of multiple data sources** – in cases where the final statistical information must be built by integrating data sources held by different (and possibly competing) private parties, statistical authorities should arrange for adequate protection of business secrets and data confidentiality. To this end, they could consider adopting advanced privacy-preserving technologies. Alternatively, the authorities may decide to use a trusted intermediary.

2.3.5. Cost of reuse

The cost of reusing privately held data for statistics includes the following items:

- **Infrastructure cost** – for statistical authorities, IT infrastructure cost may be reduced by relying on existing infrastructure on the premises of the data holders, when available. However, using existing infrastructure also has a cost for the data holder that may need to be compensated. Using existing infrastructure may also create a dependency, as decisions made by the data holder regarding the IT infrastructure may affect availability for development tools, storage space, system upgrades, etc.
- **Development and implementation cost** – when part of the data processing takes place on the premises of the data holder, the development of data processing software modules involves initial costs. Part of these costs can be saved if the data processing methods that are already in place can be reused, at least partially, for statistical production.
- **Processing cost** – this can be a recurring cost for projects involving regular statistical production over long periods.
- **Maintenance cost** – these costs can be divided into maintenance of the IT environment and updates to the processing software, to track changes in the data. It also includes staff costs for replying to queries about project execution. This will likely take the highest share of the budget for continuous production.
- **Cost for using pre-established data services** – when the statistical authority agrees to use existing data services made available by the data holder, then the service fee already incorporates all the elements described above. Data-as-a-Service (DaaS) or an API based cost model are the typical models adopted by private data holder.

2.4. Business interest

To ensure active support by businesses for data reuse for official statistics, the various dimensions of “business interest” should be considered. Going beyond the minimum safeguard provisions, it would be preferable to establish conditions in which the reuse is perceived by the business as beneficial. In other words, businesses and statistical authorities should aim to

establish partnerships where both parties collaborate to advance knowledge and understanding in society in a way that benefits them all.

This entails (i) balancing the costs versus the benefits; (ii) minimising organisational and societal risks; (iii) maximising the business case behind the partnership; and (iv) enabling data innovation that can be leveraged beyond the initial partnership.

Clauses should be put in place to enable private data companies to maximise the benefits (to their business interests) resulting from the reuse, e.g. in terms of improved public reputation.

Below, a distinction will be made between the risks and the opportunities perceived by businesses.

2.4.1. Risks perceived by business

The 4R framework : Revenue – Risk – Reputation – Regulation can be taken as reference to characterise the risk incurred by business when engaging in data reuse for official statistics.

- Revenue loss and opportunity costs
 - Companies care about revenues generated from their data, and are worried about possible revenue losses and opportunity costs resulting from making statistical information derived from their data publicly available.
 - Such worries can be prevented or at least reduced if the set of statistical products that can be published by the statistical authorities is clearly defined in advance in all dimensions (e.g. spatial resolution, frequency, variables, etc.). In most cases, this will be sufficient to reassure the data holder that the official statistical products will be sufficiently distinct from other “data products” it offers commercially to its customers or beneficiaries. To this end, it is important to define the “portfolio” of official statistical products without ambiguity.
 - Revenue loss can also occur by spending time and resources on preparing the data for the statistical authorities, especially if the expected software, hardware, processes and methodologies are different from the ones already deployed by the data holder. This calls for ensuring that the data holder receives fair compensation for such costs.
- Another source of revenue loss originates from the difference between raw data and contextualised information, as defined by the DIKW pyramid⁽³⁰⁾ (see Figure 2): contextualised data providing usable information requires extensive data processing and quality checks that generate significant and regular costs. The perception of revenue loss may therefore be different when talking about raw data or contextualised information: raw data may be perceived by some data holders as transferable for free, while they will expect compensation for transferring contextualised data.

⁽³⁰⁾ https://en.wikipedia.org/wiki/DIKW_pyramid

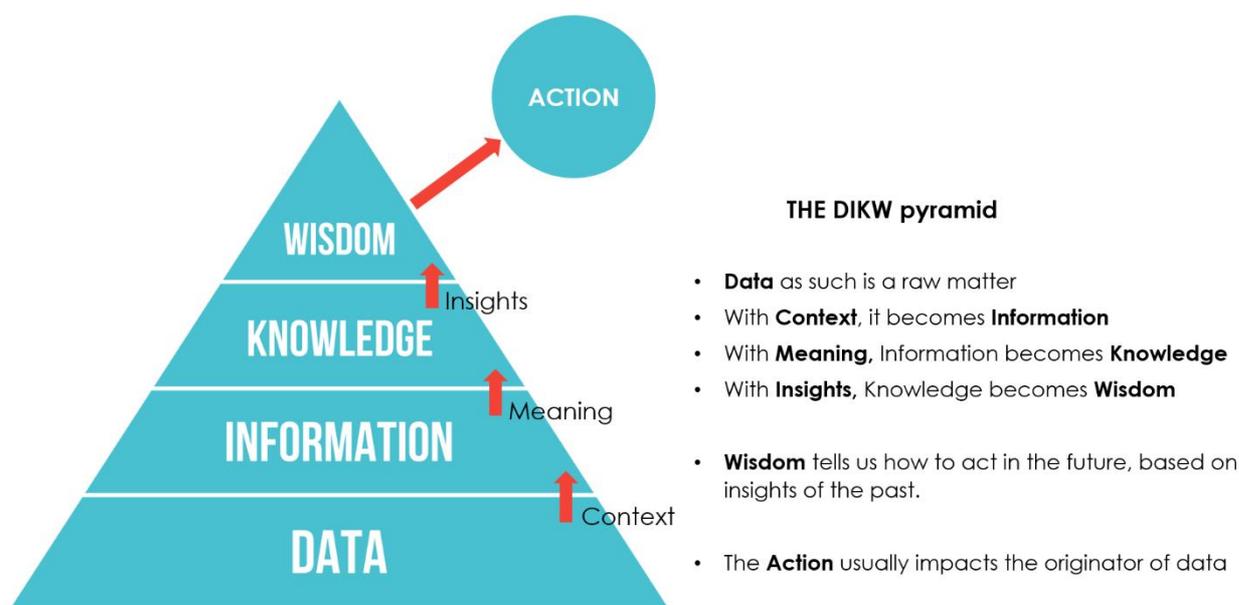


Figure 2: The DIKW pyramid

- Risks of privacy disclosures
 - As defined by Dr Daniel Solove's Taxonomy of Privacy⁽³¹⁾ (George Washington University, 2006), data can be subject to vulnerabilities at 4 different stages of their lifecycle: during their collection, processing, dissemination, or through hostile invasions (see Figure 3).
 - Transferring privately held data to governments firstly raises concerns of dissemination vulnerabilities for the private data holder that may lead to a breach of confidentiality, disclosure, exposure, increased accessibility, blackmail, appropriation, and distortion.
 - The processing of the data by the statistical offices, and especially the blending of the data with external sources, then raises concerns about processing vulnerabilities: aggregation, (re-)identification, insecurity, secondary use, and exclusion.
 - Finally, because statistical offices are government bodies, concerns about information collection vulnerabilities may arise, leading to surveillance and interrogation risks.
 - By disseminating the data, private data holders lose control of it. They worry about their responsibilities and accountability in the event of problems. Who is responsible? Who is accountable? Who is liable?

All these risks can be addressed by adopting strong protection measures at all stages of the data reuse cycle. Privacy-enhancing technologies are a particularly appealing solution in this context as they can enable pre-defined computation on confidential data without sending them to any other party. In this sense, they allow the data holder to maintain control over the reused data and reduce the (actual or perceived) risks of undue surveillance.

⁽³¹⁾ <https://www.privacysecurityacademy.com/wp-content/uploads/2018/02/Handout-Foundations-and-Themes-Professor-Soloves-Taxonomy-of-Privacy-01.pdf>

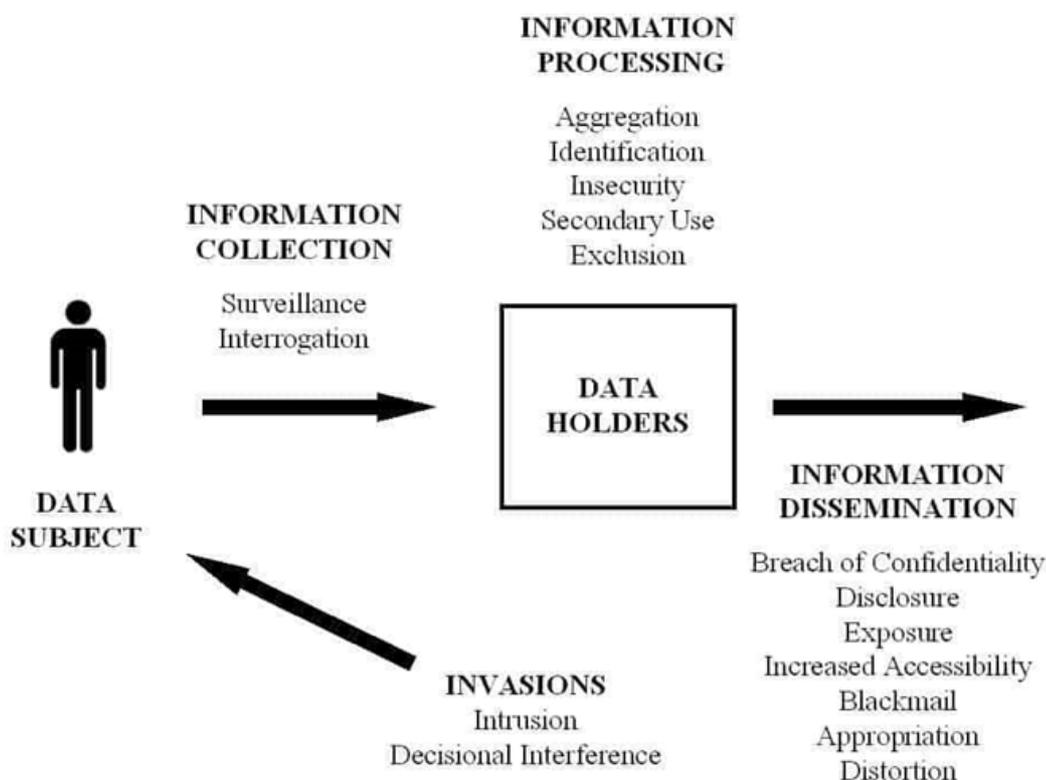


Figure 3: Solove's taxonomy of privacy

- Reputation
 - How would the general public, and more particularly the data holder's customers, react in the event of data leaks, or to the simple fact that they share their data with governmental offices? The particular political situation in a Member State could increase the perception of this as a problem. Again, adopting advanced privacy-enhancing technologies could be part of the solution.
- Regulation
 - Each new regulation initially brings an extra compliance burden for businesses. It is only after compliance has been successfully achieved that the regulation is seen as a competitive advantage or a basis for fair competition. The regulation may introduce a uniform and transparent set of rules for all stakeholders, which would be beneficial for businesses.
 - Similarly, a new regulatory framework on the reuse of privately held data for public purposes, and specifically for official statistics, may influence investment in negative or positive ways. For instance, it may lead to decreasing investment, if business model assumptions are negatively influenced or to increasing investment, if the framework provides for long-term certainty.

2.4.2. Opportunities perceived by business

Opportunities can be represented by the 9R framework⁽³²⁾, as defined by The GovLab (see Figure 4).

Business Case for Data Reuse in the Public Interest

Andrew J. Zahuranec, Andrew Young, and Stefaan Verhulst

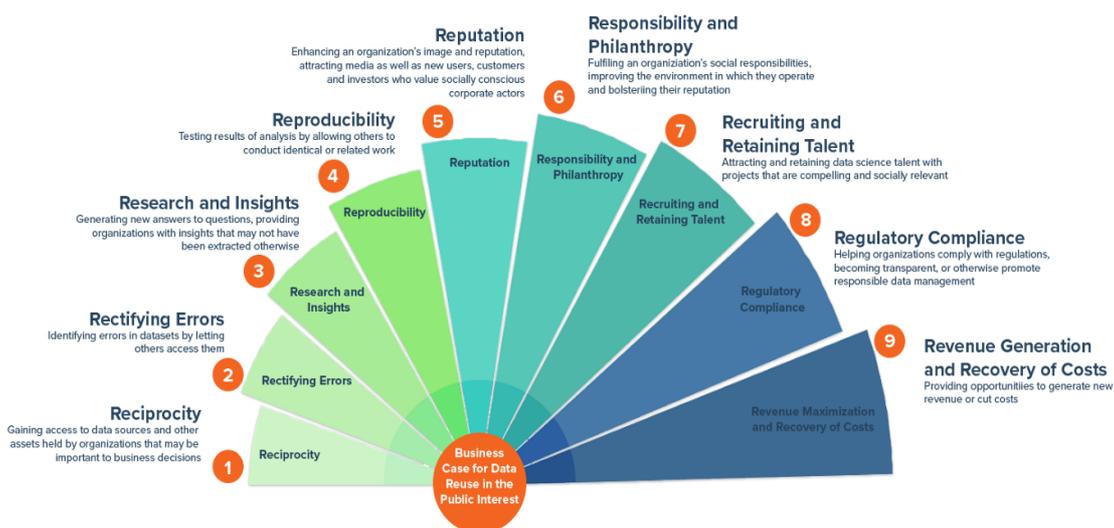


Figure 4: The 9Rs framework: the business case for data reuse in the public interest

- **Reciprocity**
By sharing their data, private data holders can gain access to data sources and other assets held by organisations whose data may be important for their own business decisions. For example, access to statistics generated by the statistical authorities based on the data from several (and possibly competing) private data holders is beneficial for all businesses.
- **Rectifying errors**
When data are reused for statistical production, the private data holder gains the chance to have trusted specialists from statistical authorities assessing (directly or indirectly) the quality of the data. With appropriate feedback, this could lead to error rectification and data quality improvement, with mutual benefit for both parties.
- **Research and insights**
Private data holders are always searching for insights hidden in their data. Letting other people access, use and reflect on their data could generate new answers to questions, and provide organisations with insights and ideas that may not have otherwise been found or thought of.

⁽³²⁾ <https://medium.com/data-stewards-network/the-9rs-framework-establishing-the-business-case-for-data-collaboration-26585455ccc0>

- **Reproducibility**
In the same way that science gains credibility by allowing different teams to reproduce experiments and replicate the same result independently, by sharing their data, private data holders can let others conduct identical processes on their raw data and verify (or correct) their conclusions, increasing their data reliability.
- **Reputation**
Sharing data can enhance an organisation's image and reputation, attract media, new users, customers, and investors who value socially conscious corporate actors.
- **Responsibility and philanthropy**
Sharing data for the public good can fulfil an organisation's social responsibilities, improve the environment in which it operates, and bolster its reputation.
- **Recruiting and retaining talent**
The visibility acquired by sharing privately held data can also increase the willingness of skilled data scientists to sign up and stay on board for projects that are compelling and socially relevant.
- **Regulatory compliance**
Preparing the data for sharing can help organisations comply with regulations, become more transparent or otherwise promote responsible data management.
- **Revenue generation and recovery of costs**
Under the right conditions, sharing data can provide opportunities to generate new income, and cut or share costs.

2.5. Data stewardship roles

Recent policy developments in Europe, such as the new Data Governance Act and forthcoming Data Act, call for new institutional arrangements and practices, such as data intermediaries, data spaces and data altruism.

In this context, data stewards can play a key role, in both the public and private sectors, by facilitating data collaboratives and partnerships focused on providing functional access to and re-usability of data in the public interest. Statistical authorities need counterparts in the private sector and civil society organisations, in order to establish partnerships that are systematic, sustainable and responsible.

Data stewards can help us reach fit-for-purpose solutions at technical level as well as at business and governance operational levels. The data stewardship function can be characterised by the set of functions and skills that enable and promote access to and reuse of data for official statistics and/or non-commercial research in a systematic, sustainable, and responsible way:

- **systematic** – going beyond pilots and leveraging a tested methodology for accessing and reusing data;
- **sustainable** – understanding the costs and benefits of data collaboration to inform long-term sustainability;
- **responsible** – preventing mis-use but also missed-use of access to/reuse of data.

Five types of stewarding can be distinguished, with the following roles and powers⁽³³⁾:

1. Stewarding access to and reuse of data for official statistics and research.

This involves:

- Helping determine and assess the purposes for which access to data is needed (e.g. for research or official statistics);
- Facilitating reuse (or secondary use), which may involve acquiring informed consent (e.g. when dealing with sensitive data) as well as acquiring a social licence for reuse through other means;
- Determining and overseeing access permissions (potentially for third-party data such as privately held data) through accreditation or other means;
- Helping re-publish what is already openly available (role of official statistics as trusted intermediary);
- Conducting risk-management, for instance confidentiality for businesses and other risks;
- Establishing and aligning data documentation, including management of metadata, and the responsibility to proactively inform data users and other stakeholders when definitions change;
- Ensuring data quality and integrity that is fit-for-purpose. This may include both supportive information on how data quality is measured and/or to help establish quality indicators for data;
- Ensuring timely and sequential release of data;
- Considering the FAIR⁽³⁴⁾ principles;
- Aligning standards and definitions (interoperability);

2. Stewarding partnerships and community engagement.

This comprises:

- Putting “citizens at the centre”, enabling a more engaged relationship (see also section 2.2);
- Identifying, vetting and engaging with possible partners and other stakeholders (including reviewing research institutions and research proposals);
- Engaging key stakeholders for data products and insights, in a user-driven process (see also section 2.2);
- Achieving interoperability through coordination and collaboration (within standard-setting processes);
- Establishing data (sharing) agreements and other contractual relationships;

⁽³³⁾ For more background on data stewardships, see:

<https://thegovlab.org/static/files/publications/wanted-data-stewards.pdf>

<https://medium.com/data-stewards-network/data-stewardship-re-imagined-capacities-and-competencies-d37a0ebaf0ee>

<https://studylib.net/doc/8219376/the-ibm-data-governance-council-maturity-model--building-a>

<https://hnu.edu/wp-content/uploads/2020/03/Data-Governance-Maturity-Model.pdf>

⁽³⁴⁾ The FAIR guiding principles for scientific data management and stewardship,

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4792175/>

3. Stewarding internal resources, stakeholders, and expertise.

This includes:

- Gaining approval from and coordinating with actors within organisations (such as public institutions, research centres, universities);
- Communicating with all internal stakeholders and organisational leadership;
- Establishing a supervisory role on how data is being accessed and reused;

4. Stewarding sustainability.

This encompasses:

- Institutionalising data innovation and access to privately held data for official statistics;
- Developing the business case for data sharing and reuse (see also section 2.4);
- Measuring impact and acquiring funding (from public and/or private sources, and also from grant funds with this purpose);
- Establishing data pools for data lifecycle management.

5. Stewarding insights.

This covers:

- Ensuring the timely release of insights;
- Raising awareness of insights (and the reuse of data) with users, partners, government, and other stakeholders;
- Closing the feedback loop with users and citizens.

3

Principles for the reuse of privately held data for official statistics

In line with the Expert Group's mandate, the focus of this chapter is not on principles for official statistics in general, but on the reuse of privately held data for such statistics. Although the Expert Group proposes a number of specific principles, there are also generally applicable principles that need to be set out or adjusted to apply to the reuse of privately held data for official statistics. Section 3.1 sets out seven such generally applicable principles, their relevance for the area the Expert Group has dealt with, and a proposal for their application to that area. Section 3.2 lists a further eight specific principles regarding the reuse of privately held data for official statistics, and Section 3.3 describes two final principles specifically about enablers for the wider data ecosystem.

3.1. General principles

1. **Principle of official statistics as a public good.** Official statistics constitute a public good⁽³⁵⁾. As such, official statistics are to be developed, produced and disseminated by statistical authorities in compliance with established statistical principles and quality standards (see section 1.1.1). This foundational principle applies to all official statistics regardless of the kind of data taken as input for their production. It applies to data collected directly by statistical authorities as well as to data collected by other entities and then reused by statistical authorities, including privately held data.
2. **Principle of stakeholder involvement.** The process leading to the selection of statistical products and questions to be addressed must involve all relevant stakeholders sufficiently. This principle is currently implemented mainly through comprehensive, mostly formal consultation mechanisms, resulting in justified decisions by ESS governance bodies. If privately held data is to be reused on a large scale, however, there is an even greater need for public acceptance and increased responsiveness to new demands for official statistics. Such acceptance and responsiveness could be achieved by a three-tier approach, as shown by Figure 5:
 - a. **Principle of participation.** Inclusive and open participatory mechanisms should be encouraged in order to identify the high-value questions that should be addressed by the reuse of privately held data. This principle of participation, illustrated by the bottom layer of Figure 5, implies that decisions on what official statistics aim to measure by means of privately held data should be based on stronger engagement with citizens – often the subject of the data that is reused – and users of official statistics (e.g. policymakers, businesses and researchers). Domain experts should also be involved. On top of already existing mechanisms and structures for stakeholders' participation, instruments

⁽³⁵⁾ https://unece.org/DAM/stats/documents/ece/ces/2017/CES_4-Value_of_Official_Statistics_for_endorsement_for_upload.pdf

like citizen assemblies and public consultations could facilitate better understanding of public expectations and opinions on what it is important to measure and why it should be measured by means of privately held data. Thanks to these additional instruments, statistical authorities can more readily identify information gaps and anticipate new emerging information needs⁽³⁶⁾. Most importantly, statistical authorities can enhance the legitimacy of accessing a wider base of citizen data and gain public acceptance ('social license') for the reuse of privately held data for statistical purposes⁽³⁷⁾.

- b. **Principle of consultation.** Formal consultation mechanisms with balanced representation of all stakeholder interests should also inform the selection of statistical products based on the reuse of privately held data. Such mechanisms require careful consideration. The middle layer of Figure 5 illustrates the principle of consultation, involving official representatives of main user groups including representatives from business associations, citizen associations, research communities and policymakers. The decision process that defines the scope of official statistics weighs up the societal importance of the statistical information to be produced against the practical feasibility and prospective costs of producing it with the required continuity and quality standards, bearing in mind also the availability and maturity of data sources. In the ESS governance framework such assessment is carried out with the support of formally appointed consulting and advisory committees such as the ESAC⁽³⁸⁾ and ESGAB⁽³⁹⁾. With the prospective systematic involvement of privately held data, the composition and working mechanisms of these committees may need to be revised within the existing governance framework to take better account of the specific aspects and stakeholder interests associated with such data, for instance for greater involvement of representatives from the relevant business sectors.
- c. **Principle of responsiveness.** The definition of European statistics based on privately held data requires an agile decision process. The ESS should adapt to the needs of a changing society by taking as input the outcome of participatory and formal consultation mechanisms and working in a flexible and responsive way. This is shown in the top layer of Figure 5. The need to serve the increased demands of a more complex and faster-changing society will motivate statistical authorities to use more timely and detailed data on the input side, such as those held in the private sector, at the same time increasing the agility of the procedures that ultimately define the portfolio of statistical products on the output side. The impact, benefits and costs of the statistical products need to be periodically reassessed to identify outdated products that should be modified or purged, in addition to new products that should be added, thus keeping the statistical portfolio always relevant and up to date.

⁽³⁶⁾ This is in line with Chapter I 'Putting people at the centre of the digital transformation' and Chapter IV 'Participation in the digital public space' of the Declaration on European Digital Rights and Principles, <https://digital-strategy.ec.europa.eu/en/library/declaration-european-digital-rights-and-principles>

⁽³⁷⁾ Public institutions (such as public libraries) can also help to facilitate outreach and deliberation. Innovation in consent management can also be used to foster participation. Participation needs to be representative and free from special interests that may potentially dominate the participatory process (e.g. fringe groups). Participation should acknowledge diversity and aim at inclusion of all citizens without discrimination of any sort.

⁽³⁸⁾ <https://ec.europa.eu/eurostat/web/european-statistical-system/governance-bodies/esac>

⁽³⁹⁾ <https://ec.europa.eu/eurostat/web/european-statistical-system/governance-bodies/esgab>

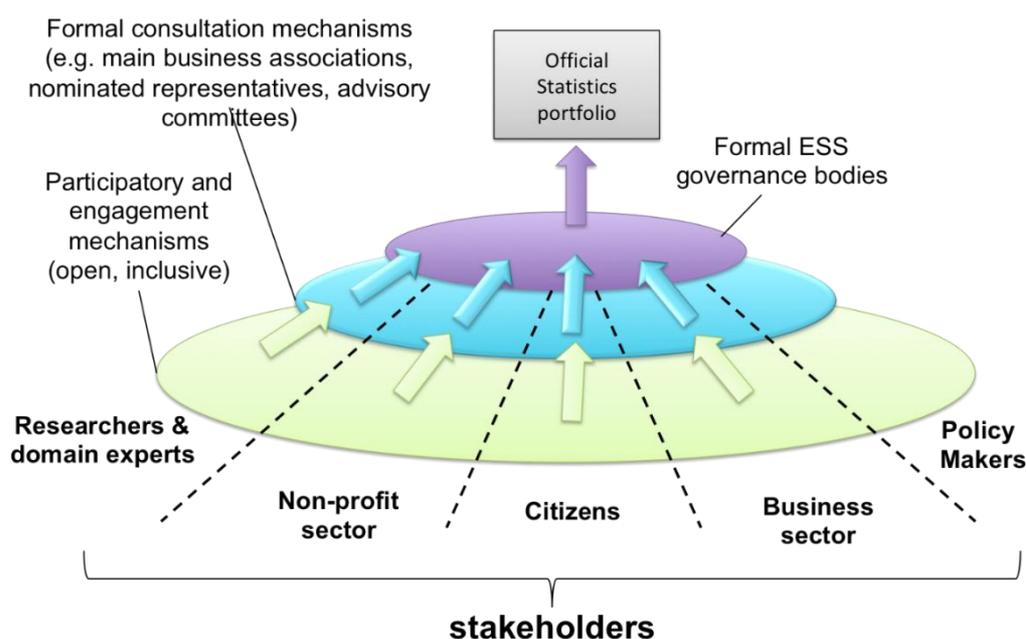


Figure 5: Representation of the decision process leading to determination of the official statistics portfolio and selection of priority areas.

Legitimacy and public acceptance are achieved through participative mechanisms and systematic consultation with relevant stakeholders.

3. **Principle of data integration.** Integration of different data sources adds value to official statistics. This principle is true for traditional data sources, but is even more important if privately held data can be reused for statistical purposes. The range of potential data sources that are relevant for official statistics comprises (i) data collected directly by the statistical authorities (e.g. survey and census data), (ii) external data generated by public sector bodies (e.g. administrative records, open government data), and (iii) external data held by businesses and civil society organisations. A single product in the portfolio of official statistics might require the combination or integration of different input data sources (multi-source statistics) and a single data source might be reused for multiple statistical products (multi-use data). Multi-source statistics, particularly statistics that are produced by statistical authorities based on the combination of privately held data with more traditional data sources (e.g. survey data), represent one additional added value of official statistics over the statistical information offered by other sources. The integration of multiple sources mitigates the weaknesses of each individual source and increases overall quality. Privately held data should be seen as one key ingredient among others in the production process for official statistics. Furthermore, if data from a single source is to be reused by multiple statistical authorities, the modalities of data reuse should be coordinated and organised with a view to minimising the burden on the data holders.
4. **Principle of sufficient product maturity.** The portfolio of official statistics should allow for the inclusion of products of varying methodological maturity. For traditional statistics, provisions for candidate products will be needed only in exceptional circumstances, since new statistics can in most cases be based on well established methods that need no further experimentation. However, for statistics based on privately held data the situation is completely different. Due to their complexity and fast-changing nature, dealing with privately held data requires more flexible and dynamic management of the

portfolio of statistical products. Such a portfolio should include mature statistical products for which established statistical methodologies are available for regular statistical production, possibly with incremental adaptation and limited adjustment, but also new candidate products for which more fundamental methodological research and development is still needed. Candidate products may be subsequently promoted to the class of mature products or dropped, depending on the outcome of the research and development activities. The legislative framework should support the research and development of new statistical products at all stages, including enabling the dissemination of so-called experimental statistics if this is deemed useful by the statistical authorities, for instance to collect early feedback by prospective users.

5. **Principle of data protection.** Data protection and security are fundamental requirements for all official statistics. Protection of data confidentiality is a core tenet of the established statistical principles and quality standards in official statistics. The same principle holds when dealing with privately held data. In particular, the agreed modalities of data reuse should comply with data protection legislation, that is, GDPR and other sectorial legislation at national and European level as relevant in each business sector (e.g. ePrivacy for telecom data). The assessment and mitigation of data confidentiality risks should be pursued with state-of-the-art techniques. As protection measures must be commensurate with the level of risk, statistical authorities may need to put in place stronger technical and organisational measures for the reuse for statistical purposes of certain kinds of privately held data involving extra high data confidentiality risks, particularly where very detailed behavioural data are concerned. To deal with such cases, statistical authorities should be encouraged to consider using novel advanced technologies from the field of privacy-enhancing technologies.
6. **Principle of open science.** Official statistics should benefit from the best practices of open science. This principle is especially important when privately held data is reused. The reuse of privately held data for official statistics should take place in a way that ensures full methodological transparency and reproducibility. Whenever possible, the use of proprietary modules in the production of official statistics should be avoided and the use of open-source code and open algorithms promoted. If privately held data are to be used in official statistics, the European Statistics Code of Practice (see section 1.1.2) should be revisited and possibly adapted to better address the new methodological issues entailed by such data – including the possibility of carrying out part of the data (pre-)processing and aggregation at the premises of the data holder – and at the same time improve alignment on the latest best practices in open science and on the FAIR principles⁽⁴⁰⁾, even if there are differences between the two domains of science and official statistics.
7. **Principle of public engagement.** Statistical authorities should actively engage with the public by fostering public literacy and communicating transparently on official statistics. This principle is particularly important in the reuse of privately held data, since this is a complex domain involving various interests. The statistical authorities should provide detailed information to the public on what data is reused and for what purposes, as well as on how it is used, why it is needed, how confidentiality is ensured, how data is processed, etc. Transparency is necessary to build trust and acceptance by the public on how the reuse of their data for statistical purposes actually takes place. Furthermore, and along with the dissemination of statistics, statistical authorities have an educational and knowledge-building role to play on how data, information and knowledge relate to each other in modern society. This role is all the more important given that data which is reused for public purposes is primarily collected for business purposes. Public

⁽⁴⁰⁾ The FAIR Guiding Principles for scientific data management and stewardship, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4792175/>

engagement efforts should reach all citizen groups, irrespective of their social characteristics.

3.2. Principles concerning the interaction between private data holders and statistical authorities

8. **Principle of collaboration.** If specific conditions are met and appropriate safeguards are in place, data holders should be legally required to collaborate with statistical authorities for the implementation of the official statistics portfolio. Collaboration with statistical authorities includes the requirement to enable data reuse – with mutually agreed operational modalities – for the development and regular production of official statistical products that cannot be reasonably produced by other means. Statistical authorities can reuse the data solely for the purpose of developing, producing and disseminating the statistical products specified in the official statistics portfolio as defined based on a formal process fulfilling all other principles. Introducing a legal requirement provides a common baseline ensuring equal treatment of private data holders in respect of costs and benefits (level playing field), and at the same time prevents undue dependence of statistical authorities on quasi-monopolistic data holders in highly concentrated business sectors. It also ensures the continuity of statistics that are dependent on reuse of the privately held data. The legal requirement to cooperate with statistical authorities should be complemented by a clear set of limitations, conditions and safeguards protecting the data holders from undue interference with their legitimate business interest.
9. **Principle of proportionality.** The requirements for data holders must be reasonable and proportionate. The societal importance of information needs to which official statistics aim to respond must be properly weighed against all costs associated with fulfilling those needs. Where privately held data are involved, the burden on and risks to the legitimate business interests of the data holder should be part of the equation. Data holders should not be subject to unreasonable burdens and constraints without fair compensation, particularly when the operational modalities for data reuse imply costs for the data holders. Data holders cannot be made to collect new additional data for statistical purposes beyond what they already collect for their own purposes. At the same time, statistical authorities must not be made to pay for the data as such.
10. **Principle of agreed modalities.** Data reuse modalities should fulfill a set of conditions and should be laid down by mutual agreement between data holders and statistical authorities. The technical and operational modalities of data reuse may vary with the specific context. Modalities may, for instance, entail the provision of direct access to pre-processed data or establishment of an end-to-end data service. The operational modalities of data reuse must be fit for purpose and must: allow for the sustainable production of official statistics; minimise the total burden, especially the burden to the private data holder; minimise costs; minimise the risks to data confidentiality and business interests; be agile in responding to changes in the production pipeline; support the application of appropriate data processing methodologies and comply with the principles adopted in the European Statistics Code of Practice, in particular ensuring that statistical quality targets are met. The agreed-upon data reuse modalities for the regular production of mature statistical products may differ from those enabling research and development for new candidate products. Relying on mutual agreements provides the necessary flexibility to adapt the operational modalities to the particular situation and ensures that all interests are properly taken into account. Due process

must apply to the establishment and implementation of mutual agreements. This guarantees that all safeguards that apply, such as on proportionality, are enforced in practice, for instance by explicit assessments. An important element of the process is the possibility of invoking dispute resolution procedures if the statistical authorities and the data holder do not agree on the data reuse modalities.

11. **Principle of appropriate methodologies.** The reuse of privately held data for official statistics must be based on data processing methodologies that are fit for purpose. Transforming raw data into final statistics involves a chain of data processing steps, from simple data preparation (or pre-processing) to the computation of the final indicators of interest, that collectively form an end-to-end statistical methodology. Data preparation is included in this methodology, since it typically has a non-negligible impact on the quality of the final indicators. Due to the typical data volumes and rates at play where privately held data is concerned, data processing is necessarily automated, hence the statistical methodology must be translated into software code. Since privately held data is generated for a primary purpose other than statistics (e.g. delivery of a private service), often involving technologies and systems that fall outside the typical skill set of professional statisticians, the definition of a complete end-to-end methodology would benefit from contributions by experts and specialists from the relevant fields (with the statistical authorities retaining final responsibility for the methodology applied and the quality of the resulting statistical products). In this sense, the methodologies for processing privately held data may be co-developed by the statistical authorities in partnership with data holders and other specialised third parties. The methodological solutions of choice must comply with data protection and privacy requirements. They should be transparent, should not contain black-boxes with no guarantee of quality, and should be auditable by peer experts.
12. **Principle of fair cost treatment⁽⁴¹⁾.** Costs incurred by data holders must be dealt with in a fair way. Although statistical authorities do not pay for data as such, financial instruments (e.g. public funding or fiscal incentives) can be made available by public institutions to cover the marginal costs of reusing privately held data if the agreed-upon data reuse modalities involve resource-intensive tasks for the data holder (e.g. support in methodological co-development activities, computationally complex preparation/processing of data). A mechanism should be in place to establish fairly the marginal costs of the agreed-on data reuse modality. In particular situations, statistical authorities may use data processing services offered on a commercial basis.
13. **Principle of business safeguards.** Appropriate safeguards must be in place to protect business secrets and interests. Statistical authorities are bound to protect the business secrets of the contributing data holders. Business data and information acquired by statistical authorities from private data providers cannot be used for any other purpose than the production or development of official statistics. Compared to more traditional business data (e.g. from surveys and aggregate financial reporting), the reuse for statistical purposes of very detailed privately held data poses increased risks for business confidentiality. Accordingly, proportionally stronger protection measures, including state-of-the-art privacy-enhancing technologies, should be put in place to ensure compliance with the business safeguards principle. Business sensitive areas and risks can be mapped in a dialogue between the data holder and the statistical authorities.
14. **Principle of consent for further reuse.** Any reuse of privately held data by statistical authorities beyond the production of official statistics should be based on agreement with the data holder. The requirement for data holders to enable data reuse for statistical purposes should be strictly limited to the development, production and dissemination of the official statistics portfolio (in accordance with the previous

⁽⁴¹⁾ This principle is in line with the current Data Act proposal.

principles). Any further reuse of such data should remain subject to voluntary agreement with the respective data holders⁽⁴²⁾. Data holders may legitimately deny consent to further data reuse if this involves serious risks for their business (e.g. loss of business opportunities, leakage of business secrets and intellectual property, reputational risks) or for their customers or beneficiaries. Moreover, statistical authorities should not be allowed to pass on the data acquired by private companies or civil society organisations to other entities, including other public authorities and research organisations, without previous agreement with their respective providers.

15. **Principle of appropriate standards.** The reuse of privately held data should be based on technical and methodological standards and on a quality framework that is fit for purpose. Standardisation of methodologies and statistical concepts should be pursued to the maximum possible extent for data sources of similar characteristics⁽⁴³⁾. The quality of the statistics concerned also stands to benefit from standardisation. Quality assurance of the entire statistical production process should be in place, overseen by the relevant statistical authorities within a standardised statistical quality assurance framework suited to dealing with the specific aspects of (the different kinds of) privately held data⁽⁴⁴⁾. The quality assurance framework should address, among other things, the issues of continuity in data availability; potential bias in the data and algorithms; and the environmental impact of data storage and processing⁽⁴⁵⁾. In cases where the agreed-upon reuse modalities provide that some part of the data processing takes place at the site of the data holder, it should be possible for statistical authorities to audit the production process on site. The mutual agreements on reuse modalities must also cover the reporting of any supplementary information that is necessary for statistical authorities to assess and optimise the quality of the statistical process and products (e.g. data description, information about the occurrence of events that may affect the availability or reliability of data, relevant changes in data collection mechanisms).

3.3. Enablers

16. **Principle of professional data stewardships.** Organisations should be encouraged to put in place data steward functions that are guided by professional principles. This principle refers to private as well as public organisations. Clear professional principles for data steward functions should be established and expressed through a code of practice (possibly inspired by the European Statistics Code of Practice) and should focus on, for instance, transparency, responsibility and accountability, public value and proportionality, confidentiality, and ethics. Transparency would refer to, among other things, public documentation of data handling (what, how, why). Responsibility and accountability could pertain to compliance with existing rules and procedures and ensuring a timely response to a request for data. Public value and proportionality may focus on public value creation and ensure that the burden aligns with the value to be generated. Confidentiality and ethics should aim at maintaining high standards of confidentiality and responsibility, and proactively engaging data subjects if agreements

⁽⁴²⁾ This is in line with current legislation, including the Data Act proposal.

⁽⁴³⁾ Methodological standardisation in official statistics would be likely to have a positive spillover impact also in the production of other statistics by the data holder or third-party companies, and in this sense should be seen as a beneficial element also for the business.

⁽⁴⁴⁾ The quality assurance framework that is currently in place in the ESS, which was developed with a focus on data collected primarily from surveys and administrative records, may need to be revised and extended to deal with privately held data reuse.

⁽⁴⁵⁾ This is in line with Chapter VI 'Sustainability' of the Declaration on European Digital Rights and Principles,
<https://digital-strategy.ec.europa.eu/en/library/declaration-european-digital-rights-and-principles>

and expectations change. A peer network of data stewards may be set up promoting the establishment of public-private partnerships and promoting peer learning over time.

17. **Principle of statistical data stewards.** Statistical authorities should be encouraged to serve as data stewards for the public sector. Statistical authorities in Member States as well as Eurostat may become more proactive in the new, complex data ecosystem. They may further develop their responsibilities in the direction of supporting other public authorities in transforming data into trustworthy information, effectively serving as data stewards for the public sector.

4

Recommendations on the reuse of privately held data for official statistics

Most principles pertaining to the reuse of privately held data for official statistics set out in the previous chapter require actions by the relevant stakeholders. This chapter lists a number of recommendations for key stakeholders. Some recommendations are addressed to a single stakeholder or group thereof, while others have more general relevance. Some recommendations call for concrete short-term action, while others are more strategic and long term. Consistent with the scope of the Expert Group, all recommendations focus on the reuse of privately held data for official statistics: some of them are intended to reinforce, adapt or extend rules and practices that are already in place in the statistical system, while others point to new actions. Taken collectively, the actions resulting from these recommendations are essential to effectively implement the principles laid down in the previous chapter. Some recommendations are already being taken up by the ESS.

4.1. Engaging with citizens and businesses

1. Statistical authorities should strengthen participatory and public consultation mechanisms to identify what to measure, using privately held data sources. Statistical authorities should seek to strengthen the participatory and consultative mechanisms in the direction of a stronger dialogue with and engagement of citizens and business associations in identification of the high value questions and the statistical products of high societal importance that warrant reuse of privately held data sources. The approaches to be implemented may be based on the involvement cycle and mechanisms set out in Section 4.2 and could include public consultations⁽⁴⁶⁾.
2. Statistical authorities should strengthen their educational role and improve public data literacy in connection with the reuse of private data for public purposes. The actions envisioned should target citizens as well as businesses. The ESS should promote the building of a comprehensive representation of data value and a shared vision that leverages the interest of private companies to be recognised as contributing to the common good. The reuse of privately held data for the public benefit, and specifically for official statistics, is thus presented as an important mission of responsible businesses.

⁽⁴⁶⁾ In analogy to the public consultations that are adopted at various stages of EU law and policy-making process, see <https://ec.europa.eu/info/about-european-commission/service-standards-and-principles/transparency/consultations>.

4.2. The approach to data

3. Statistical products and processes involving the reuse of privately held data should be strongly harmonised across the EU and globally. A high level of harmonisation across the EU is needed to achieve efficiency and effectiveness in the reuse of privately held data, contributing to limiting or even reducing the burden on businesses. Statistical products, methodologies, conditions for data reuse and statistical processes should be aligned across the EU. Furthermore, Eurostat should promote the definition and adoption of global (worldwide) statistical concepts, methods, classifications, and other standards to ensure cross-country comparability of products and processes relevant for private data sources. In that context, the collaboration of Eurostat with other international organisations involved in the global statistical system should be continued and further strengthened.
4. Security and data confidentiality should be safeguarded with the strongest available technical and organisational measures when reusing privately held data. Data confidentiality, both in terms of personal data (privacy) and business sensitivity, and protection against data misuse (including guarding against potential government intrusion) should be ensured by the strongest available measures. In particular when massive amounts of very detailed behavioral data are involved, the risks (actual or perceived) might be considerably higher compared to more traditional data (e.g. survey data and administrative records) and the safeguards must be proportionally stronger. Adoption of advanced privacy-enhancing techniques and other state-of-the-art technologies should be encouraged to strengthen organisational measures and data governance policies, especially when the statistical process involves data processing across multiple organisations (e.g. one or more data holders and the statistical authority). A framework should be developed for assessing the impact and residual risk of different data protection techniques on chosen methodologies.
5. The ESS frameworks for quality management should be reassessed and possibly adapted to better deal with the distinctiveness of privately held data sources. This recommendation pertains to the European Statistics Code of Practice and the quality assurance framework as explained in the first chapter. The solutions and rules currently applied by statistical authorities in dealing with traditional data sources (survey, census and administrative records) may not necessarily fit or suffice for the reuse of privately held data. The reuse of data produced and held by businesses involves distinctive quality aspects and specific issues that need to be addressed by the Code of Practice and quality frameworks. These include, for instance, the application of new analytic and modelling methods, the need to reduce dependencies on individual data providers, strengthening of security and privacy, the application of open science principles such as the FAIR principles and ensuring transparency and full reproducibility of the whole statistical production process when detailed data is taken as input instead of declarations and responses by the data subjects. The quality framework should provide for an official process for data holders to notify statistical authorities when there are changes in the data source, data pre-processing or data gathering methods that might influence the quality of the statistical product. Finally, since statistical quality ultimately depends on the quality of the input data, a European reference model for assessing the level of maturity and suitability of privately held data for reuse in official statistics should be developed and applied in the process of defining new candidate statistical products. Taken collectively, these actions are necessary to ensure the reliability and trustworthiness of the final statistical information.
6. **A methodological framework for the reuse of privately held data should be developed by the ESS.** A methodological framework should be developed specifically for the reuse and integration of privately held data in official statistics. In this endeavour,

statistical authorities should take a leading role with active coordination by Eurostat. Such a framework, of which quality considerations would be a core tenet, could have a potential impact also outside the scope of official statistics (e.g. for research or for the production of commercial statistics by private companies or civil society organisations). The framework should be open for input from the research community and provide a basis for better aligning official statistics on research and methodological developments in neighbouring fields. It would ease the integration of already developed methodological solutions into the future production process for official statistics.

7. Statistical authorities should be open to contributions from the private sector and civil society organisations for the development of data processing methods for privately held data without compromising on transparency. The methodological and quality frameworks should aim at maximising methodological transparency while taking into account that some aspects of the input data and processing operations, particularly at the initial stages of data (pre-)processing, may depend on proprietary technologies, involve business secrets, require highly specialised technological knowledge and may be highly specific to some particular data holder. In such cases, statistical authorities should consider the possibility of incorporating methodological solutions developed by the data holders or by other specialised third parties into official statistics production processes, in agreement with the methodological and quality assurance framework, if that contributes to the feasibility and timeliness of the statistical development and production process. Any data (pre-)processing solution developed by other parties should in any event fulfill the requirements laid down for transparent methodology and ensure the necessary quality level for official statistics. To deal with such cases, statistical authorities should put in place an auditing framework for vetting data-processing methods developed by other parties for their suitability for official statistics production. In this way, experienced data holders and specialised third-party companies can contribute to the development and implementation of better statistical methodologies under the supervision of statistical authorities.

4.3. Balanced partnerships

8. Statistical authorities and private data holders should develop a partnership approach geared towards maximising business incentives and minimising risks, based on mutually agreed operational modalities of data reuse. To gain the support of businesses, enabling conditions and incentives should be put in place to promote perks and opportunities and minimise perceived risks in the reuse of privately held data for statistical purposes. As the risks, costs and liabilities depend heavily on the operational modalities via which data reuse is concretely implemented, it is important that the operational modalities of data reuse are set by mutual agreement between the statistical authority and the data holder concerned. The data reuse agreement between the data holder and the statistical authority should clearly identify (among other things) the roles, responsibilities and liabilities of each party. Data reuse modalities need to be designed in such a way that the burden, risks and liabilities for data holders are minimised, while ensuring sustainable statistical production with the target levels of quality (fit-for-purpose). Special attention should be given to reducing the marginal risks and liabilities for data holders associated with data confidentiality⁽⁴⁷⁾. Standard procedures and model agreements (including e.g. official data request forms with a clear description of all required technical specifications, data and metadata details,

⁽⁴⁷⁾ In fact, breaches of data confidentiality are particularly dangerous for the business (e.g. in terms of reputation loss vis-à-vis their customers in the event of personal data breach, and direct leakage of sensitive business secrets).

requested pre-processing operations, responsibilities, etc.), possibly harmonised at EU level, should be developed to ease the setup of successful partnerships agreements.

9. Statistical authorities should put in place non-financial incentives and enabling conditions to facilitate the setup of successful partnerships with private data holders. This can be done in the following ways:
 - a. Statistical authorities should be encouraged to propagate useful information back to the partner data provider, through e.g. notification of data quality issues, corrections and advice for improvement. Such backward propagation of useful information serves as an extra incentive for the data holders. To this end, structured and systematic feedback mechanisms should be provided for along with the feed-forward mechanisms by which the data holder notifies quality-related issues to the statistical authority.
 - b. Statistical authorities should promote the ethical, responsible and philanthropic aspects of data sharing and develop instruments for publicly acknowledging the contribution by data holders and other relevant partners to the development and production of official statistics as a public good ('name and fame'). The resulting improvement in reputation and corporate social responsibility image would act as an incentive, since brand image can be a key reason for businesses to collaborate. The data holders that have enabled data reuse can be cited as contributors to statistics when they are released. This could differentiate them from the non-participating competitors and bring a positive message to their own employees. Furthermore, such increased visibility towards the public would lower the barriers and convince even more businesses to collaborate with statistical authorities, independently of any possible regulatory obligation to cooperate.
 - c. Foster the exchange of experiences for tried and tested operational reuse modalities. This involves enabling a dialogue between data holders, statistical offices and any relevant third party to highlight cases of successful data reuse. It also involves the creation of playbooks and handbooks that statistical authorities, data holders and relevant third parties can use to understand how and for what purposes data can be reused.
10. **A clear set of requirements and safeguards for the private data holders should be encoded in the legal framework.** This concerns the regulatory framework for reusing privately held data, which is the responsibility of the EU and the Member States. Statistical authorities should preferably seek to engage data holders in voluntary partnerships. Additionally, legal requirements should be set forth in legislation to ensure the sustainable and continued production of official statistics in situations where voluntary partnership based on incentives cannot be established. The regulatory framework should therefore include a provision that data holders can be required to collaborate with statistical authorities to enable the development, production and dissemination of the official statistics portfolio, subject to a clear set of limitations and conditions, and safeguards for the data holders. The legal framework should set out and implement procedures to ensure that safeguards and conditions for data reuse are properly implemented. The main safeguards concern proportionality between costs and benefits, mechanisms to deal fairly with costs incurred by data holders, protection of business interests and data confidentiality. The definition of the operational modalities for data reuse should in any event be left to mutual agreement between the statistical authority and the data holders. The legal framework must also include a dispute resolution mechanism.
11. **Develop a fair reference cost model and compensation scheme.** While statistical authorities should not pay for the data as such, in some cases the additional costs

incurred by private holders to enable data reuse for official statistics may be substantial. This is the case for example where the reuse modalities agreed on involve a significant amount of data (pre-)processing at the premises of the data holder, or where the statistical authorities decide to rely upon additional *data services* offered by the data holders. To deal with such cases, the European Commission should take the initiative to develop and implement a fair reference cost model and compensation scheme, in consultation with the business associations and other relevant stakeholders.

4.4. Engaging with researchers

- 12.** The ESS should develop a strategy regarding collaborative research networks that explore the reuse of privately held data for research purposes. Statistical authorities should establish solid collaborative research networks with external researchers (from academic institutions, private companies and local, national and international civil society and non-profit organisations and networks) and from a diverse range of scientific disciplines in order to foster multi-disciplinary approaches. This is a key success factor to strengthen the ability of the ESS to innovate methods and practices of statistical production based on the reuse of privately held data, since this would involve a whole new range of skills and knowledge beyond the traditional perimeter of official statistics. For research partnerships that require reuse of privately held data, the data holders should be involved and should contribute to co-defining the operational modalities of data reuse.
- 13. Promote agreements on the reuse of data that involve research partners.** Statistical authorities and private data holders should be incentivised, also with support from dedicated funding schemes, to extend data reuse agreements between them to include collaborative research activities involving external researchers and possibly other data providers. Participation in such multi-party research partnerships would be voluntary. Terms and modalities of data reuse for scientific purposes are left to agreement with the respective data holders and subject to their consent. They should incorporate appropriate confidentiality protection safeguards.
- 14. Foster cooperation and trust through research and innovation activities.** Cooperation and trust can be fostered by data holders and statistical authorities alike through actively seeking collaboration, based on shared principles, with researchers from the wider data ecosystem. This would enable data holders to further discover the potential of their data and provide input for agile portfolio development for official statistics. A transparent vetting process should be created for researchers, companies and institutions who show interest in particular data sources and seek research opportunities. Effective feedback loops should be set up where data holders can gain insights from the research conducted on their data. Thus, successful and effective research partnerships have the potential to create trust and get data holders used to collaborating with official institutions and other trusted intermediaries.

4.5. Improving the institutional environment

- 15. Arrange for data reuse for statistical purposes at European level for certain kinds of privately held data.** In some cases, the ESS should arrange data reuse directly at European level for reasons of efficiency or effectiveness. In such cases, Eurostat should establish a single agreement with the data holder(s) on behalf of the whole ESS and procure all the technical and organisational instruments to implement the agreed-upon data reuse modalities, carrying the associated costs. Eurostat should then be

responsible for cascading data reuse down to national statistical authorities. The terms and modalities of reuse should be transparently communicated to the data holders involved as well as to the public, and should be preliminarily approved by the European data protection authorities⁽⁴⁸⁾ whenever personal data are involved. This model should be adopted whenever it leads to a clear reduction of cost and burden for both the statistical authorities and the data holder(s). Hybrid models should also be enabled where Eurostat or a national statistical authority arranges direct access on behalf of a subset of, but not all, national statistical authorities. In any case, if reuse of data from the same private data holder by multiple statistical authorities is needed, a coordination mechanism should be applied to minimise the burden to the data holders.

- 16. The ESS should further invest in common services and shared infrastructures supporting the reuse of privately held data.** Such services and infrastructures would be provided at European level, serving multiple statistical authorities and data holders on a voluntary basis. The adoption of hybrid dual-use infrastructures for experimentation and regular production should be promoted to reduce the time from experimentation to deployment and at the same time save on costs. The common services and shared infrastructures should include:
- a. A set of pre-defined ready-to-use model agreements (e.g. standard contract templates) between statistical authorities and data holders, based on extensive consultation with business representatives (e.g. business associations) in the relevant sectors and civil society. These model agreements provide standard solutions for all contractual issues (e.g. liabilities) and help to (i) reduce the time and cost of formulating agreements; (ii) increase harmonisation and transparency; (iii) reduce compliance risks. The adoption of such model agreements (possibly with some adjustment) is voluntary.
 - b. Shared IT infrastructures in support of collaborative research and development of statistical methodologies, experimentation and testing (trusted secure sandboxes).
 - c. Shared IT infrastructures to enable processing and integration of confidential data from multiple, mutually distrusting sources with the highest possible security standards (e.g. based on privacy enhancing technologies). The technical specification and procurement of this infrastructure is conducted by the ESS in consultation with the European data protection authorities so as to guarantee full compliance with data protection rules.
 - d. Common methodological and technical standards defined directly at the European level and coordinated with global institutions (e.g. the UN) to the extent possible. These standards are (co-)developed by the ESS and external experts from the business and research domains under collaborative structures.
 - e. Shared communication tools and public channels for acknowledging the contribution of data holders as proposed in previous section.
- 17. Competent authorities should invest in strengthening the capacity of statistical authorities to reuse privately held data for statistical purposes.** Statistical infrastructures and processes need to be considerably strengthened to cope with the challenges and needs entailed by the regular reuse of privately held data. This strengthening requires widespread and substantial investment in data research, technology, applications, development, and product and process innovation. Public authorities at all levels should establish public funding instruments to offset initial investments and marginal costs incurred for the development and production of official statistics. At the same time, they

⁽⁴⁸⁾ These are the European Data Protection Supervisor and the European Data Protection Board.

should provide sufficient resources to enable the ESS to develop the capabilities required for playing a leading role in the new data ecosystem and the new skills and capacities of staff needed by statistical authorities to deal with privately held data sources. Moreover, the development and production of official statistics itself needs to be properly funded. At the EU level, investments in official statistics across the whole range of European institutions should be strong, carefully designed and well coordinated. The European Commission, and Eurostat in particular, should have the primary responsibility for coordinating, supporting and promoting public investment efforts at EU level to enable more effective and efficient reuse of privately held data for official statistics. At the national level, Member States should be encouraged to expand the operational capacity of their statistical authorities in order to play a more proactive role in the new data ecosystem, particularly in supporting and coordinating the use of (private and public) data in the public sector. National data strategies should be encouraged to consider that statistical authorities can act as data stewards for the public sector.

- 18. The European Commission should encourage the creation of data stewardship functions.** Statistical authorities and the private sector should be encouraged to create and implement the function of data steward and organisational structures dedicated to enabling the efficient reuse of privately held data. Large data holders should be asked to identify a role or function in their organisation to interface with statistical authorities for the purpose of dealing with data reuse for official statistics. If a data steward function is present in the organisation, it is natural to expect that it will serve also as the contact point for statistical authorities. A conceptual skills framework and code of practice (possibly inspired by the European Statistics Code of Practice) should be developed for the data steward function. It is recommended that a community or network of data stewards be developed to advance peer learning and expand research and knowledge sharing on leveraging private sector data for official statistics. In addition, it pays to invest in training and education, for instance by integrating courses on data stewardship within the European Statistical Training Programme⁽⁴⁹⁾ (ESTP) or include this at the national level in university courses for students in relevant subjects.

4.6. Next steps

What should be done next? Obviously, this depends on the reception of the proposed principles and recommendations of the Expert Group, not only by the Commission, at whose instigation this report has been written, but also by the other stakeholders addressed in the principles and recommendations. If they are well received, it is up to the stakeholders addressed to apply the principles and make plans to implement the recommendations. Given the high number of recommendations and the range of stakeholders addressed, an overview of what would be on the plate of which stakeholders might be useful. Such an overview is provided in Table 1.

⁽⁴⁹⁾ https://ec.europa.eu/eurostat/cros/ESTP_en

number of recommendation	recommendation subject	stakeholders					
		Eurostat	national statistical authorities	EU public authorities	national public authorities	private data holders	research community
1	participatory mechanisms for citizens	X	X				
2	educational role of statistical authorities	X	X				
3	harmonisation at EU and global level	X					
4	security and confidentiality	X	X			X	
5	reassessment of ESS quality framework	X	X				
6	development of methodological framework	X	X				
7	openness to private sector processing methods	X	X			X	
8	balanced partnership approach	X	X			X	
9	non-financial incentives	X	X			X	
10	requirements and safeguards for data reuse	X	X	X	X	X	
11	cost model and compensation	X		X		X	
12	strategy collaborative research	X	X			X	X
13	multi-party research agreements	X	X			X	X
14	building trust and collaboration in research	X	X			X	X
15	central data reuse on behalf of the ESS	X	X			X	
16	common services and shared infrastructures	X	X			X	X
17	need for investments and proper funding	X	X	X	X		
18	data stewardship functions	X	X	X	X	X	

Table 1: Main stakeholders to which the recommendations are addressed

The table contains a couple of well considered simplifications so that it can be of practical use. First of all, to show the essence, only the main stakeholders concerned are indicated in each row. For instance, the research community has some interest in the success of virtually all recommendations, but the table highlights the most pertinent recommendations for them. The stakeholders that have an initiating role in implementing recommendations are of course always marked. Second, not all stakeholders are explicitly mentioned, in particular data protection authorities, global institutions, or some of the organisations in the wider data ecosystem. No subgroups within the stakeholder groups are identified. Where the ESS is addressed, both Eurostat and the national statistical authorities are marked⁽⁵⁰⁾. Citizens are, of course, also main stakeholders, but adding them to the table would not add much practical value.

The table shows that, as was to be expected, Eurostat and the national statistical authorities are the main stakeholders for virtually every recommendation, and private data holders are the main stakeholders outside the ESS. For some tasks, in particular concerning legislation and funding, public authorities other than statistical authorities have a crucial role to play.

What should be the priorities in implementing the recommendations? In fact, most recommendations require a sustained effort by several stakeholders together. These

⁽⁵⁰⁾ To be precise, the ESS also comprises national authorities other than the designated statistical authorities that are tasked, on their behalf, with aspects of the development, production and dissemination of European statistics.

recommendations can be started – and in some cases have been started already – immediately and in parallel. For the short term, the recommendations to be prioritised are those that entail adjustments of the regulatory framework. These are primarily the following:

- Recommendation 8 on the balanced partnership approach;
- Recommendation 10 on requirements and safeguards for data reuse;
- Recommendation 11 on cost model and compensation; and
- Recommendation 15 on central data reuse on behalf of the ESS.

For the recommendations to result in a real boost in the reuse of privately held data for official statistics in a way that is fair to all stakeholders concerned and is supported by them, the Expert Group considers it essential that recommendations 8, 10 and 11 be implemented as an integrated package. Active support by all concerned is crucial for the success of this common endeavour for the public benefit.

Finally, the Expert Group encourages the ESS to adopt a common action plan to address the recommendations.

Annex I

Glossary of Terms

Civil society organisation	Non-governmental organisations, think tanks, research centres, labour unions, social movements, philanthropic foundations and related organisations.
Code of Practice	The European Statistics Code of Practice defines the standards for developing, producing and disseminating European statistics and represents the foundation for the statistical quality framework. ⁽⁵¹⁾
Data holder	A business that has the right, the obligation or the ability to make privately held data available for reuse.
Data reuse	Any arrangement that enables the extraction for statistical purposes of pre-defined information from privately held data. Reuse of data does not necessarily entail transmission of raw data to statistical authorities. In certain situations, raw data may be processed partially or entirely on the premises of (and possibly by) the data holder, according to pre-agreed computation operations and methods, only the final computation results (possibly representing the final statistics or some intermediate, aggregate data) being transmitted to statistical authorities. In other scenarios, PET technologies may be adopted.
Data stewardship	The data stewardship function comprises and combines a set of skills and practices that enables and promotes access to and reuse of data for official statistics and/or non-commercial research in a systematic, sustainable, and responsible way.
Eurostat	The statistical authority of the European Commission.
European Statistical System	The partnership between Eurostat and the national statistical authorities and other national authorities responsible in each EU Member State for the development, production and dissemination of European statistics. This partnership also includes the EFTA (European Free Trade Association) countries. ⁽⁵²⁾
FAIR principles	Guiding principles for scientific data management and stewardship (FAIR – Findable, Accessible, Interoperable, Reusable). ⁽⁵³⁾
Official statistics	Statistics produced by or under responsibility of government organisations with a legal mandate to produce them, subject to a set of principles and rules of conduct.

⁽⁵¹⁾ <https://ec.europa.eu/eurostat/web/quality/european-quality-standards/european-statistics-code-of-practice>

⁽⁵²⁾ <https://ec.europa.eu/eurostat/web/european-statistical-system>

⁽⁵³⁾ <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4792175/>

Open data	Data available free of charge, machine readable, provided via APIs, and where relevant provided as bulk download.
Privacy enhancing technologies	Technologies enabling the computation across data held by different organisations without revealing the input data, e.g. by the use of appropriate computation protocols based on encryption and other mathematical transformation (“sharing computation” as an alternative to “sharing data”).
Privately held data	Data held by a business (or civil society organisation), the data holder, that was generated primarily for some (business-related) purpose.
Statistical authorities	Government organisations at (sub)national, supra-national or international level tasked with providing official statistics.

Annex II

Members of the High-Level Expert Group

- Ms Silvina Bakardzhleva, Senior Legal Advisor at the Bulgarian Maritime Administration
- Mr Luis Cardo Jalón, Partner at Open Ideas, Spain
- Mr Florent Diverchy, Data Marketing Expert at Black Tiger, Belgium
- Mr David Dreyer Lassen, Prorector and Professor at the University of Copenhagen, Denmark
- Ms Wieteke Dupain, Head of Knowledge, R&D at Euclid Networks, the Netherlands
- Mr Paolo Garonna, Secretary General at the Italian Banking Insurance and Finance Federation, and Professor at the Luiss G. Carli University, Italy
- Mr Christoph Gerlinger, Senior Director at Bayer AG, Germany
- Mr David Gonzalez Martinez, Group Head of Big Data, Advanced Analytics and AI at Vodafone Business at Vodafone Group, Spain
- Mr Emmanuel Letouzé, Director and Co-founder of Data-Pop Alliance and Marie Curie Fellow at the Pompeu Fabra University, Spain
- Mr Mart Mägi, Chairman of the Executive Board at Omniva, Estonia
- Ms Geta Mitrea, Lecturer PhD at the National Defence University Carol I, Romania
- Ms Felicia Pelagalli, Founder and CEO of Culture srl, Italy
- Mr Dominik Rozkrut, President of the Polish Statistical Office
- Mr Geert Somers, Partner at Timelex, Belgium
- Ms Kaisa Vent, Head of Methodology Team, Positium, Estonia
- Mr Stefaan Verhulst, Co-founder and Chief of R&D at The GovLab, New York University
- Mr Erik Wetter, Assistant Professor at the Stockholm School for Economics and Co-founder and Chairman of the Flowminder Foundation
- Ms Susanna Zaccarin, Professor at the University of Trieste, Italy
- Mr Dimitris Zissis, Associate Professor at the Greek University of the Aegean and Head of R&D for MarineTraffic, Greece

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: https://europa.eu/european-union/contact_en

On the phone or by email

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),
- at the following standard number: +32 22999696 or
- by email via: https://europa.eu/european-union/contact_en

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu/european-union/index_en

EU publications

You can download or order free and priced EU publications at: <https://op.europa.eu/en/publications>. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/european-union/contact_en).

EU law and related documents

For access to legal information from the EU, including all EU law since 1952 in all the official language versions, go to EUR-Lex at: <http://eur-lex.europa.eu>

Open data from the EU

The EU Open Data Portal (<http://data.europa.eu/euodp/en>) provides access to datasets from the EU. Data can be downloaded and reused for free, for both commercial and non-commercial purposes.

Empowering society by reusing privately held data for official statistics — A European approach

FINAL REPORT PREPARED BY THE HIGH-LEVEL EXPERT GROUP ON FACILITATING THE USE OF NEW DATA SOURCES FOR OFFICIAL STATISTICS

This publication is the final report of the High-Level Expert Group on facilitating the use of new data sources for official statistics. The Expert Group, which consisted of independent experts of various backgrounds, was set up by the Commission to advise and assist Eurostat regarding Business-to-Government data sharing for official statistics in the context of the data strategy of the Commission.

For more information

<https://ec.europa.eu/eurostat/>

