



BANCA D'ITALIA  
EUROSISTEMA

# Questioni di Economia e Finanza

(Occasional Papers)

Principled data access: building public-private data partnerships  
for better official statistics

by Claudia Biancotti, Oscar Borgogno and Giovanni Veronese

July 2021

Number

629





BANCA D'ITALIA  
EUROSISTEMA

# Questioni di Economia e Finanza

(Occasional Papers)

Principled data access: building public-private data partnerships  
for better official statistics

by Claudia Biancotti, Oscar Borgogno and Giovanni Veronese

Number 629 – July 2021

*The series Occasional Papers presents studies and documents on issues pertaining to the institutional tasks of the Bank of Italy and the Eurosystem. The Occasional Papers appear alongside the Working Papers series which are specifically aimed at providing original contributions to economic research.*

*The Occasional Papers include studies conducted within the Bank of Italy, sometimes in cooperation with the Eurosystem or other institutions. The views expressed in the studies are those of the authors and do not involve the responsibility of the institutions to which they belong.*

*The series is available online at [www.bancaditalia.it](http://www.bancaditalia.it).*

ISSN 1972-6627 (print)

ISSN 1972-6643 (online)

*Printed by the Printing and Publishing Division of the Bank of Italy*

# PRINCIPLED DATA ACCESS: BUILDING PUBLIC-PRIVATE DATA PARTNERSHIPS FOR BETTER OFFICIAL STATISTICS

by Claudia Biancotti\*, Oscar Borgogno\* and Giovanni Veronese\*

## Abstract

Official statistics serve as an important compass for policymakers due to their quality, impartiality, and transparency. In the current post-pandemic environment of great uncertainty and widespread disinformation, they need to serve this purpose more than ever. The wealth of data produced by the digital society (e.g. from user activity on online platforms or from Internet-of-Things devices) could help official statisticians improve the salience, timeliness and depth of their output. This data, however, tends to be locked away within the private sector. We argue that this should change and we propose a set of principles under which the public and the private sector can form partnerships to leverage the potential of new-generation data in the public interest. The principles, compatible with a variety of legal frameworks, aim at establishing trust between data collectors, data subjects, and statistical authorities, while also ensuring the technical usability of the data and the sustainability of partnerships over time. They are driven by a logic of incentive compatibility and burden sharing.

**JEL Classification:** C82, F53, L38.

**Keywords:** data governance, business-to-government data access, official statistics.

**DOI:** 10.32057/0.QEF.2021.0629

## Contents

|  |    |
|--|----|
| 1. Introduction: official statistics in a fast-changing world.....                         | 5  |
| 2. Data sources created by the digital society as a new input for official statistics..... | 6  |
| 3. The challenges and enablers of access to new-generation data sets.....                  | 8  |
| 4. Principles for public-private data partnerships (DPs) .....                             | 11 |
| 4.1 General principles .....   | 11 |
| 4.2 Principles for official statistical agencies.....                                      | 13 |
| 4.3 Principles for private-sector data collectors .....                                    | 13 |
| 5. Conclusions .....   | 14 |
| Annex 1 - New-generation data for the environment.....                                     | 15 |
| Annex 2 - Barcode scanner technology for the calculation of price indexes in Europe.....   | 15 |
| References .....   | 16 |

---

\* Bank of Italy, Economic Research and International Relations.



## 1. Introduction: official statistics in a fast-changing world\*

In 2014, the General Assembly of the United Nations endorsed the Fundamental Principles of Official Statistics<sup>1</sup>. The very first principle states that “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”.

In order to serve their purpose, statistics must be available at the right time (timeliness), must represent correctly the phenomenon of interest for the users (relevance) with the minimum possible level of uncertainty (accuracy), and must be free from any kind of political or economic bias (professional independence, impartiality).<sup>2</sup> These features, together with production based on scientific methods and transparent processes, make official statistics trustworthy.

Digitalization and globalization have accelerated social and economic dynamics. The Covid-19 pandemic required the prompt deployment of emergency measures, while introducing marked uncertainty for governments, businesses, and individuals. This does not change the mission of statistical authorities, yet it compels them to adapt their toolbox to evolving user needs and expectations. Some indicators need to be published with higher frequency in order to stay timely. New activities, behaviours and phenomena have to be precisely defined and measured if statistics are to keep their relevance. In the age of social media, contested truths and fake news, accuracy, professional independence and impartiality are essential like never before.

Statistical authorities can best face this challenge by leveraging one specific trait of the digital society – the abundance of data on a very large and diverse set of domains of human life, including some that have never been quantified before. The amount of raw information that is produced daily by all sorts of devices, from smartphones to industrial machinery, is unprecedented and increasing exponentially. This data, together with new analytical techniques, can greatly advance the cause of evidence-based policymaking. Official statistical agencies, which represent social rather than private interests, need to play a primary role in this process. Depending on circumstances, they need to be given access to certain privately collected information that is relevant to specific policy objectives or, at the very least, a right to use it at various degrees of aggregation. When policymakers choose to resort to indicators sourced by the private sector, official statisticians must be able to audit their methodological quality.

In this paper, we aim to provide a foundation for broader, structured access to new data sources on the part of official statisticians, for the purpose of evidence-based policymaking and research. Whenever the data refers to individuals, such access has to be accompanied by a careful assessment of privacy and confidentiality implications for individuals, and the implementation of safeguards to guarantee full anonymity and preserve trust. These considerations also extend to other entities such as businesses that have sensitive and economically valuable data under their custody about their operations and their relations, including but not limited to intellectual property. We focus on

---

\* The paper benefited from extensive discussions and suggestions from colleagues in many G20 and non G20 central banks, finance ministries and statistical offices. In particular, we thank Ulf von Kalckreuth and Robert Kirchner for their contribution at the drafting of a first version of this work. We also thank for their comments Craig Hammer, as well as Gabriel Quiros-Romero. The usual disclaimer applies. Any opinions expressed in this paper are personal and are not to be attributed to the Bank of Italy.

<sup>1</sup> As originally adopted by the U.N. Statistical Commission in 1994.

<sup>2</sup> In Europe, this set of characteristics constitutes the definition of “quality” of statistical information. See Regulation 223/2009 on European Statistics and European Statistics Code of Practice (available at <https://ec.europa.eu/eurostat/web/quality/european-statistics-code-of-practice>) as well as the Public commitment on European Statistics by the ESCB.



information collected and processed by private companies, which represents the majority of “big data”<sup>3</sup>.

The proposals put forward in this paper do not aim to undercut existing initiatives for public-private cooperation in the data sphere, especially those stewarded by international organizations for statistical capacity building in low-income countries. Rather, our goal is to enhance multilateral consensus and coordination in this field, and to provide renewed impetus to the multiple discussions that are occurring both at the national and the international level among statisticians as well as policy-makers. In this respect, in their April 2021 Communiqué, G20 Finance Ministers and Central Bank Governors made a clear call to expand access to new data sources: “*We also recognize that improving data availability and provision, including on environmental issues, and harnessing the wealth of data produced by digitalization, while ensuring compliance with legal frameworks on data protection and privacy, will be critical to better inform our decisions.*”

We acknowledge that jurisdictions differ in how they understand and regulate the data sphere and we are fully aware of how closely data governance choices intertwine with national security. This paper contributes to the on-going discussion on data governance by setting out a way forward to design a workable framework for Business-to-Government data access that could be followed across G20 countries and beyond. For this reason, we stop short of any detailed prescription and rather outline a set of very broad, high-level principles that we see as compatible with a wide variety of existing data governance models.<sup>4</sup> Notably, we are aware that interventions in the direction we propose are likely to require at least in some jurisdictions a revision of the legal basis for the activity of official statistical agencies, although this discussion is out of the scope of the present article.<sup>5</sup> Finally, as the focus of the work is on the principles that should underpin Business-to-Government data access within any jurisdiction, we do not touch upon the issue of cross-border data flows.

The paper is organized as follows. Section 2 describes how data sources created by the digital society can provide useful input for official statistics. Section 3 discusses the main challenges and enablers of access to these sources on the part of statistical authorities. Section 4 proposes three sets of principles that should guide data access.

## **2. Data sources created by the digital society as a new input for official statistics**

Before the widespread digitalization of economies and societies, official statistical agencies enjoyed an advantage over the private sector in terms of access to data on most domains. Information generally had to be collected on purpose, with costs and challenges that private companies had no incentive to

---

<sup>3</sup> Improving access to administrative records for statistical purposes is also an important goal. We do not address this issue here, since it involves specific reference to domestic administrative structure of each jurisdiction. For an overview of the problem from a US perspective, see Lane (2020).

<sup>4</sup> A precursor to some of our proposals was set forth in 2017 by the Global Working Group on Big Data for Official Statistics, within the UN Statistical Commission, in a set of “Recommendations for Access to proprietary data”, to support international collaboration in the development of official statistics using new data sources and to help countries measure the Sustainable Development Goals (SDGs).

<sup>5</sup> Here and thereafter, “domestic” refers to the appropriate level of decision-making on data governance, not always corresponding with the nation-state. For example, in the European Union (EU) many decisions in the field are the prerogative of supranational authorities. As to the strategy of the EU on business-to-government data access and the experience of France (where the law for a digital republic allows the public sector to access certain private sector data of general interest), see: High-Level Expert Group on Business-to-Government Data Sharing (2020). For an economic assessment of the matter at the EU level, see: Bertin and Duch-Brown (2020).



sustain on a large scale<sup>6</sup>. This advantage started to fade as more and more data were created automatically together with the provision of digital services, and flowed to the businesses providing such services. The advantage further receded with the rise of giant, multi-purpose platforms and the advent of the Internet of Things. Today the private sector has a lead when it comes to volume, scope and granularity of micro-level data on many aspects of economic and social life. It also attracts top talent in artificial intelligence. The big data industry is now routinely extracting new correlations among multiple data pools.

Access to private data can improve official statistics in all dimensions of quality described in Section 1. This is best exemplified by looking at the most challenging policy decision that governments had to face during the Covid-19 crisis – how tightly to restrict personal mobility and economic activity in order to slow the spread of the virus. The choice required simultaneous consideration of a very complex set of factors, ranging from the epidemiological to the economic and social ones.

Data was needed on each piece of the puzzle. In most countries, micro-level information on contagion, hospitalizations and deaths was accessible to authorities in a relatively prompt fashion, although quality was not homogeneous across different areas. This was not the case for other domains:

- Indicators on compliance with social distancing obligations. Governments could only produce very rough indicators, such as the number of individuals caught in violation of stay-at-home orders. Private businesses such as Google, Apple, and Facebook could publish detailed “mobility reports” at the subnational level for most jurisdictions, based on data sourced by billions of users. These did not come with sufficiently transparent methodological documentation or any possibility of independent audit;
- Indicators on consumption habits. The pandemic had a sharp and sudden disruptive effect on consumption habits, as entire markets were shut down and others moved online. It also slowed down some official surveys on consumer spending. Major online retailers such as Amazon or Alibaba had important information on what households were buying and how, if at all, prices were changing in response to shifts in demand. They knew in real time which supply chains were working well and which were suffering.

A similar situation emerged for other key magnitudes, such as the number of people looking for work or social sentiment vis-à-vis policy decisions<sup>7</sup>. Relevant micro-level data was produced in abundance, but it was generally not available to statistical authorities – although in some countries successful experiments were carried out. In South Africa, the central bank reviewed debit and credit card data from a payments clearinghouse, tracking consumption trends on a daily basis. In other countries, new digital administrative records became available for statistical use.

Public-private competition in the space of statistical indicators is an inevitable and healthy feature of the digital society. For decades now specialized companies have been providing information and analytics on various business domains, with market research and financial analytics offering perhaps the most obvious examples. In the age of giant platforms with their rich multidimensional datasets, and especially in times of crisis where speed is of the essence, this can however turn into a forced

---

<sup>6</sup> Data collectors faced challenges such as designing cost-effective samples, reaching potential respondents to surveys, persuading them to cooperate, and validating information based on incomplete recollection or records. This required significant financial and human resources, along with a reputation for reliability and legal means to enforce any obligation to report. Only official statisticians could benefit from all such resources in a timely fashion.

<sup>7</sup> Also see Biancotti et al. (2020).

reliance on the private sector, as official statisticians face challenges to get access to information and assess quality of indicators sourced by corporates<sup>8</sup>.

Some successful examples also exist. Eurostat, the European statistical agency, has an agreement with AirBnB, Booking.com, Expedia and TripAdvisor to receive aggregate data on tourism<sup>9</sup>. This allows for a finely grained assessment of how the tourism sector, hit very hard by the pandemic, is faring. Some European phone network operators shared 2020 mobility data with the European Commission<sup>10</sup>. Similar data was also analyzed by researchers from the International Monetary Fund<sup>11</sup>. Beyond the Covid-19 emergency many companies, including the largest platforms, participate in “data for good” initiatives whereby they provide information to governments, academia<sup>12</sup>, and international or regional organizations<sup>13</sup> in the interest of social and economic development. See Annex 1 for another non-Covid example of how private data can benefit public policy, directly feeding into international policy top priorities on climate and sustainability.

However, this is not enough. These initiatives lack a systematic dimension. With scant governance frameworks to enable the scaling up and persistence in their use, the engagement of official statistics producers with the private sector and vice versa is still underdeveloped and fragmented<sup>14</sup>. This has the potential to create uncertainty and additional costs on both sides, thus undermining incentives for partnership formation. Statistical offices and the private sector thus need to consider engaging in sustainable partnerships that unlock coordinated flows of high-quality and high-frequency data to inform evidence-based policy-making and the broader debate<sup>15</sup>. International organizations have a role to play in stewarding such partnerships<sup>16</sup>.

### **3. The challenges and enablers of access to new-generation data sets**

Enabling access to and/or use of new-generation data sets for official statistics purposes on the part of statistical authorities presents challenges in three broad domains.

#### **Trust.**

- Between data subjects and statistical authorities. In many jurisdictions, citizens are increasingly attentive to why and how their data is collected, processed, and shared. Periodically, data abuse scandals erupt involving private companies, governments, or both. This is a truly global issue, independent of geographical differences in data protection

---

<sup>8</sup> In some cases, authorities combined indicators provided by digital platforms with traditionally sourced data to create new statistics. One example is provided by the new weekly activity index (WAI) for Germany, developed in 2020 by the Bundesbank as part of the response to Covid-19. See Deutsche Bundesbank (2020).

<sup>9</sup> See European Commission (2020), Commission Reaches Agreement with Collaborative Economy Platforms to Publish Key Data on Tourism Accommodation.

<sup>10</sup> See Iacus, et al. (2020), How Human Mobility Explains the Initial Spread of Covid-19, JRC Methodological Report.

<sup>11</sup> Caselli et al. (2021), The Disproportionate Impact of Lockdowns on Women and the Young, Voxeu.org.

<sup>12</sup> See, for example, the Opportunity insights economic tracker at Harvard University (available at <https://tracktherecovery.org/>). Anonymized data provided by a diverse set of private actors, e.g. credit card processors and payroll management firms, are used to build real-time economic indicators across different geographical areas, industries, or income groups.

<sup>13</sup> See for example the Development Data Partnership, which brings together the World Bank, the International Monetary Fund, the Organization for Economic Cooperation and Development, and most major digital platforms (available at <https://datapartnership.org/>).

<sup>14</sup> Also see OECD (2019), Enhancing Access to and Sharing of Data. Reconciling Risks and Benefits for Data Re-Use Across Societies.

<sup>15</sup> While micro-level data may be needed for some purposes, aggregate information might be sufficient in other cases.

<sup>16</sup> See for example Hammer et al. (2017), Big Data: Potential, Challenges and Statistical Implications, IMF Staff Discussion Notes.

frameworks. Any broadening in data access must happen in such a way that citizens' trust is preserved. This requires first of all that citizens are properly informed their data can be shared with official agencies. Individuals should be able to trust that official statistical agencies apply sufficient safeguards to guarantee anonymity and relevant privacy requirements. Moreover, official statistical agencies might have access to other data sources and a combination of these data sources could pose additional privacy risks. This means additional precautions such as information barriers or application of privacy enhancing technologies. Best practices in cybersecurity are especially important. Concentration of data at central data points should be prevented.

- Between statistical authorities and the private sector. Policymakers have the questions, and private data may be able to provide the answer, but only the companies that hold it have full knowledge of their potential for a given purpose. Official statisticians need to trust that all information relevant to building or integrating a given statistical indicator is disclosed. Symmetrically, private companies need to trust that granting statistical authorities data access is not going harm their rights and legitimate interests, including economic ones. They need to trust that their reputation, competitiveness or economic performance is not going to be damaged by authorities misusing information in any way (see also Annex 2 for an example).

**Usability.** Privately collected data is seldom an immediate fit for use on the part of official statisticians:

It may refer to identified or identifiable individuals – typically, customers. In this case, it needs to be thoroughly anonymized prior to sharing, and/or analysed with privacy-preserving techniques. For complex, multidimensional personal information this may be a difficult task, although new solutions are constantly being developed<sup>17</sup>. Delving in the very complex issue of re-identification possibilities offered by AI, privacy-preserving analysis, and various definitions of anonymity is beyond the scope of this work. We acknowledge the need to refer to impartial standard setting bodies and/or appropriate research task forces to establish ground rules on this subject and continuously monitor their adequacy in the light of technical development.

- It may cover only a portion of the population that is of interest to authorities, or be biased for a host of other methodological reasons, which would make it unfit for direct and immediate policy use. For example, data on consumption collected by online retailers or providers of digital payments over-represents younger, digitally literate, and affluent households, while under-representing the elderly and disadvantaged communities<sup>18</sup>. If statisticians aim at leveraging this type of data to produce population-wide consumption measures, they will need to either integrate this type of data with conventional sources or apply model-based corrections. This is of the highest importance, since indicators that capture only the behaviour of certain groups can inadvertently lead to discriminatory and distorted policy decisions;

---

<sup>17</sup> The application of anonymization techniques to data and the choice of the relevant level of aggregation (if any) ultimately depend on the type of data at hand and on the geographic reference as well. For example, firm-level data can be anonymized by removing geographical information (and/or outliers) or re-aggregated at higher level (e.g. sector level) to share them while preserving anonymity, but the appropriate level of sector aggregation will be different depending on whether the data are regional or national and on the size of the geographic area. On the risk of possible re-identification of individuals in highly granular and multidimensional datasets, see for example Torra and Navarro-Arribas (2016).

<sup>18</sup> According to the European ICT surveys, one in six people do not use the Internet and more than one in three users lack basic digital skills. See INSEE – 2019 (available at <https://www.insee.fr/en/statistiques/4280364>).

- Companies optimize data formatting, labelling and storing for their own internal purposes, sometimes through the use of proprietary systems. Reorganizing and reprocessing may be needed before granting access to official statisticians.

**Sustainability.** The potential of new-generation data can be fully leveraged by statistical authorities only through long-term, systematic access and use. This requires sustainable modes of cooperation, where the responsibilities of all parties involved are clear and consistent over time, the economic and operational costs are assessed beforehand and transparently allocated between the parties involved, and the incentives are appropriately aligned in the context of broader domestic data governance frameworks. It also requires a credible, long-standing commitment from both sides. Adequate provisions need to be in place especially for high-value data that may need to be preserved for a long time.

The challenges of trust, usability, and sustainability should be met through the development of private-public data partnerships guided by globally shared principles. The word “partnership” here is meant to be legally neutral, encompassing frameworks of cooperation that may be widely different in how they address the economics of data access and use. The concept of “Partnership” highlights that a joint effort is needed to pursue the common good of improved statistical information in the age of big data. Legal frameworks are an important part of the equation, but so is pursuit of a socially cooperative spirit – embraced by data subjects, data collectors, statisticians, and policymakers – and of a common ethical aspiration towards information use for the public good.

Choices on which arrangements are legally acceptable and/or desirable should fall within the remit of individual governments<sup>19</sup>, without compromising on the independence of official statistics and compliance with international standards. There is a continuum of options, ranging from scenarios where partnerships are a purely voluntary engagement between authorities and companies (e.g. contracts for the provision of data access and accessory services) to an approach whereby the private sector is mandated by law to share information and expertise at no cost. Individuals may also be directly involved, e.g. through opt-in schemes whereby they request private-sector collectors to share a portion of their data with statistical authorities for certain public interest purposes. *Ex ante* regulation could also be an option in any of these fields or segment thereof. Countries may choose to adopt one or more principles as non-binding best practices, transform any or all of them into positive obligations with attendant degrees of legal liability, or any intermediate solution.

We expect that most jurisdictions are going to opt for something in between, perhaps differentiating the approach depending on the understanding of the role of government, the type of statistical need, and the type of data. Any scheme chosen needs to incorporate an assessment of incentive compatibility, so that quality objectives are met. When multinational data collectors are involved, coordination between national authorities is desirable for international comparability purposes and transnational use cases (e.g. tourism statistics, balance of payments, price comparisons across

---

<sup>19</sup> See Note 6.

countries<sup>20,21</sup>). International, supranational and regional institutions could help with this coordination role in their respective sphere of competence<sup>22</sup>.

Another key enabler of wider access to information that should be discussed at the country level is a clear framework on data governance, starting from the definition of usage rights over data (i.e. what companies, governments, and individuals can do with a given piece of data and under what conditions). This would greatly help in the implementation of data access, by anchoring it to formal concepts of economic and social optimality. In addition to data collectors and authorities, the data subjects or their representatives should ideally be involved in defining a well-balanced framework.

A gradual phase-in of the principles through a number of carefully designed use cases, ideally to take place under different legal frameworks and to be monitored by an international organization, such as the IMF or the OECD, might be very helpful in refining the proposal.

#### 4. Principles for public-private data partnerships (DPs)

Proposed principles should be understood as best practices and can be divided in three broad groups. Group 1 sets the stage by stating general principles for DPs, applicable to all parties involved. Group 2 introduces principles for official statistical agencies (OSAs). Group 3 outlines principles for private-sector participants. Principles are inspired by a logic of burden sharing in order to avoid that either the public sector or the private one are faced with excessive and unreasonable tasks<sup>23</sup>.

##### 4.1. General principles

- **Purpose and time limitation**<sup>24</sup>: DPs should be directed at clearly specified purpose(s) that are demonstrably in the public interest, e.g. production of official statistics on a precisely defined set of phenomena, over a specific period, or training of a specific algorithm for a given purpose. They may also include a limitation of duration for the use of the data.
- **Proportionality**: DPs should be limited to data that is essential for the stated purpose(s). In any DP, the cost and effort required for the supply and use of data should be reasonable and proportionate to the benefits pursued and to the interest of the stakeholders involved.
- **Controlled flexibility**: exploratory analysis of information collected by private participants in a DP might be desirable to determine which data items are best suited to the stated purpose(s), or to determine new purposes based on emerging needs. Such analysis should be

---

<sup>20</sup> The International Comparison Program, managed by the World Bank under the auspices of the UN Statistical Commission, is one especially fitting example (available at <https://www.worldbank.org/en/programs/icp>). One of the largest statistical initiatives in the world, it aims at reconstructing purchasing power parities across the globe, and it would benefit from an international sample of prices from e-commerce platforms and other internet services.

<sup>21</sup> For an example of a major international initiative, see the United Nations Global Platform – Data for the World (available at <https://unstats.un.org/bigdata/un-global-platform.cshtml>).

<sup>22</sup> For an important reference on many of the issues discussed throughout this Section see OECD (2021), Recommendation of the Council concerning Access to Research Data from Public Funding.

<sup>23</sup> While this article focuses on public-private data partnerships, application of the relevant subset of principles to the increasing number of private-private data partnerships as well would ultimately contribute to making public-private partnerships more sustainable.

<sup>24</sup> In this article, we adopt the EU-GDPR denomination for this principle. Other denominations, such as “necessity” as used in Canada, convey an equivalent meaning.

carried out on synthetic datasets that disclose which items are potentially available without revealing the actual informational content.

- **Anonymization of individual data and/or use of privacy-preserving analytical techniques:** any privately sourced data originally referring to an identified or identifiable natural person or legal entity can be used in a DP only if it has been successfully anonymized by the data collector prior to sharing and/or if it is analyzed with appropriate privacy-preserving techniques. Special care should be taken in case of any data pooling<sup>25</sup>. Given the technical complexity of this problem, techniques for anonymization and privacy-preserving analysis should be chosen among those indicated by independent standard setting bodies and/or relevant research task forces.
- **Application of professional standards:** all participants in a DP should act in full accordance with scientific and ethical standards of the profession, as codified by relevant international and national statutes and/or industry bodies.
- **Transparency:** all participants in a DP should be transparent vis-à-vis each other with respect to process and results. This includes but is not limited to providing clear descriptions of data, metadata, methodology, risks of discontinuity in time series, algorithms, and code deployed by each of the parties, and the impact of any related choice on the final product.
- **Fair and ethical data use:** all participants in a DP should access and use data in an ethical, legitimate, fair and non-discriminatory manner, with full respect for the choices made by individuals on how their data can be used<sup>26</sup>. Data coming from different sources should be processed in such a way as to limit biases, including “selection bias”, to the extent possible.
- **Risk mitigation and safeguards:** DPs should contain appropriate safeguards in order to protect the rights, such as privacy, data security and non-discrimination, of the individuals whose data is used. DPs have to guarantee the legitimate interest of all stakeholders – “legitimate interest” also meaning that the way data is provided should not aim at protecting the interests of a specific group<sup>27</sup>. The private data collector should bear no negative economic consequences from participation in the DP and should be able to continue to use and monetize the data in its own business operations.
- **Timeliness:** all participants in a DP shall make sure that data are shared in a timely manner to make sure that data are relevant for the purpose.
- **Accountability:** all participants in a DP should be accountable for using and sharing data in a responsible and lawful way and should be able to demonstrate compliance.

---

<sup>25</sup> See for example the experience of Korea, which only allows linkage between individual data under strict criteria and only in dedicated centers operated by the OSA.

<sup>26</sup> At present, there is no internationally shared reference framework for ethical data use, yet many different proposals are being put forward especially with respect to AI-based techniques. The workstreams on data access/use and on ethics and fairness need to be tightly linked going forward.

<sup>27</sup> Also see the Five Safes approach adopted in the United Kingdom (available at <https://blog.ons.gov.uk/2017/01/27/the-five-safes-data-privacy-at-ons/>).

## 4.2. Principles for official statistical agencies (OSAs)

- **Overall responsibility and accountability on process and output:** OSAs should be ultimately responsible and accountable for the whole production process and statistical results deriving from the data access partnership, regardless of any processing steps taking place at the private data source, i.e. computation being “pushed out”. This implies full transparency of the methods, procedures and technical systems (e.g. Application Programming Interfaces) used, including sustainability of data flows, the possibility to monitor quality aspects, ensuring confidentiality, and avoiding any harm to private-sector parties.
- **Overall responsibility on methodological standard setting:** OSAs should have the overall responsibility on methodological standard setting, and make sure that procedures are in place to guarantee that standard concepts, definitions, and classifications are consistently applied over domains, time and, as far as possible, across countries. Where this is not realistic, any concept, definition, and classification should be a good approximation of those required for statistical purposes.
- **Safeguard of individual interests:** use of privately held data by OSAs should not harm the rights of individuals. Data collection and dissemination by OSA should have sufficient safeguards in place to guarantee anonymity and fair treatment of individuals.
- **Safeguard of business interests:** accessing and using privately held data for official statistics should not compromise the reputation and business of the private data holder. OSAs should ensure neutrality and impartiality and avoid harm to business interests, especially if private data has a market value or it comprises sensitive information. Data shall be processed so to ensure adequate protection of know-how and trade secrets of business involved. Broad-ranging “indemnity”<sup>28</sup> clauses may be considered to safeguard the position of private participants in DPs.
- **“Collect only once”:** OSAs should coordinate domestically and, as far as possible under national law, internationally in order to standardize DPs with the aim of achieving similar conditions and to avoid requesting the same data multiple times from the same private provider. Where applicable, this can be facilitated by the setup of dedicated infrastructure as part of the DP. Such coordination and standardization may be efficient for data holders and beneficial to OSAs.

## 4.3. Principles for private-sector data collectors.

- **Data and metadata availability:** private participants in DPs should cooperate with OSAs to identify, list and make available all the data and metadata that are necessary to the stated purpose(s) of the DP.
- **Technical accessibility:** private participants in DPs should provide OSAs with any information necessary to the security and continuity of IT solutions for data and metadata access.

---

<sup>28</sup> By means of indemnity clauses, OSAs could assume responsibility to compensate and reimburse private data collectors for certain losses such as monetary costs and expenses resulting from unauthorized data breaches or any other specific violations on the part of the OSAs.



- **Cooperation on formats and standards:** private participants in DPs should cooperate with OSAs in deploying formats and standards that are adequate for sustainable interchange and interoperability. The use of global benchmarks such as XBRL or SDMX should, where possible, be privileged in order to ensure the interoperability and use of common metadata. Private participants may also offer to cooperate in defining new standards that reflect frontier technologies in storage and analytics.
- **Mitigation of limitations:** to address the potential limitations of private-sector data, including potential inherent bias, private sector actors should offer reasonable and proportionate support to help maintain and assess its quality for the stated purposes wherever appropriate.
- **International comparability:** private participants in multiple DPs involving OSAs in different jurisdictions might usefully contribute to establishing compatibility of methods, technical systems, and standards as far as allowed by the specific purposes of the OSAs and by national legislation, for the purpose of international comparability and without prejudice to the independent management of statistical infrastructure and process on the part of OSAs. For the sake of international comparisons, the sharing of new datasets could be embedded in international data sharing arrangements.<sup>29</sup>

## 5. Conclusions

Official statistics are a key compass for policymakers, especially in the current scenario of high uncertainty and proliferating disinformation. More than ever, they need to be readily accessible, impartial, reliable, timely and accurate. They must also be flexible enough to capture emerging phenomena, and to remain salient even in the presence of major disruptions. The Covid-19 crisis revealed both weaknesses and gaps in our statistical systems.

Digitalization provides a challenge, but also a unique opportunity. The abundant data it generates could feed into official statistics, improving them in many respects. Today, this happens to a limited extent. Official statisticians have little access to new data sources, which remain largely confined with the private sector, fragmented across collectors and bound by commercial and regulatory ties. It is hard to underestimate the welfare-enhancing value of this data. Were privately held data available for official statistics, they could vastly improve public decision-making.

In this paper, we proposed the development of private-public data partnerships for better official statistics, driven by globally shared principles. Those aim at establishing trust between data collectors, data subjects, and statistical authorities, while also ensuring technical usability of the data, and sustainability of partnerships over time. They emphasize burden sharing and incentive compatibility.

The word “partnership” here is meant to be legally neutral, encompassing frameworks of cooperation that may be widely different in how they address the economics of data access and use. It should accommodate G20 jurisdictions and beyond. Choices on which arrangements are legally acceptable and/or desirable should fall within the remit of individual governments, without compromising on the independence of official statistics and compliance with international standards.

---

<sup>29</sup> This is in line with recommendation II.20 of the second phase of the G20 Data Gaps Initiative (see <https://www.imf.org/en/Publications/SPROLLS/G20-Data-Gaps-Initiative#sort=%40imfdate%20descending>).

## **Annex 1 - New-generation data for the environment**

Policymakers in many jurisdictions aim for a data-driven approach to environmental sustainability, but they are slowed down by limited availability of relevant data<sup>30</sup>. Examples of information that is typically held by the private sector and should be shared with authorities include, but are not limited to:

- data flows collected by electricity and gas companies from “smart meters” installed by residential and business consumers. Those provide a very detailed, real-time overview of energy consumption;
- data on take-up of insurance against natural disasters and related claims. Those are useful in order to assess the magnitude of environmental risk and how it is shared across the financial and non-financial sectors;
- data on consumer preferences for green and sustainable products and services, e.g. as observed by online retailers and advertisers;
- data on sentiment relating to environmental issues, e.g. as measured through textual analysis of social media posts.

## **Annex 2 - Barcode scanner technology for the calculation of price indexes in Europe**

Barcode scanner technology enables retailers to capture detailed information on transactions at the point of sale, including but not limited to quantities sold and prices paid for each product. Scanners are privately owned and they were introduced for commercial reasons, such as stock management. The data they provide are, however, also used by official statisticians for the compilation of consumer price indexes.

In Europe, the importance of scanner data is embedded in legislation: “Upon the request of the national bodies responsible for compiling the harmonised [price] indices, the statistical units shall provide, where available, electronic records of transactions, such as scanner data”<sup>31</sup>. Currently, 10 EU Member States and all 3 EFTA countries use scanner data to some extent.

Nevertheless, in many countries the negotiations with retailers to obtain the scanner data take several years and sometimes fail. A relationship of trust with a retailer has to be established and nurtured, as the scanner data contains highly sensitive business information. Retailers require authorities to adopt the strictest security measures to transmit, store and process the data. In return, authorities need guarantees from the retailers for the continuous, timely and complete transmission of the data.

---

<sup>30</sup> The “Risk Data Hub” of the European Commission, a non-financial database including both macro-level indicators of environmental risk and micro-level information such as environmental-social-governance (ESG) ratings for businesses, provides a first step in the right direction (see <https://drmkc.jrc.ec.europa.eu/risk-data-hub#/>).

<sup>31</sup> Regulation (EU) 2016/792 of the European Parliament and of the Council of 11 May 2016 on harmonised indices of consumer prices and the house price index, Article 5(4). “Statistical units” refers to business that are included in the sample selected by authorities for measuring price dynamics.

## References

- Biancotti, C., Kirchner, R., Mouriaux, F., Rosolia, A., and Veronese, G. (2021) “Covid-19 and Official Statistics: A Wake-up Call?”, *Occasional Papers of Bank of Italy Series* n. 605, [https://www.bancaditalia.it/pubblicazioni/qef/2021-0605/QEF\\_605\\_21.pdf?language\\_id=1](https://www.bancaditalia.it/pubblicazioni/qef/2021-0605/QEF_605_21.pdf?language_id=1).
- Bertin, M., Duch-Brown, N. (2020) “The economics of business-to-government data sharing”, *JRC Digital Economy Working Paper* 2020-04, <https://ec.europa.eu/jrc/sites/jrcsh/files/jrc119947.pdf>.
- Caselli, F., Grigoli, F., Rente Lourenço, P., Sandri, D. and Spilimbergo, A. (2021), “The Disproportionate Impact of Lockdowns on Women and the Young”, Voxeu.org, <https://voxeu.org/article/disproportionate-impact-lockdowns-women-and-young>.
- Deutsche Bundesbank (2020) “A weekly activity index for the German economy”, Monthly Report, May 2020.
- ESCB (European System of Central Banks), Public commitment on European Statistics, [https://www.ecb.europa.eu/stats/ecb\\_statistics/governance\\_and\\_quality\\_framework/html/escb\\_public\\_commitment\\_on\\_european\\_statistics.en.html#:~:text=Public%20commitment%20on%20European%20Statistics%20by%20the%20ESCB,statistics%20in%20the%20areas%20under%20the%20ESCB%E2%80%99s%20responsibility](https://www.ecb.europa.eu/stats/ecb_statistics/governance_and_quality_framework/html/escb_public_commitment_on_european_statistics.en.html#:~:text=Public%20commitment%20on%20European%20Statistics%20by%20the%20ESCB,statistics%20in%20the%20areas%20under%20the%20ESCB%E2%80%99s%20responsibility).
- European Commission (2020), Commission Reaches Agreement with Collaborative Economy Platforms to Publish Key Data on Tourism Accommodation, [https://ec.europa.eu/commission/presscorner/detail/en/ip\\_20\\_194](https://ec.europa.eu/commission/presscorner/detail/en/ip_20_194).
- General Assembly of the United Nations (2014) Fundamental Principles of National Official Statistics, A/RES/68/261 from 29 January 2014, <https://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx>.
- Global Working Group on Big Data for Official Statistics (2017) Recommendations for Access to proprietary data, [\">https://unstats.un.org/unsd/statcom/48th-session/documents/BG-3d-recommendations-for-access-to-proprietary-data-E.pdf.\](https://unstats.un.org/unsd/statcom/48th-session/documents/BG-3d-recommendations-for-access-to-proprietary-data-E.pdf)
- Hammer, C., Kostroć, D., Quiros-Romero, G. (2017), Big Data: Potential, Challenges and Statistical Implications, *IMF Staff Discussion Notes*, <https://www.imf.org/en/Publications/Staff-Discussion-Notes/Issues/2017/09/13/Big-Data-Potential-Challenges-and-Statistical-Implications-45106>.
- High-Level Expert Group on Business-to-Government Data Sharing (2020) Towards a European strategy on business-to-government data sharing for the public interest, <https://digital-strategy.ec.europa.eu/en/news/experts-say-privately-held-data-available-european-union-should-be-used-better-and-more>.
- Iacus, S., Santamaria Serna, C., Sermi, F., Spyrtatos, S., Tarchi, D. and Vespe, M. (2020), “How Human Mobility Explains the Initial Spread of Covid-19”, *JRC Methodological Report*.
- Lane, J. (2020), “Democratizing Our Data: A Manifesto”, MIT Press. <https://mitpress.mit.edu/books/democratizing-our-data>.
- OECD (2021), Recommendation of the Council concerning Access to Research Data from Public Funding, <https://www.oecd.org/sti/recommendation-access-to-research-data-from-public-funding.htm#:~:text=On%2020%20January%202021%2C%20the,shown%20in%20the%20figure%20below>.

OECD (2019), Enhancing Access to and Sharing of Data. Reconciling Risks and Benefits for Data Re-Use Across Societies, <https://www.oecd.org/publications/enhancing-access-to-and-sharing-of-data-276aaca8-en.htm>.

Torra, V. and Navarro-Arribas, G. (2016) “Big Data Privacy and Anonymization” in Privacy and Identity Management: Facing up to Next Steps, ed. A. Lehmann., <https://hal.inria.fr/hal-01629153/document>.