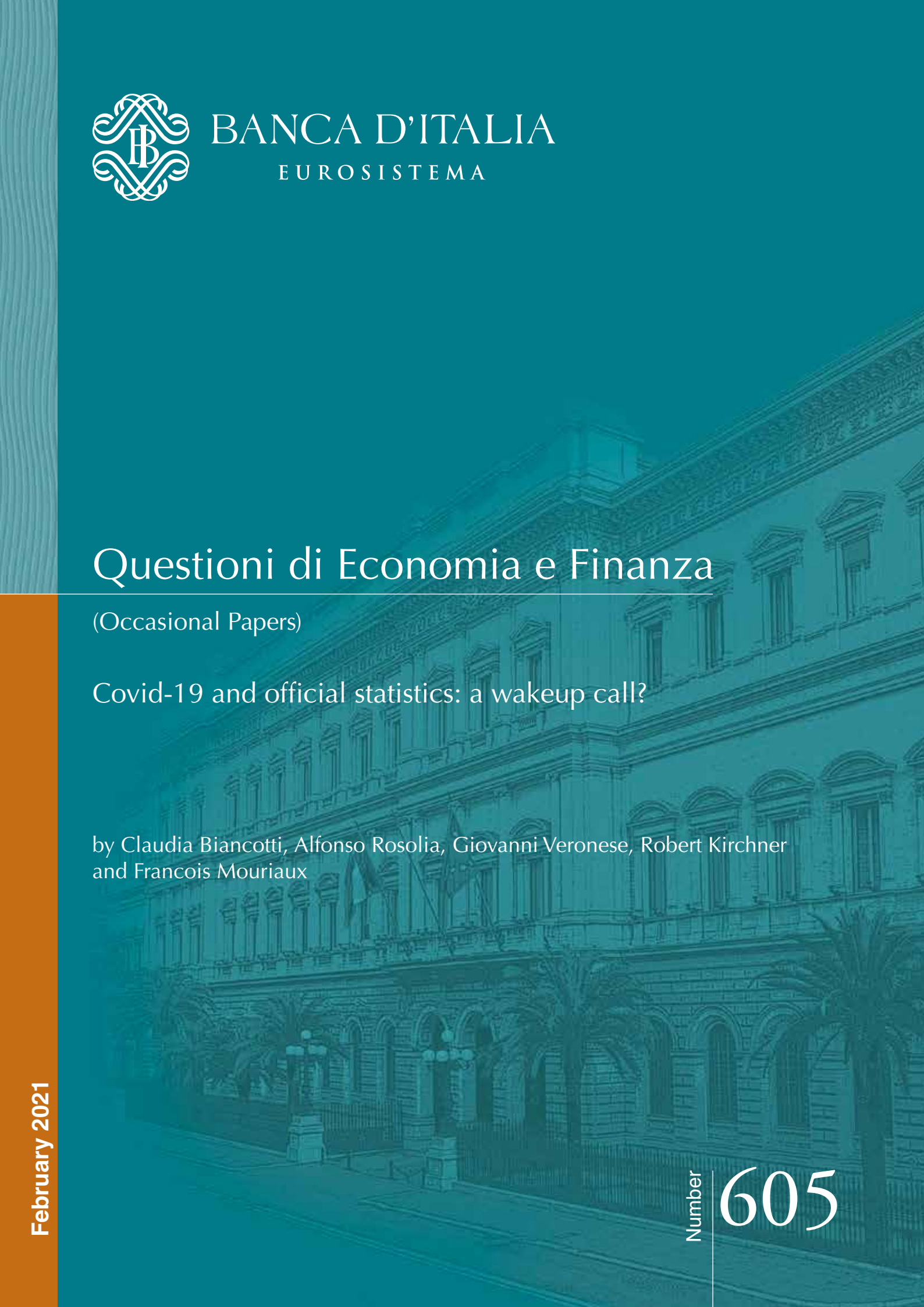

BANCA D'ITALIA
EUROSISTEMA

Questioni di Economia e Finanza

(Occasional Papers)

Covid-19 and official statistics: a wakeup call?

by Claudia Biancotti, Alfonso Rosolia, Giovanni Veronese, Robert Kirchner and Francois Mouriaux

February 2021

Number 605

BANCA D'ITALIA
EUROSISTEMA

Questioni di Economia e Finanza

(Occasional Papers)

Covid-19 and official statistics: a wakeup call?

by Claudia Biancotti, Alfonso Rosolia, Giovanni Veronese, Robert Kirchner and Francois Mouriaux

Number 605 – February 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 .

ISSN 1972-6627 (print)
ISSN 1972-6643 (online)

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

COVID-19 AND OFFICIAL STATISTICS: A WAKEUP CALL?

by Claudia Biancotti*, Alfonso Rosolia*, Giovanni Veronese*,
Robert Kirchner** and Francois Mouriaux ***

Abstract

As COVID-19 spread globally, fast political decisions and the implementation of drastic measures were necessary to slow down proliferation and counter the economic disruption. The demand for broad, timely, high-frequency statistics about economic and health developments surged. At the same time, the pandemic outpaced the frequency at which most conventional statistics become available. Unconventional data helped to bridge these time lags, and to supply information on aspects of society not suitably covered by traditional official statistics, but that the need of the day suddenly made prominent for decision makers. The lesson from the COVID-19 crisis is that greater preparedness and flexibility in facing "future unknowns" is essential. Enabling users of statistics to quickly tap on data dimensions and relationships needed for their decisions when confronted with exceptional circumstances, is essential for guaranteeing salience and, ultimately, trustworthiness of official statistics.

JEL Classification: C82, F60.
Keywords: high frequency statistics, data access, official statistics.
DOI: 10.32057/0.QEF.2021.605

Contents

* Bank of Italy, Directorate General for Economics, Statistics and Research.
** Deutsche Bundesbank, Bundesbank Directorate General Statistics.
*** Banque de France, Directorate General Statistics, Economics and International relations.

Introduction [1]

COVID-19 spread rapidly around the world. Fast political decisions and the implementation of drastic measures were necessary to slow down its proliferation and counter the economic disruption. To prompt and calibrate unprecedented policy measures, the demand for broad, timely, high-frequency statistics about economic and health developments soared.

At the same time, official statistics were faced with disruptions to the data flow from enterprises and households. The rapid diffusion of the pandemic outpaced the frequency at which most conventional statistics become available. Unconventional data and indicators helped to bridge these time lags, and to supply information on aspects of society not suitably covered by traditional official statistics, but that the needs of the day suddenly made prominent for decision makers.

Compared with the 2007/2008 financial crisis, far more granular data are actually digitalized and stored somewhere. Digitalization boosted the amount of data available, expanding the potential information set, not only as to the timeliness and frequency, but also to the many other dimensions previously not even remotely measurable. However, accessing and exploiting data from new information sources often requires access to privately held data and government data. National or regional legal initiatives fostering new frameworks to more effectively access data in the digital age (e.g. as announced by the EU Commission) can be used to bring in the public interest of producers of official statistics (statistical offices and central banks to name a few) with regard to business-to-government and government-to-government data relationships.

The role of official statistics is to provide policy makers and the public with independently compiled and relevant data of high quality, as expressed in the UN Fundamental Principles of Official Statistics. The major lesson from the COVID-19 crisis is that greater preparedness and flexibility in facing 'future unknowns' are essential. Enabling users of statistics to quickly tap into the data dimensions and relationships needed for their decisions when confronted with exceptional circumstances is essential for guaranteeing the salience and, ultimately, the trustworthiness of official statistics.

This calls for a well-defined framework for international and interdisciplinary cooperation. In this perspective, a data initiative after the end of the Second Phase of the G20 Data Gaps Initiative (DGI-2) in 2021 could be very useful for addressing emerging data needs. Access to privately held data and government data could be further elaborated in a follow up of the DGI-2 recommendation. Based on the progress in thematic fields supporting the sharing of micro data (such as anonymization, access regimes, and techniques) as well as on learning from international initiatives of data sharing in non-statistical areas (e.g. payments, supervision and taxation) could be useful drivers. In addition, it would be useful to assess the effectiveness of a statistical data standard, promoted by international organizations and underwritten by states, as an important step towards overcoming the current hurdles to international data sharing and data access while safeguarding data privacy. Fostering standardization (e.g. LEI, SDMX, INEXDA templates) would help economists and researchers to analyse the data and therefore to expedite the process of turning data into knowledge for political decision makers.

To fulfil their role in the digital age, official statistics need to transform the whole statistical value chain: from benefiting from new data for enterprises in the collection phase, more cooperation with private sector data firms, using artificial intelligence in the compilation phase, to standardized data publication and international data sharing in the distribution phase.

[1] The views expressed are those of the authors and do not necessarily reflect those of the affiliated institutions.

An upgrading of the final phase of dissemination and outreach of official statistics – to make them not only more accessible but also more easily understood by the general public– can also act as a first line of defence against misinformation in the digital age. International sharing of experiences on how to improve the use of digital techniques throughout the statistical value chain can inspire statistics producers globally.

1 - The COVID-19 impact on the demand and supply of statistics

The unfolding of the COVID-19 crisis triggered an unprecedented rise in macroeconomic uncertainty, as economic outcomes became closely intertwined with those, unfathomable because of the novelty of the virus, on the health front. This led to an unprecedented demand, from policy makers and users at large, for statistics capable of providing a timely and genuine picture of developments on the ground.

Responding to this demand proved to be particularly challenging for producers of statistical indicators, which include but are not limited to official statistics. This challenge was more acute in ensuring timely delivery of public health related statistics, as well as in providing adequate information on the "real sector" of the economy, including the development on the credit and finance side, notably households and small and medium enterprises (SMEs henceforth). Instead, the business continuity of monetary, banking and financial statistics was largely guaranteed, thanks also to well-established business continuity practices of reporting entities.

1.1 Heightened demand for statistics at large

The rapid diffusion of the pandemic outpaced the frequency at which most conventional statistics become available. Unconventional data and indicators (e.g. from digital platforms, mobile networks, traffic tolls, energy consumption) helped to bridge time lags in the availability of statistics and to support a timelier forecast of economic indicators, as well as to supply information on aspects of society not suitably covered by the official statistics, but that the need of the day suddenly made salient for decision makers.

The mere fact that alternative sources flourished suggests that conventional providers had a difficulty in meeting this increased demand. The proliferation of alternative sources of statistical information raises the question of trustworthiness, in particular of official statistics – not for lack of rigor in providing quantitative objective measures, but rather for their speed in producing knowledge on new phenomena relevant for society itself, especially in times of radical change, and in remaining innovative and communicative vis-à-vis all components of society (a point emphasized by former OECD Chief Statistician Giovannini, 2009). Yet, the magnitude of the crisis exacerbated the expectation of the users – be they the policy-makers or the general public – to have access to timely, if not real-time, and more importantly informative data.

In the performance of their tasks – monetary policy, financial stability - central banks are intensive users of economic statistics, from official and private (e.g. Purchasing Managers' Indices) sources, to analyze the developments of the real sector and the financial sector. In particular, the fulfillment of the mandate of ensuring price stability implies the observation of price developments and data on the economic developments underlying price formation: wage and related labour market data, output gap indicators, etc.

As the COVID-19 crisis suddenly tied public health developments –unsaddled from their slow moving trends– to economic conditions, also health-related statistics became essential information to central banks. For example, meaningful macroeconomic projection exercises had to be based on scenarios that properly accounted for the developments on the public health front and their interaction with policy decisions on social distancing.

Detailed mobility information, until the pandemic hardly of interest for macroeconomic policy, has suddenly become essential to understanding consumer behaviour, or to gauging the impact of social distancing on different occupations and the future of working from home in different sectors.

Going forward, a key challenge for policymakers is to understand the structural implications of the COVID-19 crisis to assess the shape and speed at which structural changes will take place and their distributional implications. In this respect, the multifaceted economic implications of the digital transformation accelerated by the epidemic remain something that official statistics can inform us on, albeit difficulties already known before Covid remain (e.g. valuing intangible assets to name one) and are becoming more prominent today. The role of official statistics will have to rise to the challenge.

1.2 The supply of official statistics

The pandemic had a direct impact on European official statistics.[2] In many EU countries the health and public safety measures introduced by national governments interrupted regular face-to-face interviews, slowed down the flow of data from reporting agents and generated disruptions also to primary data coming from administrative registers. Under these conditions, Eurostat (the EU statistical office) and the providers of national statistics (National Statistical offices, Central Banks, etc.) have striven to continue publishing relevant statistics, according to their original release calendars. In particular, efforts were concentrated on providing statistics to track the evolving situation from March 2020 onwards and the impact of the COVID-19 outbreak.

Most NSOs developed innovative approaches to overcome the difficulties in providing all of the data with the conventional processes, since some categories of survey respondents were unable to report and certain surveys could not be carried out entirely.[3] Some large-scale in-depth surveys did not take place according to the calendar and had to be postponed entirely.[4]

The restrictions in the movement of people and the closure of outlets had a direct and indirect impact on household consumption, both in terms of level and composition, and thus on consumer prices in many EU countries. Data from e-commerce platforms could have been useful to statisticians seeking to understand how consumption habits were changing and how prices were evolving in the few markets that remained open, also complementing the informative value of official indicators such as the Harmonized Index of Consumer Prices (HICP) during lockdowns.[5]

Statistical production by Central Banks was also strained, especially for what regards balance of payments, surveys on household consumption, and non-financial companies. In the domains based on the compilation of multi-source statistics such as the balance of payments, which rely heavily on surveys, efforts were devoted to the substitution of the unavailable data sources with alternatives (e.g. using mobile phone data to estimate travel) or to account statistically for the more or less extensive under-reporting of data. It was also observed that the potential to maintain continuity of surveys was much better for surveys managed by phone or mail[6].

However, as regards monetary, banking and financial statistics, it seems that only limited part of the processes has suffered from disturbances. As a matter of fact, the key reporting entities are financial institutions that have well-established business continuity practices.[7] Understandably, many respondents asked for temporary reliefs

[2] Other advanced economies faced similar problems: the United States Census Bureau has announced a postponement of 2020 census activities, and the UK Office of National Statistics has pre-emptively informed the public of anticipated delays in regular data publication in the wake of the crisis. See also the United Nations Committee for the Coordination of Statistical Activities (CCSA) report.

[3] See for example the INSEE article on how French Statistics responded to the COVID-19 operational challenge, https://blog.insee.fr/official-statistics-and-the-challenge-of-the-current-health-crisis/.

[4] This was the case of the annual EU survey on income and living conditions (EU-SILC) for which data collection was postponed by about half a year, as it requires an initial "face-to-face" survey interview with the household at home; in some countries, specific indicators were not published despite efforts to guarantee the business continuity. For example, in April Istat suspended entirely its consumer and business confidence surveys, cancelling the update scheduled for the end of April, while the May survey resumed regularly. See: Istat press release, April 2 2020. https://www.istat.it/en/archivio/240812 and Istat press release, May 28 2020, https://www.istat.it/it/archivio/243511.

[5] According to calculations by Eurostat, prices of items accounting for about 32% of consumption expenditure of euro area households were imputed due to the corona crisis in April 2020. In the April and May 2020 HICP press releases, Eurostat published an explicit disclaimer informing users that the price collection had been affected. See Eurostat, Information on imputations made related to COVID-19, https://ec.europa.eu/eurostat/web/hicp/methodology.

[6] This was notably the case as regards Banque de France surveys for balance of payments, where the responding rate and quality of data matched the expected standards for the monthly reports from large non-financial companies.

[7] Reporting obligations helped to guarantee the flow of information to Central Banks.

in their reporting obligations. Therefore, the key challenge was to communicate clearly with the respondents about the importance of the continuity of statistics.[8]

2 – Lessons from the crisis

The COVID-19 shock has once again underscored that when facing unprecedented events (the pandemic, the global financial meltdown, 9/11, etc.) both public and private decision makers suddenly realize gaps in their information sets. Filling these information gaps may require analyzing previously less important dimensions of commonly charted phenomena, as the need of higher frequency information on the real economy to track the quick unfolding of the Covid19 shock. It may also require analyzing new data domains, as epidemiological data, and social distancing and mobility indexes in the case of Covid19, or risk exposures and within- and cross-border financial connections data in the case of the global financial crisis.

Going forward, it is very likely that - having hopefully leveraged the lesson of this one - our societies will have readily available statistics to address the next pandemic. It is also very likely that a currently unfathomable shock (thus, not a pandemic nor a global financial crisis) will reveal other major information gaps.

Yet, a more general lesson to be drawn from the COVID-19 (and from the Global financial crisis) is that these newly discovered information gaps not necessarily reflect data gaps, that is the very lack of relevant data. In an information society well beyond its infancy, such as ours, data are constantly generated and being collected and stored by public and private actors in pursuing their endeavors. Be it public agencies recording administratively relevant events, private businesses collecting data relevant to their functioning, individuals connecting to mobile networks, weather stations automatically recording environmental data, satellite imagery, data exist on many more phenomena and along many more dimensions than those commonly exploited by analysts, statisticians and decision makers.

During the COVID-19 crisis, some of these non-standard data sources or data dimensions were, not always easily, exploited and helped statistics producers and decision makers to fill in some gaps, either by shedding a light on relevant, albeit uncharted domains, or by complementing and supporting the production process of official statistics (also see Annex 1). The fact that these non-standard data sources (at least from the perspective of the production and use of economic statistics) were commonly used by actors often far away from the community of economic analysts and statistics producers helped in establishing a dialogue and setting up exchange frameworks.

Therefore, albeit unpredictable as of today, the information gaps that, going forward, will arise when facing the next unknown unknowns will be more easily filled in if a general framework has been set up that allows public and private users to quickly tap on data dimensions and relationships that will have turned out to be salient in those circumstances, be it a higher frequency of measurement of common phenomena or unusual statistical domains required by the circumstance.

Setting up this framework needs not start from scratches. For one thing, data generated within the perimeter of governments are increasingly exploited for analytical and statistical production purposes and experiences with linking even very distant data sources are rapidly growing also thanks to the fast development of Machine Learning (ML) and Artificial Intelligence (AI) techniques. Also, legal initiatives fostering access to existing information already held by public as well as private actors are a promising way forward (e.g. the EU Commission initiative towards a Data Act).

2.1 – Administrative data – opportunities and challenges

In pursuing its many undertakings, the public sector generates huge amounts of data covering a wide spectrum of phenomena. While some of this data are disseminated to the broader public in the form of aggregate statistics

[8] See for instance the ECB communication on reporting agents:https://www.ecb.europa.eu/pub/pdf/other/ecb.200415_communication_reporting_agents_on_statistical_info_covid~563fccc78b.en.pdf.

or shared with NSIs and CBs to be used in their statistical production process, this wealth of information remains in general largely underused.

First and foremost, the use of many of these data sources may be subject to legal limitations. For example, it may be the case that the law states that data collected by a certain administration can be only used to pursue that administration's tasks; similarly, individual data may be subject to privacy regulation. Second, existing exchanges with NSIs and CBs are usually governed by clearly stated rules about, for example, content, frequency, and dissemination delays, that ensure compliance with relevant regulation. These are however the result of often lengthy interactions and as such are difficult to adapt to fast evolving environments and information needs of decision makers and the society. Third, data collected for administrative purposes are often not immediately serviceable for economic analysis and statistical production. A clear understanding of the rules that govern data collection and of the underlying administrative process is required to effectively bridge the gap between the administrative data and the statistician's or economist's object of interest. Fourth, these difficulties are often compounded by the collecting administrations' failure to recognize the broader interest that the data may have for statisticians, economists and analysts. This may lead to data management practices, for example in terms of storage and documentation, which mostly respond only to the primary purpose of the data collection and significantly limit their re-usability.

While the need of real-time high-frequency data for monitoring and policy purposes may be specific to the current emergency, being able to quickly access previously untapped sources already maintained by the public sector in order to fill in new information gaps and guide policy making has a value that certainly goes beyond it. Statistical agencies and central banks as well as the research community and the general public would greatly benefit from active dissemination strategies, even if these were not specifically tailored to their information needs. For example, public bodies collecting data would greatly help the community by sharing at suitable frequencies aggregations of their data constructed so as to strike an efficient balance between the need of preserving individual confidentiality and that of maximizing the information content of the aggregations being shared.

Clearly, more public discourses are needed to make public bodies aware of the value of their data, beyond the purposes they were originally collected for. An open dialogue between users and the public bodies disseminating their administrative data could help fine-tuning what is being shared and define the data access regimes necessary for preserving privacy.

2.2 – Private sector data – opportunities and challenges

A plethora of alternative data sources also lie in the hands of the private sector. Large internet platforms have access to information on human behavior that is unprecedented in terms of size and scope. They also possess troves of highly detailed non-personal information, e.g. image data underlying mapping services. They have the ability to collect information on individuals in their digital life, but also in many dimensions of the physical, ranging from where they go to how fast their heartbeat is. These "datasets of everything" make for a prime resource to tackle unknown unknowns – whatever the shock that hits us next, chances are that the dominant platforms have access to information related to it. Most importantly, such datasets allow to explore correlations across different domains at a speed that would not be otherwise possible. In addition to these behemoths of multidimensional information, the private sector can also provide smaller, more specialized datasets (e.g. from mobile phone companies, or platforms that specialize in a single activity) that can be extremely useful for subject-specific deep dives.

These data can fruitfully complement traditional sources of statistics such as censuses, surveys and administrative systems. However, with scant governance frameworks to enable the scaling up and persistence in their use, the engagement of official statistics producers with the private sector is still underdeveloped, and fragmented. Statistical offices thus need to engage in sustainable partnerships that offer a path toward coordinated flows of high quality data to inform the policy-making.

What is the value that can be extracted from these new sources of private data? There are at least five dimensions of value, which can increase the "*salience*" of official statistics: i) they can increase the *scope and breadth* and thus the insights offered by existing evidence for policymakers; ii) they can enhance the *quality* of existing data (cross-validation); iii) they can increase the *timeliness* and relevance of more outdated information in official statistics (social media streams, for example, can provide real-time insights into public behavior); iv) they can *lower costs and increase other efficiencies* for statistical organizations; v) they might *lower the administrative burden* for society in the medium term since official surveys – while remaining essential as an "anchor" for all other datasets – could go through a new cycle of reengineering, such as a review of frequency, size of sample, modalities of data collection…-. Examples of benefitting from private sector data in statistics are given in Annex 2.

The greatest societal value from accessing these kinds of private data can be derived from *reusing* and *combining* previously separate datasets.[9] However, performing this blending across datasets can be very challenging. Significant interoperability issues still impede the combination of data from different sources within sectors, and even more so between sectors. Furthermore, linking datasets from different sources can be problematic if this linking needs also to ensure individual anonymity.

The need to leverage privately collected data to improve the timeliness and accuracy of official statistics has not been mentioned often in the big-picture discussions around fair data use, but it might just be the type of argument that appeals to most if not all sides – at least within Europe – and allows different stakeholders to build a sharing framework that reflects most if not all concerns (see Annex 3 for a brief review of related literature).

3 - Solutions and crosscutting issues to be addressed by the international community

The COVID-19 pandemic has demonstrated the need for flexibility and adaptability on the part of official statisticians, especially when it comes to quickly acquiring and integrating data from multiple sources. This raises the fundamental long-term question of how to transform the whole statistical value chain in the information society: from making use of real time data in enterprises in the collection phase via using artificial intelligence in the compilation phase to standardized data publication and international data sharing in the distribution phase.

However, making use of such data typically requires access of official statistics to privately held data and government data. National or regional legal initiatives fostering digitalisation, as those recently announced in the European Union (EU Commission 2020a) shall be used to bring in the well justified public interest of statistical offices and central bank statistics with regard to business-to-government and government-to-government data relations.

Digitalisation is an investment on the side of reporting companies, but also in statistical offices and central banks. Step-by-step procedures, starting with pilots (e.g. cooperation with enterprises which have an interest in changing/revising their IT) for specific statistics and learning from experience would help to manage the process and develop more general timetables. In the transition period of the years to come, statisticians and users of statistics will have to deal with a mixture of different stages of developments: digital high frequency data deliveries/statistical products and traditional reporting[10].

[9] The same value has been attributed to combining dataset compiled by official statistical agencies. See the recommendation on "linking datasets" made as a part of the G20 data Gaps Initiative and included in the Report to the G20 FMCBG in Baden-Baden (March 2017), available at http://data.imf.org/api/document/download?key=61400076, p. 11 ff.

[10] Digitalisation is an accelerator for the production of timely and high frequency data. For example, the implementation of a flexible and adjustable production system, built on a microservices architecture, allows the primary statistical value chain to be digitalized and to be produced in a more timely manner. Within such a system, single services (microservices) can act independently from each other to fulfill specific tasks along the value chain. Therefore, the system stays flexible in the future even when external and internal requests or an increase in data quantity require the production system to be

The way forward could benefit from international cooperation. In this perspective, a data initiative after the end of the Second Phase of the G20 Data Gaps Initiative (DGI-2) in 2021 could be very useful to address emerging data needs. Access to privately held data and government data could be further elaborated in a follow up of DGI-2 recommendation. Based on the progress in thematic fields supporting the sharing of micro data (like anonymization, access regimes, and techniques) as well as on learning from international initiatives of data sharing in non-statistical areas (e.g. payments, supervision, and taxation) could be useful drivers. In addition, it would be useful to investigate whether a statistical micro data standard, promoted by International Organizations and signed by jurisdictions, could be an important step towards overcoming current hurdles of internationally data sharing and data access while safeguarding data privacy.

The way forward would also benefit from strengthened and streamlined interdisciplinary co-operation. COVID-19 provided a clear example of how data from different domains must be analyzed together to gain an understanding of new phenomena and make sound policy decisions. In this specific case, economic forecasting could not do without health data. Future shocks may require the integration of entirely different sets of disciplines. A framework for cross-field collaboration and data sharing is required.

Sharing experiences internationally on how to improve use of digital techniques through all the statistical value chain, including those on how to adapt the outreach capacities of official statistics in the world of mass data, could inspire data producers all over the world. Finally, yet importantly, fostering standardisation (LEI, SDMX, INEXDA template for micro datasets) would help economists and researchers to analyse the data and, therefore, to speed up the process of turning data into knowledge towards political decision makers.

4 - Conclusions

Official statistics have an important mandate in modern society: both policy makers and, via the media, the general public needs to be informed in an independent and transparent way about the state of affairs and ongoing developments. This is as a fundamental precondition for rational and efficient decision making by policy makers and choices (economic and beyond) by the general public.

In recent years, two important changes have come to challenge the status quo. First, important questions come and go at a fast pace today: the Great Financial Crisis has posed many new and fundamental questions, and the current pandemic adds new dimensions. The time to market for new statistical answers is rather long. Second, digitalization has created both a host of new technological opportunities and a large space of information accessible to the public that is an additional source to the information provided by official statistics.

Official statistics need to embrace those new possibilities – by deploying new technologies in its production processes, becoming both more flexible and faster, and by making more use of "alternative" data sources – being called alternative mainly because they are only partly being used by statisticians. Going this way, we will have to find new ways of coordinating with market participants, such as tech companies, and with each other, by working out and implementing micro data standards.

Statistical production should also meet the changing needs of the general public. Shocks and developments, be they economic or not, may change even dramatically the information required by the public to make informed decisions. These needs should be met by statistical producers, who should broaden their mission

adjusted over time. Deployments could be rolled out within minutes due to test automation and is focused only on specific microservices modules rather than the whole production system. Finally, machine learning algorithms (built as a microservices modules) can be integrated, in the production cycle to support and to speed up data quality management (see T. Cagala (2017), Improving Data Quality and Closing Data Gaps with Machine Learning, IFC Bulletin, 18-19- May, Brussels.) Bundesbank already had prototyped such a flexible system in security statistics and started the project "Digital Statistics" for the fully productive implementation and operation. First indicative results suggest that for selected primary statistics data processing could be reduced by approximately 75%, i.e. from a couple of weeks to a few days.

beyond the, albeit more efficient and timely, production of official statistics to unofficial domains that may become relevant from time to time to the public and to policy making bodies. This agility can be achieved by building experience with unofficial/unconventional data sources and methodologies and developing a shared framework to pool the huge amounts of data produced daily by our digitized societies (Bean, 2016).

Last but not least, in an age characterized by extensive access to multiple sources of information, official statistics need to be disseminated more effectively, to make them not only more accessible but also better understood by the general public. Good statistics, as dramatically shown during the COVID-19 crisis, can be crucial to provide a first line of defense against misinformation (Biancotti et al., 2020). While data enable it, by multiplying the opportunities for the targeting of fake news and divisive messages, they can also provide an antidote.

Ahead of us is a fundamental choice: Embracing the new options will increase the social benefit of statistics. We will be faster, more flexible and more relevant to politics and the society at large.

Annex 1 - Examples of higher frequency data used or developed during the pandemic

"Non-standard" and timely information was already being used inside Central Banks. This was the case for the Bank of Italy, where nowcasting models including electricity consumption already had a long-established tradition (Bodo and Signorini, 1987), and information on gas pipelined to the industrial sector, car and freight traffic along Italian highways and POS transactions had been exploited before the health emergency (Aprigliano et al., 2019).[11] At Banque de France, credit card data and mobile phone data have being used for estimates of the travel item of the Balance of Payments for several years.[12]

The crisis spurred efforts to improve the availability of such "non-standard" indicators with respect to timeliness and frequency. The German truck toll mileage index is a case in point. While monthly data were published since 2018, a daily index was developed at the beginning of the crisis which is updated every working day with a publication lag of five to nine days to provide information as early as possible.[13] In general, an advantage of improving existing indicators/statistics over developing new indicators is that they (i) often date further back in time and (ii) are already well-established and proven for monitoring economic developments. Nevertheless, the availability of such indicators with a higher frequency came along with new challenges, for example related to the seasonal and calendar adjustment of daily or weekly data.[14]

As for implementing ad hoc surveys, an example was the one conducted by the Bank of Italy to collect a broad range of qualitative data on households' economic situation and expectations, as the standard information released by Istat on consumer confidence was missing, after the relative survey was temporarily interrupted in April.[15] At the Banque de France, supplementary questions have been added to the Monthly Business Survey, about the loss in activity compared to normal, the outlook for a return to normal or the use of remote working.[16] Moreover, text mining techniques were applied to the additional comments collected in the context of the Monthly Business Survey.[17]

Apart from the processing of data at a frequency far higher than usual, the COVID-19 crisis prompted the analysis at a more granular level, e.g. the regional dimension to capture the geographical evolution of the crisis. Following the COVID-19 crisis, also experimental synthetic indicators to measure real economic activity in a timely manner were introduced, both at the Bundesbank and at the Bank of Italy. The Weekly Activity Index (WAI) for the German economy consists of nine high-frequency indicators, such as electricity consumption, toll (road charge), air pollution, flight indicators and Google search terms alongside monthly industrial output and quarterly GDP.[18] Similarly, the Italian Weekly Economic Index (ITWEI) is based on ten variables

[11] For more details on how the Bank of Italy coped with forecasting in the context of the health emergency, see https://www.bancaditalia.it/media/notizie/2020/Previsioni_al_tempo_del_coronavirus_Locarno_Zizza.pdf (in Italian).

[12] An assessement of the use of mobile phone data for balance of payments has been published in Eco&Stat https://www.insee.fr/fr/statistiques/3706178?sommaire=3706255

[13] See Cox, Michael/Triebel, Jürgen/Linz, Stefan/Fries, Claudia/Flores, Luis Federico/Lorenz, Andreas/Ollech, Daniel/Dietrich, Andreas/LeCrone, Julian/Webel, Karsten. *Daily truck toll mileage index based on digital process data from toll collection system* In: WISTA Wirtschaft und Statistik. Ausgabe 4/2020, page 63 ff (Only available in German), https://www.destatis.de/DE/Methoden/WISTA-Wirtschaft-und-Statistik/2020/04/lkw-maut-fahrleistungsindex-042020.pdf?__blob=publicationFile .

[14] See Deutsche Bundesbank, Experimental seasonal and calendar adjustment of daily economic indicators, in: Monatsbericht, August 2020, page 67ff.

[15] See https://www.bancaditalia.it/pubblicazioni/relazione-annuale/2019/sintesi/2019-box/CH5-Box8.pdf and https://www.bancaditalia.it/media/notizie/2020/Evi-preliminari-ind-straord-famiglie.pdf (in Italian).

[16] See https://www.banque-france.fr/en/statistics/business-surveys/business-surveys and https://www.banque-france.fr/en/economics/macroeconomic-projections-france.

[17] See Gerardin, M. and M. Ranvier (2020). https://blocnotesdeleco.banque-france.fr/en/blog-entry/how-firms-are-adapting-lockdown-insights-text-mining.

[18] See Deutsche Bundesbank, Weekly activity index for the German economy, https://www.bundesbank.de/en/statistics/economic-activity-and-prices/weekly-activity-index

accounting for different aspects of the economy.[19] In contrast, at Banque de France the joint use of credit card data and electricity consumption allowed the identification, at a very early stage, of the decoupling between the consumption of households and the output of enterprises. [20] Cellular mobility data was successfully used for the compilation of balance of payments statistics at the Bank of Italy, which has recently started to use properly aggregated and anonymized mobility information to estimate the "foreign travel" item of flows.[21]

Annex 2 - Examples of benefitting from private sector data in statistics

In Europe pilot applications of using data from private sources exist, and in some cases have become a standard input for official statistics. For example, Eurostat and Statistics Belgium have experimented with leveraging call detail records provided by the major Belgian telecom network operator. The project not only revealed the complementarity of this information to the one obtained from traditional census statistics, but also showed it could be used to increase the ability of real-time monitoring of economic activity.[22] Statistics Netherlands (CBS) recently launched a Center for Big Data Statistics (CBDS) in partnership with many large private companies.[23] Other NSOs are already using scanner data from supermarkets for monitoring consumer prices and as an input for the computation of consumer price indices (Istat in Italy, Destatis in Germany and Insee in France).

A recent milestone was reached with the agreement last March between Eurostat and Airbnb, Booking, Expedia and Tripadvisor to permit access to reliable and unique data about holiday and other short-stay accommodation offered in these platforms.[24] Similarly, the Deutsche Bundesbank regularly analyses prices of transactions for package tours purchased by German tourists based on a contract with an IT provider for the travel and tourism industry.

As to statistics produced by Central Banks, an example of how private data freely accessible through the internet are used for the balance of payments purposes is an approach developed by the Deutsche Bundesbank in cooperation with the Federal Statistical Office to estimate cross border services digitally received by households.[25] Unlike services provided or delivered by enterprises, services received by households usually remain far below the existing statistical reporting thresholds in terms of value and therefore cannot be collected via traditional sources. However, although the value of each transaction is very small, they can sum up to billions of Euro and will certainly become more important in reaction to the COVID-19 pandemic. However, the use of such alternative data sources to enhance and complement official statistics is risky because private providers can stop at any time the publication of the necessary information. It is therefore of importance that the data access is either contractually secured, which could be costly, or is mandatory.

Private data from mobile operators or internet platforms is not the output of instruments designed to produce valid and reliable data amenable for scientific analysis (Lazer et al. 2014), nor for official statistics, but instruments and scientific methods exist to bridge these gaps (Gelman and Hennig, 2017). In this respect, the main issue concerns the quality of private data. Data held by private companies often represents a particular demographic subset, while ignoring others, creating so-called "data invisibles". Caution needs to be exercised

[19] The dataset includes gas pipelined to the industrial sector, consumption of electricity, POS transactions, traffic flows on motorways, value added tax paid on imports, three expenditure indexes, the purchasing managers' indices for both manufacturing and services. See Delle Monache et al. (2020).

[20] Banque de France, 2020 (https://blocnotesdeleco.banque-france.fr/en/blog-entry/covid-19-and-monitoring-economic-activity-contribution-high-frequency-data).

[21] See https://www.bancaditalia.it/media/notizie/2020/Nota-Covid19-Viaggi-BoP.pdf (in Italian).

[22] See https://ec.europa.eu/eurostat/cros/content/assessing-quality-mobile-phone-data-source-statistics_en.

[23] See https://www.cbs.nl/en-gb/about-us/innovation.

[24] See https://ec.europa.eu/commission/presscorner/detail/en/ip_20_194

[25] See Deutsche Bundesbank, Households' digital purchases in the balance of payments, in: Monthly Report, March 2020, available at: https://www.bundesbank.de/en/publications/reports/monthly-reports/monthly-report-march-2020-828872

in extrapolating general observations, with consequences for the population at large, from such data. In addition, the time-series length of the data collected by private companies is often too short –at most limited to the period in which the particular company is active– complicating use in statistical models.[26]

Furthermore, the metadata accompanying this kind of private digital data differ varies greatly. This contrasts with the traditional production of official statistics, where substantial investment is made on the system of metadata providing rigorous and unambiguous definitions for each of the variables collected. Instead, digital data is generated without statistical metadata, and embrace variables with a vague connection with target indicators and aggregates. Digital data is generated to provide some kind of service, extraneous to statistical production. Thus, for the new data to be used in the production of official statistics, a non-negligible amount of preprocessing is required before effectively incorporating these new sources of information.

Annex 3 – A brief review of the literature on data ownership and sharing

As digitalization progresses, the collection and analysis of data becomes a key element of productive processes in all sectors (Aronson and Cowhey, 2017). While economists have long been attuned to the fact that information is valuable, today's datasets are nothing like pre-smartphone ones, and understanding their social and economic role requires a different toolkit.

Foundational questions must be asked all over again. For decades, the data sphere in the private sector was ruled by a major unspoken assumption – information belongs to those who collect it. As data becomes more abundant and more detailed, and the platform model takes hold in the internet economy, unease with this assumption is growing. It has been alternatively challenged as unfair to data subjects, leading to inefficient economic outcomes, and overlooking the nature of data as a public good.

There is a diffuse sense that something must be done to address these issues, although no consensus emerged as of yet on solutions. In the following, we briefly outline two ongoing debates closely linked to policy decisions on access to and sharing of data: they concern (a) property rights over data and (b) competitive implications of data accumulation. These discussions do not directly address the needs of statisticians, yet still serve them by shedding light on the nature of contemporary datasets and on the limits of the status quo.

Property rights

A sizable share of the fuel for today's information economy consists of personal data, i.e. data that concerns individuals. Digital companies routinely collect information on their users and deploy it towards a number of profitable ends, ranging from product improvement to ad targeting. Is it fair that they appropriate the attendant revenues in their entirety?

The "data dignity" movement, pioneered by Lanier (2013) and popularized by Posner and Weyl (2018), posits that income from personal data ultimately belongs with the data subjects. Internet companies, the argument goes, practice what is tantamount to exploiting free labor. Instead, they should compensate users – with money, not with online services – for their data. Union-like organizations should mediate negotiations.

This view is not exempt from drawbacks. What one person uploads online not always pertains to them alone (see e.g. Immorlica, Jackson and Weyl 2019), making it difficult to establish clear property rights. Private data markets are also liable to reinforce digital inequalities, since low-income individuals are less likely to opt out of selling their privacy compared to wealthier ones. Nonetheless, the data dignity angle captures a very real problem of asymmetry in power and information between data collector and data subject.

An alternative approach to the same problem, which takes a macro perspective instead of a micro one, uses fiscal policy as a conduit. In the recent past, the European Commission (2018) proposed a tax on revenues that internet companies extract from data-centric operations, as a way of giving back some of the value to the

[26] For example, restaurant reservations collected by OpenTable, were particularly useful to gauge how social distancing impacted restaurants activity, but the time series depth of the data was inherently too short for statistical usage.

users[27]. The proposal did not go through as a EU-wide measure, but some member states have introduced policies to this effect. A similar policy, called a digital dividend, was proposed in California in 2019.

Competition

A variety of contemporary productive processes build on data-intensive technologies, such as machine learning. Many valuable datasets enabling such technologies, however, are concentrated in the hands of few companies. According to one school of thought, this constitutes a barrier to entry that stifles competition, innovation, and growth (e.g. Autorité de la Concurrence and Bundeskartellamt, 2016; Bourreau, de Streel, and Graef 2017). A textbook market failure should be rectified by public intervention.

To quote just a few of the many policy proposals in this field, Hall and Pesenti (2017) suggest the introduction of "data trusts", i.e. standard contractual agreements for the sharing of data between entities, guaranteed by "neutral and expert organizations" acting as trusted third parties. Mayer-Schonberger and Ramge (2018) go one step further, suggesting that large platforms such as Google, Facebook and Amazon should be mandated by governments to share at least part of their data with competitors[28].

Several other reflections on the subject of data and non-competitive rents – also focused on traditional antitrust solutions such as the breakup of internet oligopolies, or the recasting of data as essential facilities – have been carried out in academia, the public debate, and governmental task forces. One notable example is the long series of public hearings on "Competition and consumer protection in the 21st century" organized by the US Federal Trade Commission in 2018 and 2019.

A broader policy challenge

The debates sketched above are but a part of a broader discourse also relating to privacy, security, cross-border trade in data, and the governance of cyberspace. A balance must be found that allows the public to benefit more from privately collected data, while maximizing incentives for innovation and safeguarding individual rights (OECD 2019). There have to be open discussions about market failures, data responsibility, and the balance between technology and law in strengthening user control over personal information.

References

Aprigliano, V., G. Ardizzi and L. Monteforte (2019) "Using Payment System Data to Forecast Economic Activity", International Journal of Central Banking, vol. 15(4), 55-80.

Aronson, J. and P. F. Cowhey (2017), Digital DNA: Disruption and Challenges for Global Governance, Oxford University Press.

Autorité de la Concurrence and Bundeskartellamt (2016), Competition Law and Data.

Bean, C. (2016), Independent Review of UK Economic Statistics.

Biancotti, C. and P. Ciocca (2019), Opening Internet Monopolies to Competition with Data Sharing Mandates, Peterson Institute for International Economics Policy Brief 19-3.

[27] "The Digital Services Tax is a tax with a targeted scope, levied on the revenues resulting from the supply of certain digital services characterised by user value creation. […] There are several ways in which user participation can contribute to the value of a business. For instance, digital businesses can derive data about users' activities on digital interfaces, which is typically used in order to target advertising at such users, or which can be transmitted to third parties for consideration. Another way is through the active and sustained engagement of users in multi-sided digital interfaces, which build on network effects where, broadly speaking, the value of the service increases with the number of users using the interface." (European Commission, *ibid.*)

[28] Data sharing mandates of varying intensity are an interesting solution from the point of view of competition, and also with a broader look at reducing asymmetries in the digital world. However, trade-offs with privacy must be carefully evaluated (Biancotti and Ciocca 2019), as re-identification of individuals in highly detailed, highly multidimensional data sets is still too easy.

Biancotti, C., A. Rosolia, F. Venditti and G. Veronese (2020), Saving Economic Data from COVID-19.

Bourreau, M., A. de Streel, and I. Graef (2017), Big Data and Competition Policy: Market Power, Personalized Pricing and Advertising. Brussels: Centre on Regulation in Europe (CERRE).

Bodo, G. and L.F. Signorini (1987) "Short-term forecasting of the industrial production index", International Journal of Forecasting, 3, issue 2, p. 245-259.

Cagala, T. (2017), "Improving Data Quality and Closing Data Gaps with Machine Learning", IFC Bulletin, 18-19- May, Brussels.

European Commission (2018), Proposal for a Council Directive on the Common System of a Digital Services Tax on Revenues Resulting from the Provision of Certain Digital Services.
European Commission (2020a), EU Communication "A European strategy for data" COM(2020) 66 final.

European Commission (2020b) EU Report of the Expert group on B2G data sharing, https://ec.europa.eu/digital-single-market/en/news/meetings-expert-group-business-government-data-sharing.

Delle Monache, D., S. Emiliozzi and A. Nobili (2020) "Tracking economic growth during the COVID-19: a weekly indicator for Italy", Bank of Italy, manuscript.

Gelman, A, and C Hennig (2017), "Beyond subjective and objective in statistics", *Journal of the of the Royals Statistical Society,* 180: 967-1033.

Giovannini, E. (2009), "Measuring Society's Progress: A Key Issue for Policy Making and Democratic Governance". https://www.oecd.org/site/progresskorea/globalproject/41684236.pdf

Hall, W. and J. Pesenti (2017), Growing the Artificial Intelligence Industry in the UK, UK Department for Digital, Culture, Media & Sport and Department for Business, Energy & Industrial Strategy.

Immorlica, N., M. O. Jackson and E. G. Weyl (2019), Verifying Identity as a Social Intersection, mimeo.

Lanier, J. (2013), Who Owns the Future?, Simon and Schuster.

Lazer, D, R Kennedy, G King and A Vespignani (2014), "The Parable of Google Flu: Traps in Big Data Analysis", *Science* 343(6176): 1203-1205.

Mayer-Schonberger, V. and Ramge, T. (2018), A Big Choice for Big Tech, Foreign Affairs.

OECD (2019), Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-Use across Societies.

Posner, E. A. and E. G. Weyl (2018), Radical Markets: Uprooting Capitalism and Democracy for a Just Society, Princeton University Press.

United Nations Committee for the Coordination of Statistical Activities (2020), https://unstats.un.org/unsd/ccsa/