Data Governance: issues for the National Statistical System

Irena Krizman and Bruno Tissot[1]

Executive summary[2]

The IFC together with the International Statistical Institute (ISI) organised a High-Level Meeting on Data Governance in Tunisia on 22 November 2019. The event was a key occasion to discuss the issues faced by public institutions in general, and national statistical authorities in particular, **when dealing with, and using, official data**. It also proved to be a useful opportunity to show how statistics can play a decisive role in measuring, monitoring and evaluating the implementation of major international initiatives supporting development. This is particularly the case in respect of United Nations (UN) 2030 Agenda for Sustainable Development Goals (SDGs), which requires countries to produce a standardised list of indicators, and the Agenda 2063 for the development of Africa.

A first main message from the meeting is **the need to have an all-encompassing approach to data governance when collecting, managing, disseminating and making use of official statistics.** This should cover all the related principles, policies and procedures, structures, roles and responsibilities.

Another key lesson is the importance of proper data governance frameworks for those organisations composing national statistical systems (NSSs), especially National Statistical Offices (NSOs) – or National Statistical Institutes (NSIs) – and central banks' statistical departments,[3] to **reap the full benefits of the ongoing "data revolution"**. Such governance frameworks should cover the entire organisations and be an integral part of their strategic plans.

As regards first **data collection**, traditional statistical surveys and censuses can be usefully complemented with new types of information, eg alternative data sources including administrative records and "the internet of things" (big data) – described by some as the new oil of the 21st century (The Economist (2017)). This can be a great opportunity for those less developed statistical systems, not least considering the high costs associated with setting up and maintaining standard exercises. One risk from this perspective is the hoarding of the vast data resources collected outside the

---

official perimeter of statistics if one fails to incorporate them properly to support the measurement of economic indicators.

Second, there are clear challenges related to NSSs' **management of the evolving data ecosystem**. In particular, what is unclear is how to deal with "organic" information sources, whether private commercial data sets or public registers that were not initially set up for a statistical purpose, and which may not pass the test of time. Sticking to long-established and internationally agreed practices and standards, preserving sufficient "traditional" statistical capacity in the NSS and favouring a complementary use of both traditional and alternative data sources are **central to maintaining well-founded trust in official statistics**.

Third, turning to **data dissemination**, digitalisation techniques allow for easier, almost cost-free access to information for the public. However, the increasing complexity of economic and financial activities in a data-rich world puts a premium on statistical education and financial literacy. In addition, official statistics are essential to provide reference, objective information and in turn support economic development and well-being.

Fourth, there has been a growing interest globally **for the better use of data for policy purposes**, especially when designing, calibrating, assessing and modifying policy actions. But the development of such indicator-based frameworks is facing important obstacles, reflecting existing limitations to effective and seamless data access and the sharing of official statistics (IAG (2017)) – for instance, when trying to make use of information collected from supervisory reports. One way to go is to promote the exchange of experience among institutions and countries in addressing these challenges in an effective and practical way.

Looking forward, well-defined data governance frameworks can be instrumental in supporting official statisticians' task to collect and analyse data of the highest quality possible. However, an **institution-level approach to data governance should be complemented by a broader focus covering the entire production and use of statistics, including alternative sources**. For instance, ensuring the following of adequate Codes of Principles by private data providers, clarifying the responsibilities in the national governance landscape, and establishing proper international guidelines and cooperation mechanisms.

Perhaps more importantly, while **NSS organisations are facing a decline in their traditional function of "data collectors", they have a key role to play as reference custodians of the quality of the data used by society**. Needless to say, establishing sound data governance frameworks can be a central element in supporting this "data curator approach".

> "It is essential to consolidate and make sure that agreed figures only are used. The utmost confusion is caused when people argue on different statistical data."
>
> Winston Churchill[4]

1. Introduction

Several international initiatives have underlined the **role of economic and financial statistics in supporting evidence-based policy making** in the recent past. This was particularly the case with the statistical response to the Great Financial Crisis (GFC) of 2007-09 in the context of the Data Gap Initiative endorsed by the G20 to enhance economic and financial stability (FSB and IMF (2009, 2015)). The 2030 UN Agenda for sustainable development (UN (2015))[5] has also recognised the crucial role that statistics can play to monitor, evaluate and track progresses on economic development. As regards Africa more specifically, the implementation of the development Agenda 2063 has been set up as "a results-based approach with concrete targets that are measurable and can be tracked and monitored" (African Union Commission (2015)). More recently, the Covid-19 pandemic underscored the need for good and reliable statistics on a wider range of topics that are not properly covered by the "traditional" statistical apparatus, especially on environmental topics (eg climate change) and socioeconomic factors (eg inequalities).

A holistic approach to data governance

The renewed interest in the interaction between economic policy and factual evidence highlights the sheer **importance of establishing proper governance when dealing with, and using, official statistics**. Yet this concept of data governance can be unclear, reflecting the proliferation of approaches developed over time across institutions and countries. In some places, data governance may be considered in a narrow way, by focussing essentially on data protection and its associated legislation. Yet the focus could be broader, covering various intertwined topics from data management to data use.

A **holistic approach to data governance** is to consider "everything designed to inform the extent of confidence in data management, data use and the technologies derived from it" (British Academy and The Royal Society (2017)). The frameworks put in place to govern corporate data ecosystems will typically include three layers: strategic, tactical, and deployment (delivery); cf Table 1 in the case of the public sector.

For **NSS institutions, data governance will in practice refer to the various organisational features put in place for dealing with official statistics and cover all the related principles, policies and procedures, structures, roles and**

---

[4] *Note to the secretary of the War Cabinet*, 8 November 1940, as quoted by Prévost and Beaud (2012).

[5] Cf The SDGS in action website.

responsibilities.[6] The objective is to ensure the "quality" of this information considered as "a strategic institutional asset" (Diokno (2019)).

Table 1: Data governance frameworks in the public sector – the OECD three-layer approach

Layer	Main components	Core elements
Strategic	Leadership - vision	→ Data strategy & policies → Leadership roles → Policy levers
Tactical (coherent implementation)	Capacity for coherent implementation	→ Data committees & communities, data stewards → Skills & competences, training & funding → Data innovation, value extraction
	Regulation	→ Data-related rules & guidelines (data openness, publication, protection & sharing)
Delivery (day-to-day deployment)	Data architecture	→ Standards, reference data, interoperability, relationships
	Data infrastructure	→ Data registers, catalogues, lakes, IT solutions
	Data value cycle (from production to openness & reuse)	→ Actors, roles and technical skills (eg data validation, sharing, integration, ownership, integrity)

Source: OECD (2019).

It is certainly the case that there is **no single definition of "data quality", the safeguarding of which is the main objective pursued by data governance frameworks** (cf Box 1). In practice, one can define quality as referring to the various elements that support the Fundamental Principles of Official Statistics. "Quality" will therefore be understood as a generic term covering the various characteristics sought for official statistics, ie their accuracy and trustworthiness, integrity and security, and that they are documented and easy to find/access. Yet the approach will also cover the user side, for instance to ensure that the data are fit-for-purpose, that their value is maximised, and that they can be traced, reused, and eventually adequately deleted. Perhaps more importantly, it will also include more ethical aspects, such as preserving public trust in official statistics and ensuring that the information collected is not misused and does not undermine personal privacy, confidentiality or democratic principles.

---

[6] Cf for instance in the EU/Eurostat case: "**data governance** entails defining, implementing and monitoring strategies, policies and shared decision-making over the management and use of data assets"; "**data policies** are a set of broad, high level principles which form the guiding framework in which data assets (...) can be managed. More specifically, data policies govern data management, data interoperability and standards, data quality, data protection and information security"; (European Commission (2020b)).

Progress on data governance appears necessary so that NSSs can reap the full benefits of the ongoing "data revolution". The reason is that to make use of "non-traditional" types of data sources, statisticians need to revisit and preserve trust in the four main steps involved in the information chain, ie the collection, management, dissemination and use of data.

Addressing the four main steps in the information chain

As regards **first data collection, "traditional" statistical surveys and censuses can increasingly be complemented with new information sources**, especially big data and administrative records – the (micro) data revolution (Bean (2016)). This can be a great opportunity for those less developed statistical systems, such as in many African states, not least considering the high costs associated with setting up standard data collections.

Yet there are a **number of governance challenges posed by accessing these new data**, for instance in terms of quality and accuracy, privacy, ownership, and reputation. In particular, one issue of key relevance for public statisticians is how alternative information sources (being private commercial data sets or public registers

that were not initially set up for a statistical purpose) and the data producers located outside of the NSS feature vis-à-vis the Fundamental Principles that have been defined for the perimeter of Official Statistics (UN (2013)). For example, the production of the statistics in a professionally independent way, based on scientific methods and rigorous quality criteria. Indeed, a major point is that, by using alternative data, authorities would be perceived as endorsing the methodologies applied by third parties; hence, any concerns about the quality of these data could damage the credibility of official numbers, as well as of data-driven public policies. This also raises the question of whether and how national statistical legislation and quality assurance frameworks should be revised accordingly (UN (2019)).

Turning to the **second area of data management, the combination of new and more heterogeneous data types and evolving analytical needs requires more, and sometimes different types of resources** (in terms of IT equipment, staff skill mix, budget for acquiring new databases, etc; cf IFC (2020)). This may prove particularly difficult in the least advanced countries, where an important issue is how to build and keep sufficient technical expertise in the NSS. In the light of frequent and acute budget constraints, capacity building can be ensured both by joining forces between relevant national bodies and drawing on all international support available. Another important topic for official statistics is related to the implications of globalisation: national compilers have to access information sources that are not locally available, and this can be difficult given budget constraints, technical considerations, and/or data sharing limitations.[7]

Third, disseminating data also poses clear challenges. On the one hand, new digitalisation techniques allow for easier, almost cost-free access to information, which can be a great opportunity for populations with limited resources, especially in developing countries. On the other hand, the increasing complexity of economic and financial activities puts a premium on statistical education and financial literacy. National financial inclusion policies have been particularly sensitive to this issue, given that measures to facilitate households' and small firms' access to credit and payment services can only be effective if economic agents have a good understanding of the related financial aspects including data (IFC (2018a)). Again, this topic is clearly relevant for a region like Africa, where financial inclusion has been progressing only slowly.

Fourth, there has been **increasing interest globally in using data better for policy purposes, especially when designing, calibrating, assessing and modifying public actions** (Tissot (2017)). Their experience has underlined the importance of having highly granular and flexible data sets. This can facilitate the capturing of distinct, specific dimensions in the data that are relevant for analysing policy actions – in turn supporting ex-ante impact assessment, understanding feedback effects and unintended consequences, and identifying areas of improvement (IFC (2021b)). Yet what is still unclear is how such indicator-based frameworks should be developed so that adequate lessons can be drawn from the actions of public authorities. Moreover, the implications for official statisticians are yet to be fully incorporated, especially as regards the national rules governing data access and data sharing possibilities and associated challenges in terms of data

---

[7]   Data on derivatives trades reported to trade repositories (TRs) constitute a case in point (IFC (2018b)). Transactions can be reported to TRs in different countries and no single jurisdiction alone can have a complete global overview, making it difficult to assess group-level positions on a consolidated basis.

protection, privacy and confidentiality (Norwegian National Research Ethics Committees (2019)).

The contribution of the central bank statistical community

The above considerations clearly **underscore the importance of revisiting data governance frameworks for organisations in charge of official statistics – a matter particularly high on the agenda of central banks, given their distinctive dual role of being both producers and users of official statistics** (Lane (2021)). A relevant initiative from this perspective was the joint organisation by the IFC and the ISI of a High-Level Meeting on Data Governance in Tunisia on 22 November 2019. This meeting, hosted by the National Institute of Statistics of Tunisia and benefiting from the support of the African Union (AU) Commission, targeted the heads of the statistical function in the NSSs. The aim was to provide a platform for facilitating cooperation between the NSOs and statistics divisions in the central banks in jointly analysing data governance issues and with a specific focus on Africa. Participants from more than 60 organisations, including central banks, NSOs, international organisations, the private sector and academia, convened to discuss innovative strategies that would help improve official statistics in a structural way.

This event proved to be another **important milestone in the IFC's ongoing work on data management – including rapid improvements in technology (the "big data revolution") and issues related to data dissemination, sharing and standards** (eg IFC (2017,20,21a)). Moreover, it provided another occasion to share experiences and highlight existing best practices and potential opportunities, especially to support data-driven policymaking. It was also an opportunity to take stock of the challenges to be addressed as a priority, with a focus on the African perspective. In order to make concrete progress, participants concentrated on the four main topics at the core of data governance frameworks:

(i)   data collection, with the need to complement "traditional" statistical surveys and censuses with "new" data sources (Section 2);

(ii)   data management, as the combination of new data types and evolving analytical needs requires more, and sometimes different types of processes and resources (Section 3);

(iii)   data dissemination, with a focus on the public usability of the data produced and the associated ethical aspects (Section 4); and

(iv)   the use of data for policy purposes, especially when designing, calibrating, assessing and modifying policy actions (Section 5).

These governance issues also highlight the need to reflect on the evolving role of the NSS looking forward (Section 6).

2. Collecting data and dealing with new information sources

The impact of the data revolution

Big data is playing an increasingly important role in official statistics and this was particularly evident when the Covid-19 pandemic struck in 2020. Data producers

were confronted with sudden and acute data disruptions and had to develop new methodologies to address compilation challenges (De Beer and Tissot (2020)). This "wake-up call" reinforced the ongoing push for using alternative indicators to add up to the "traditional" offering of the NSS (Biancotti et al (2021)). Indeed, the opportunities provided by connecting different statistical sources are not new. In Uganda, the accuracy of poverty estimates (primarily derived from the NSO's household surveys) has been enhanced by linking to other types of data derived from an international programme measuring health conditions.[8] Turning to the United Kingdom, the population census has gradually evolved in recent years to make a greater use of administrative data sets. The aim has been to produce results that are more up to date, more continuous and granular, and more relevant to address the needs of users (and potential users). These needs are becoming more complex, more detailed, and less stable over time.

This trend is clearly gaining momentum. In Thailand, for instance, the central bank has played an active role in trying to reap the benefits of the "data revolution" by setting up an adequate data acquisition process. This has been done by:

(i)   seeking all the new data sources available in the country (eg internal, regulatory, administrative, collaborative and "next generation" sources);

(ii)   developing new types of data science and analytical methods, not least to avoid a situation where the institution is "drowning in data, but starving for knowledge"; and

(iii)   designing appropriate data management rules and processes, especially to deal with data quality, security and privacy, as well as auditing and monitoring tasks.

Similarly, in Tunisia, the **central bank has launched a fundamental upgrade of its statistical system so as to incorporate data and technology innovation**. The goal was to integrate a variety of internal and external sources (based on clear data exchange agreements), manage them according to high-level principles (eg confidentiality, transparency), and produce reliable data in compliance with international standards (eg the IMF Special Data Dissemination Standard (SDDS)). The concrete implementation relied on a business intelligence (BI)-based statistical information system. This system had a unique data and metadata repository, automatic collection and validation processes ensuring data quality, and the offering of an "hybrid" BI architecture for data management tools (with both a centralised data warehouse for "standard" production processes and data lab facilities for more complex analytics). It also had easy data access (by internal senior management as well as external stakeholders).

Substituting or complementing official statistics?

One question is **whether new data sources being contemplated will simply replace traditional statistical exercises**, eg those surveys or censuses that are complicated and/or time consuming and that can be particularly difficult to organise in poor economies and/or remote areas. For instance, smartphone location data can

---

[8]   Namely the demographic and health surveys (compiled in the context of the international programme implemented by ICF International for developing countries), which do not collect data on household income or expenditure but do record several household characteristics that are likely to be related to standards of living (Development Initiatives (2020)).

now help to monitor geographic commuting patterns or consumer habits; in the past, getting such information required detailed questionnaires, filled in by only a limited sample of households and at infrequent points in time. Experience in Rwanda also shows that existing records such as satellite images and price scanner data can replace expensive surveys for compiling, respectively, agricultural and price statistics.

However, **it would be a step too far to declare that traditional statistical exercises have become irrelevant**, as they are likely to continue to be in demand despite the development of the "internet of things". First, there are still lingering uncertainties as regards the robustness and accuracy of alternative data, their associated underlying methodologies, and data quality assurance processes to be set up (eg to control for potential bias or population under coverage; Tissot (2019)). In view of this uncertainty, official surveys/censuses represent irreplaceable benchmarks against which to judge whether new big data sources are exact ("precision"), correct ("accuracy") and representative of the situation analysed ("veracity"). Such elements should be carefully assessed before adopting any alternative indicators to compile official statistics – cf the difficulties faced by Google in setting up advanced indicators of the spread of influenza a few years ago.[9]

Moreover, **one can be concerned about the future availability of sources that would be selected as an alternative to official statistical exercises**. The risk is that the statistical infrastructure would depend on public or private organisations whose primary objective is not data collection (House of Commons Science and Technology Committee (2012)). This could lead to sudden compilation changes and breaks in series; it could also constrain the ability to study new or unexpected phenomena as times goes on.

Furthermore, **sticking to long-established and internationally-agreed practices and standards can be essential in maintaining well-founded trust in official statistics**; for instance, as regards private data protection and professional secrecy and independence. In contrast, external sources can be characterised by insufficient documentation and transparency, potential conflicts of interest (at least in the perception of the general public) and the absence of ethical standards that would provide a degree of confidence comparable to the one provided by the Fundamental Principles of Official Statistics (Box 2).

The first implication of the above is **the need to preserve sufficient "traditional" statistical capacity in the NSS** – including sampling methodology, ability to set up questionnaires, training for staff like census/survey enumerators. This can be an important issue in developing countries, where NSSs with high workload may have insufficient time and resources to explore alternative sources without compromising the core of their mandate, ie the compilation of "conventional" official statistics.

---

[9]   The Google Flu Trends indicator initially intended to provide estimates of influenza activity based on Google Search queries. It was discontinued in the mid-2010s, reflecting the difficulty of identifying stable patterns in the data (Lazer et al (2014)).

Box 2: From the Fundamental Principles of Official Statistics to private sector quality assurance processes

Fundamental principles govern the production of appropriate and reliable official statistics that have to adhere to certain professional and scientific standards. A number of private firms involved in the production of alternative sources have also adopted dedicated and transparent mission statements and principles pursuing similar objectives. Yet these various initiatives are still fragmented and the official statistical community may play a role in fostering their development in a more comprehensive way. **The main goal should be to set up a framework allowing the usability of (non-official) private data sources for national statisticians in supporting public policy**. This Box summarises key ethical considerations related to the statistical profession, documents some recent private initiatives and lists a number of key issues when dealing with alternative data sources.

1. Key elements of the Fundamental Principles of Official Statistics:

The 10 Principles of Official Statistics adopted in 1994[10] were specifically targeted to support the role of high-quality official statistics in analysis and policy decisions. They stress the **essential importance of public confidence and call for effective fundamental values to govern statistical work (UN (2013)). The following are the key objectives** (summarised excerpts):

1. Being the indispensable element of the information system of a democratic society, by compiling and making available statistics of practical utility.
2. Retaining trust in official statistics / following strict professional considerations.
3. Facilitating a correct interpretation of the data.
4. Commenting on erroneous interpretation and misuse of statistics.
5. Drawing data on all types of sources.
6. Preserving strict confidentiality of data about individuals and ensuring data is used exclusively for statistical purposes.
7. Making public the laws and rules under which the NSS operates.
8. Ensuring coordination among statistical agencies.
9. Using international concepts / classifications / methods.
10. Promoting international statistical cooperation.

A number of regional initiatives have been developed to promote adherence to these principles, for instance the European Statistical System (ESS) adopts 15 principles set out in the European statistics code of practice to cover three main statistical areas (ie institutional environment, statistical processes, and statistical output) and complemented by an official Quality Declaration (Eurostat (2017)).

2. Professional values shared by statisticians

In 2010 the ISI adopted a declaration on professional ethics (ISI (2010)). Its principles are comparable to the UN principles but with the **aim of covering the wide range of the statistical profession working within a variety of economic, cultural, legal and political settings**. It notes that the definition of who is a statistician goes well beyond those with formal degrees in the field to include a wide array of creators and users of statistical data and tools.

The declaration consists of a statement of three shared **professional values (ie respect, professionalism, and truthfulness and integrity) and the following set of 12 ethical principles** derived from these values:

- Pursuing objectivity
- Clarifying obligations and roles
- Assessing alternatives impartially

- Conflicting interests (to be avoided)
- Avoiding pre-empted outcomes
- Guarding privileged information
- Exhibiting professional competence
- Maintaining confidence in statistics
- Exposing and reviewing methods and findings
- Communicating ethical principles
- Bearing responsibility for the integrity of the discipline
- Protecting the interests of subjects

3. Examples of recent private sector initiatives

Commercial data providers have been also working **on providing transparent information on their data offerings**, following the example of Google that has developed public "FAQ about Google Trends data". Most of the principles put forward by the industry focus on particular aspects of data governance. For instance, public communication by Apple – which produces mobility trends derived from smartphone location data[11] – emphasises that privacy is a fundamental human right and one of the core values of the firm. Yet, while such disclosure practices are welcome, private data producers do not generally refer to a set of universally accepted principles governing data management and that would be comparable to the established ones for official statistics. Indeed, there have been a number of calls among statisticians for this gap to be filled.[12]

Certainly, a number of private data providers have already taken steps to **establish "principles" that govern their activities with the objective of ensuring trust in the data they provide**.[13] For instance, and as regards data privacy specifically, four key privacy principles have been set up by Apple (Apple (2021)):

- data minimisation: collecting only the minimum amount of data required;
- user transparency and control: making sure that users know what data is shared and how it is used, and that they can exercise control over it;
- on-device processing: processing data on the device, wherever possible, rather than sending it to the firm's servers, to protect user privacy and minimise data collection;
- security: hardware and software working together to keep data secure.

Another example is the US online real estate company Zillow, whose data are now being used by the Board of Governors of the US Federal Reserve System for official statistical purposes (see Hume McIntosh (2021)). Zillow's Research Mission Statement and related Principles states that its research "aims to be the most open, authoritative source for timely

---

[10] These Principles were initially developed by the Conference of European Statisticians in 1991, when countries in Central Europe began to change from centrally planned economies to market-oriented democracies, to ensure that NSSs in such countries would be able to produce appropriate and reliable data that adhered to certain professional and scientific standards. As statisticians in other parts of the world soon realised that the principles were of much wider, global significance, the very same set of principles were adopted in 1994 by the UN Statistical Commission as the UN Fundamental Principles of Official Statistics. The principles were then formally endorsed by the UN General Assembly in 2013. They have been also complemented by comprehensive guidance on implementing quality assurance frameworks (UN (2019)).

[11] See covid19.apple.com/mobility for information on the Apple Mobility Trends Reports; and www.apple.com/uk/privacy/ for the firm's commitments on privacy.

[12] See the call by Nuti et al (2014) for "a discussion and collaboration between Google Inc. and the research community [...] to create a set of best practices to ensure that the [Google Trends] tool is being used responsibly".

[13] As regards financial institutions, the Basel Committee on Banking Supervision has developed principles to ensure that a strong governance framework, risk data architecture and IT infrastructure are in place (BCBS (2013)).

and accurate housing data and unbiased insight. Our goal is to empower consumers, industry professionals, policymakers and researchers to better understand the housing market". The related research principles comprise:

- the provision of unbiased data and analysis in a transparent way including, whenever possible, by making code and underlying data available so work can be replicated independently and clearly explaining methodology;
- being independent from the firm's business goals, and not being a revenue centre;
- being open about how data sources and data quality impact research, and transparent about potential issues with data used;
- benchmarking findings against outside data sets whenever possible to ensure accuracy and appropriate context; and
- respecting the integrity of data and using it honestly, and never manipulating data to create a desired result; being data- and fact-driven.

4. Towards a set of self-commitments for providers of alternative data?

One way to go would be for commercial data providers to simply state publicly that they adhere – for the relevant part of their activities – to already existing principles such as the fundamental ones governing official statistics or the ISI Principles, and to document how this is being organised in practice. Another solution would be **to adopt a set of recommendations that would be specific for the private data industry**. For instance, a key objective should be that the data involved are unbiased and produced with high professional standards and in a transparent way (eg with adequate metadata information). Another would be to ensure that the provision of this kind of information is made independent of the data providers' business goals. A third aspect is to favour benchmarking exercises, especially against other (non-alternative) data, to ensure accuracy and appropriate context. A final point would be to certify that the integrity and confidentiality of the data are respected and that the information is used honestly.

NSSs and international organisations could play a role in developing such a "principle-based" approach and support good data governance. A key point is to build on the comparative advantage of high-quality official statistics when the necessary proof of quality may be lacking for other data sources (Eurostat (2016)). Policy makers have indeed already underlined the importance of this topic to address the challenges associated with the Covid-19 pandemic. In particular, the G20 Finance Ministers and Central Bank Governors have stressed that, "harnessing the wealth of data produced by digitalisation, while ensuring compliance with legal frameworks on data protection and privacy, will be critical to better inform our decisions". They invited the main international financial organisations to reflect on a possible new Data Gaps Initiative that could address these issues (G20 Italian Presidency (2021b)). Such public/private cooperation could be guided by a set of high-level principles covering three main groups, ie all the parties involved, NSS members, and private data providers (cf Table 2).

In conclusion, by highlighting the urgent need for new statistics to support a move towards sustainable development, the Covid-19 crisis has shown that official statistics could play a new, essential role to enhance the public statistical infrastructure (SJIAOS (2020)). **Key to this is creating new forms of cooperation with other data players located outside the NSS**, possibly by: (i) setting up an adequate "Code of Practice"; (ii) providing appropriate labels/quality certificates based on the know-how/reputation of NSS bodies; (iii) and/or providing data management support services (eg IT tools and methodology advice).[14]

---

[14] In a way that could be similar to the role played by the official statistical international community in providing a range of free, open source and commercial software tools to support the Statistical Data and Metadata Exchange (SDMX) standard, as listed on the SDMX website.

Table 2: High-level principles for public private partnerships supporting better official statistics

For all the parties involved	• Purpose and time limitation • Proportionality • Controlled flexibility • Data anonymisation / privacy-preserving techniques • Application of professional standards • Transparency • Fair / ethical data use • Risk mitigation / safeguards • Timeliness • Accountability
For official statistical agencies	• Overall responsibility / accountability on process & output • Overall responsibility on methodological standard-setting • Safeguard individual & business interests • Collect only once
For private data providers	• Data & metadata availability • Technical accessibility • Cooperation on formats and standards • Mitigation of limitations • International comparability

Source: G20 Italian Presidency (2021a).

A second consequence is to **favour a complementary use of both traditional and alternative data sources**. The Covid-19 pandemic underlined the importance of the financial big data chest built since the GFC (eg large micro databases) as well as of the "internet of things" to support policies at a time of high uncertainty (Lane (2021)). Indeed, alternative data can be exploited to improve the modelling of existing surveys, reduce their costs/frequency, provide more up-to-date estimates, or act as statistical buffer in times of disruption / high uncertainty (INSEE (2020)).[15]

Coordination between data providers

Achieving such complementarity **puts a premium on fostering coordination between members of the NSS and non-traditional data providers**, especially to clarify the sources involved and the underlying methodologies. In Ghana, for instance, this helped to provide confidence in the robustness of private telephony records, in turn enhancing the official estimates of quarterly GDP.

In practice, good **coordination often requires a formalised framework undersigned by the respective parties**. A key point is to have clear agreements/partnerships to ensure that the NSS can leverage big data by accessing all the best sources possible in a transparent way. This must be done while ensuring confidentiality protection, safeguarding the legitimate interests of the data owners

---

[15] For concrete applications, see the specific page set up by the German Federal Statistical Office (Destatis) to display "Corona statistics", including experimental indicators such as the "truck toll mileage index".

and avoiding the perception that the data may have been altered because of organisational deficiencies and/or business considerations.

Such a framework would be **essential to reassure official statisticians willing to control the quality of the new data sets incorporated**.[16] It is particularly important for those countries where the NSS is not entitled by law to have access to other national data providers, or when the legal arsenal at their disposal is not well adapted to the peculiarities posed by new big data sources (especially when this information is not stored within domestic borders). For instance, "community-based data-sharing agreements" have been developed as part of the OECD project on enhancing safe and ethical access to data, with "data trusts" being set up as legal structures to provide independent stewardship of data (OECD (2019)). Further, such partnerships can cover additional aspects including the compilation of "new" indicators, the technologies involved and dissemination channels (UNECE (2018)). The experience of Kenya is that this approach can successfully help to exploit untapped data sources and integrate them in the production of official statistics, in turn reducing their collection burden and enhancing their accuracy.

Yet the experience of Rwanda shows that even when the NSI is empowered with a sound legal basis to access private databases, **practical considerations also matter when ensuring effectiveness**. Formal partnerships may not be sufficient, and good coordination at the working level is required to ensure that data produced outside the NSS are correctly integrated into the national statistical framework. The reason is that official statisticians need to have a good understanding of the new and evolving data universe, access relevant documentation ("metadata") and conduct research work before going to the production stage of new indicators. This puts a premium on having a constant dialogue with external data producers, which can be challenging in practice. Such a dialogue is reported to be easier with public institutions compiling administrative data sets compared to private big data firms, suggesting that government staff may be better placed to understand the challenges faced by official statisticians and their specific requirements.

3. Managing data: facing new capacity requirements

Data governance as a way to structure data work

The official statistical system is facing an avalanche of data that can be difficult to manage properly. On the supply side, rapid IT innovation means that databases are becoming larger, easier to access (eg "open data") and also more numerous and heterogeneous – cf the expansion of non-structured data. This calls for more capacities to store, preserve, link, transform and analyse the data. It also requires setting up adequate and standardised procedures for combining all the information together (eg taxonomy-based data dictionaries, catalogues, registries, transformation

---

[16]   See the example of Statistics Netherlands which launched the Center for Big Data Statistics in 2016 to develop (in close collaboration with knowledge institutions and market players) experimental statistics or "beta products"; important attention is put on examining the stability of the data sources, validating methods and assessing operational requirements.

rules and cross-sector interoperability standards).[17] On the demand side, the potential usefulness (or "value") of the new information has gone up especially for policymaking. For instance, real time statistics can help understand complex changes to the economy faster and improve how policy decisions can be made. Moreover, data is an asset that must be adequately safeguarded and controlled, to ensure that it remains safe, reliable and appropriately used, and also to protect confidentiality, privacy and intellectual property (eg prevention of data leaks and compliance with legal frameworks).[18]

These evolutions are clearly **putting pressure on NSSs' IT and budget resources as well as on staff skills**. In the case of Senegal, for instance, the implication of the information revolution is not just that more indicators have to be produced by the NSI: for the data to be useful, they need to be transformed and exploited. Moreover, there are specific issues posed by the integration of alternative sources. These include issues related to data documentation (or "metadata", to have reassurance on the quality and fitness for purpose of these data), calibration (to ensure that the new indicators compiled provide an accurate view of the underlying reality) and curation (to ensure the adequate organisation and integration of data collected from various sources, including the ones collected by other data communities). In addition, the processing and analytical tools at disposal are subject to rapid innovation, while new data usages are constantly emerging. Furthermore, official statistical bodies have to deal with the specific responsibilities entailed in accessing new, potentially sensitive data, especially in terms of transparency and public accountability.

A data governance framework can alleviate these challenges as it can help to better understand these various issues and address them effectively. In Portugal, for instance, the central bank has launched an integrated data management programme to make better use of the information available in the whole organisation and rationalise the various processes involved in its collection. More concrete objectives are to combine multiple data sets, promote data sharing, allow for both regular data exploration and (irregular) experimental analysis, and maximise the usefulness of the data for internal and non-internal (research community) users. The implementation of this programme included two main aspects:

- the definition of a governance model with clear differentiation of responsibilities between the different types of staff involved in data management; and

- the development of a logical data architecture encompassing the data (comprising all potential sources and "core" master data), storage facilities (with a data lake for all raw data and a corporate data warehouse repository), and data exploration tools (with both data lab facilities based on data science techniques and BI-based tools supporting corporate production tasks).

---

[17] Cf Witt and Blaschke (2019) for the importance of these elements supporting the Data Intelligence Service Centre (DISC) platform set up at the ECB.

[18] See for instance the ECB communication on commitment to data protection, which is a fundamental right enshrined in the European Union (EU) by various legal texts. These include the Charter of Fundamental Rights, the Treaty on the Functioning of the EU and the data protection rules set out in the General Data Protection Regulation (GDPR). These apply to any organisation established in the EU, as well as to organisations based outside the EU that intentionally offer goods or services to the EU, or that monitor the behaviour of individuals within the EU.

Implementation

A **key feature of a sustainable data governance framework should be comprehensiveness in covering the entire institution** so that it can become more data-driven. The objective is to make the best use of the data available, enable their access, sharing and integration, and increase overall efficiency and accountability (OECD (2019)). In the BIS case, the developed framework includes a wide range of data policies, standards and guidelines that can support extracting value from all of the data assets in the institution. Its set up comprised the design of a general data processing architecture (with the structural role played by the Statistical Data and Metadata Exchange (SDMX) information model (IFC (2016)), the clarification of the business cases across various units, the development of appropriate IT solutions, and the integrated implementation of related software and hardware components. Comprehensiveness was ensured by defining clear data governance principles, setting up data stewards mandates, and clarifying the main technology underpinnings (cf Box 3).

Because of their broad scope, **data governance frameworks are often considered as an essential part of the organisations' strategic plans**. In the case of the BIS, for instance, it is a key building block supporting its overall Innovation 2025 strategy, ensuring its robustness and comprehensiveness. It can also speed up its secure implementation by acting as a potential "car break", allowing the pursuit of multiple ambitious objectives in parallel – eg improved user services, novel data-driven capabilities, (operational and cyber) resilience and business agility.

The **European Commission** has also developed a fully fledged "European strategy for data", based on four pillars:

- a cross-sectoral governance framework for data access and use, based on a set of high-level guiding principles;

- enablers (investments in data and strengthening Europe's capabilities and infrastructures for hosting, processing and using data, interoperability);

- competences (empowering individuals, investing in skills and in SMEs); and

- common European data spaces in strategic sectors and domains of public interest (see European Commission (2020a).

Turning to the **IMF**, its overarching strategy on data and statistics focusses on three "I"s: Integration, to connect the various data provision work streams; Innovation, to take advantage of big data; and Intelligence, to leverage artificial intelligence (AI) to analyse data and statistics (IMF (2018)). Moreover, two governing bodies have been created: the Standing Committee on Data and Statistics to shepherd the implementation of the data strategy; and the Data Governance Group develops and monitors the implementation of good management policies, practices and guidelines.

The data governance framework supporting the architecture developed for BIS statistical processes in the context of the Bank's Innovation 2025 programme relies on three key components: high level data governance principles; selected key IT underpinnings; and dedicated staff resources (eg data stewards).

1. High-level data governance principles

- Data is an asset
- Data has an owner
- Data that has shared value should be shared
- Data is accessible
- Data quality is actively managed
- Data is described with a common vocabulary and data dictionaries
- Data security is actively managed
- Interoperability: software and hardware should conform to defined standards that promote interoperability for data, applications, and technology (to ensure consistency and improve the ability to manage and design systems, the protection of existing IT investments and user satisfaction).

2. Key IT underpinnings

- Interactive Data Exchange and Analytics (IDEA): consolidated platform architecture supporting the statistical data lake / factory / lab
- SDMX-based information model, with two key aspects:
  (i) Content-Oriented Guidelines (COG) as recommended practices for creating interoperable data and metadata sets using the SDMX technical standards in all statistical domains.
  (ii) A focus on harmonising specific concepts and terminology that are common to a large number of statistical domains; such harmonisation supports a more efficient exchange of comparable data and metadata.
- Data catalogue
- BI tools (dashboards)
- BIS data portal and dissemination toolsets

3. Leading role of dedicated data stewards in:

- Promoting data governance principles
- Developing / maintaining data catalogue
- Managing metadata
- Promoting awareness of data sets
- Selecting and implementing IT tools

The various experiences above show that success will often depend on a few "best practices", including high-level commitment in the organisation, a transparent inventory of all data assets,[19] the design of controls to ensure that agreed procedures are followed and staff resources specialised in data management tasks. In

---

[19] With the structural role of establishing **proper data catalogues** to support this inventory; cf Zaidi and De Simoni (2019).

particular, there is a need to differentiate clearly between responsibilities, for instance with the following split:[20]

- system owners: in charge of using specific information systems;

- data owners: responsible for information sets in the data warehouse;

- data custodians: responsible for the safe custody, transport, storage of the data;

- data stewards: their role is to ensure the quality of the data assets stored, in line with the governance processes established by the organisation; as "day-to-day" contact points between business areas and the information management programme, they often play a key role in the success of data governance frameworks;

- data experts: specialised in specific business contents; and

- data users (cf Sections 4 & 5).

It is also important that the **various stakeholders are closely associated, both within the organisation and outside** (other data producers, business communities, cybersecurity experts, etc). Yet the major change is often cultural, by promoting organisational changes and developing effective partnerships between business and technology areas. In other words, data governance is not exclusively a technical issue confined to IT departments, as it applies to all of the organisational structures of an institution. This calls in particular for establishing strong data coordination groups.

The benefits of cooperation

In practice, the establishment of a fully-fledged data governance framework may be hindered by the important resources to be mobilised – an issue of clear importance in view of the budget constraints faced by many NSSs. **One way to go is cooperation**. Setting up formal associations of statisticians as well as more informal forums can be a good first step to share experiences and spread innovative practices – in turn avoiding the so called "NIH (not invented here) syndrome" when different institutions seek to address problems in parallel. This can be particularly useful in developing regions like Africa, where statistical capacity is limited: (big) data skills can be scarce, and local talents may be attracted by more lucrative positions in other countries. Addressing these challenges requires considering a wide range of issues, ie not just training for internal staff, but also the available infrastructure in terms of budget, IT equipment and project management tools, as well as development opportunities (both from an academic and managerial perspectives).[21]

Cooperation can take place at various levels to support capacity building, as documented by the AU Commission. First and foremost, it can be organised within

---

[20] This list is just indicative and the exact repartition of these interdependent roles will in practice depend on the specific organisation of the corporate data governance framework; see for instance the leading role played by data owners, stewards and users in the case of the European Commission / Eurostat (European Commission (2020b)). In addition, organisations may wish to establish a single function of chief data officer responsible for enterprise-wide governance and the utilisation of information as an asset, and/or dedicated governing bodies eg data sharing / access committees.

[21] Cf UNECE's work on the various aspects related to human resources, organisational frameworks and cultural change at the very core of the modernisation of official statistics.

the NSS, with a leading role to be played by the NSO and the central bank as key compilers of official statistics; this can be facilitated by establishing proper legal frameworks, effective statistical planning and strong coordination mechanisms. Second, cross-country cooperation can help to publicise "best practices" or "success stories". For instance, the NSI of Rwanda has decided to set up a Data Science Campus to investigate the use of alternative sources and promote new generation data technologies, leveraging the experience of the UK Office of National Statistics (ONS).[22] Third, NSS capacity challenges can be addressed though the support of international organisations. As an example, the Partnership in Statistics for Development in the 21st Century, or PARIS21, has developed a comprehensive approach to implement and monitor capacity development to strengthen national statistical governance. Statistical planning has already significantly improved with the development of National Strategies for the Development of Statistics, though limited funding remains a challenge.

4. Disseminating data: How to make everyone benefit from the data revolution

Official statistics as reference information

A major feature of official statistics relates to their contribution to the public good, underlining the importance of adequately disseminating them to the general public. Indeed, one can argue that "communicating evidence effectively is as important as the quantity and quality of evidence itself "(CSQ (2019).

A first aspect is that official statistics should provide a reference and an objective basis for answering questions posed by society. This means presenting evidence-based facts effectively so that they can be understood, accepted and used without question, as argued by the former Head of the UK ONS John Pullinger in UNECE (2018). Data compilers should not only communicate in a clear and understandable way but also present a comprehensive picture of the analytics provided, including the degree of uncertainty associated with the data and techniques used. It also puts a premium on developing statistical literacy in the population, and on ensuring that the public understands what official statisticians are doing and is comfortable with that. These issues have clearly gained prominence with the emergence of big data and "black box evidence", reflecting the difficulties of communicating on the basis of complex and automatised analytical tools relying on AI (Wibisono et al (2019)). Lastly, it may require NSS bodies to challenge misrepresentations or bad use of numbers put forward by private producers in the public debate.

There is evidence that the role of the NSS in **providing reference information can be instrumental in supporting economic development**. Indeed, the conduct of IMF multilateral surveillance exercises requires cross-country comparable data, and past financial crises have underlined the importance of providing better and transparent information to markets (Fischer (2002)). One example is the research

---

[22] See the ONS Data Science Campus initiative that aims to develop data science projects and capacity building in partnerships with academics, industry and government departments at www.ons.gov.uk/aboutus/whatwedo/datasciencecampus.

conducted by the Deutsche Bundesbank that shows that information asymmetry problems between investors and project owners, a key obstacle to FDI expansion in poor countries, can be relieved by the dissemination of accurate and trusted data. The importance of reference statistics is also one key driving force of the push made in recent years to improving countries' statistical production in line with the IMF data dissemination standards.[23] Indeed, Choi and Hashimoto (2017) estimate that such data transparency efforts can significantly reduce the spreads of emerging market sovereign bonds.

Communicating statistics

A comprehensive governance framework can help to strengthen the reference role of official statistics. For instance, the German NSO has developed a communication strategy to ensure that the data produced are understandable and explicable as well as relevant and easily accessible. This strategy relies on key data governance elements, such as:

- strong official statistics "brand" especially in terms of quality and trust;

- improved data access: eg user friendliness, search engines, demand-driven content, and link management tools;

- focus on meeting users' needs, by moving away from a culture of users retrieving information to delivering data to targeted groups;

- dialogue with external stakeholders: users' forum, newsrooms, social media, feedback surveys; and

- improved data comprehensibility: eg visualisation, literacy, comprehensiveness.

Another successful experience is the open source approach adopted by the Statistics Department of the African Development Group as part of its capacity-building program. The increasing demand for official indicators is addressed by a unified statistical portal (the "Africa Information Highway"[24]) based on live open data platforms electronically linking all African countries. This approach has been particularly effective to support the monitoring of the SDGs in Africa. As part of its new strategy on data and statistics, the IMF is also working to establish "global data commons", ie an integrated cloud-based network of country websites publishing data essential for surveillance (IMF (2018)).

Yet, **in practice communicating statistics also requires paying concrete attention to the specific groups of users involved**: the data compiled are of little use for them if they do not meet their expectations. NSSs' bodies should thus clearly identify and understand users' information needs before deciding on new dissemination projects as well on respective priorities. In turn, the receivers of statistical information should be able to digest it properly. This calls for helping them

---

[23] See dsbb.imf.org/. The IMF dissemination standards comprise, amongst others, the Special Data Dissemination Standard (SDDS), established in 1996 to guide countries that have, or that might seek, access to international capital markets in the dissemination of economic and financial data to the public; and the enhanced General Data Dissemination System (e-GDDS), established in 2015 to guide countries in data dissemination by supporting transparency and helping to create strong synergies between data dissemination and surveillance – with a particular focus on developing countries in Africa and statistical capacity development.

[24] See dataportal.opendataforafrica.org/.

to understand, access and make use of the data through adequate visualisation and dissemination tools (eg infographics, dashboards and other BI tools; cf IFC (2019b)). To this end, purely quantitative information needs to be transformed to be knowledge accessible for non-statisticians, by providing visualisations, narratives or "stories", and short "key facts" pages (Drozdova (2017)).

Attention has focussed on specific groups of users of statistics, in particular academic researchers (and also policy makers, cf Section 5 below). To make further progress in this area, a number of central banks, NSOs and international organisations have joined forces in the context of the International Network for Exchanging Experience on Statistical Handling of Granular Data (INEXDA) (IFC (2019a)). The network's ultimate aim is to help others use granular data for analytical, policy and research purposes. Yet one difficulty in practice is to balance the utility of making information more accessible and the risks involved in terms of confidentiality (eg risk of identification or data leak). One successful example is the Research Data Center and Innovation Lab unit set up by the Bank of Italy to promote the dissemination of microdata and address these challenges. Since all the data cannot easily be anonymised, different processes have been developed depending on the type of information considered. For instance, by producing so-called Public Use Files (PUFs) consisting of micro-level records prepared in such a way that individual entities cannot be identified; this allows working on micro data sets but at the price of a loss in information value. Other solutions include the set-up of remote restricted access facilities or the provision of on-site access for selected researchers.[25]

5. Increasing the use of data for policy purposes

A specific aspect of data dissemination relates to policy makers, who have proved to be increasingly important consumers of official statistics. Data governance frameworks should facilitate the provision of good quality, trustworthy and relevant evidence, and ensure that this evidence is correctly used at the right time. This in turn leads to better decisions by supporting the design, calibration, assessment and modifications to public actions. These tasks can be facilitated by the ongoing data revolution, with the increased availability of new data sources and statistical tools, including predictive analytics such as propensity score matching techniques[26] that can support the assessment of policy effectiveness. Yet three fundamental issues arise. First, what are the key elements to focus on when supporting policy makers in making use of data? Second, what are the implications for the NSS in interacting with public authorities? An third, how can countries learn from each other, since most public policies are undertaken at the domestic level?

As regards the first issue, the starting point is to provide information that is relevant to policymakers. An important requirement is to have real-time data as much as possible. Timeliness can be enhanced through better coordination among the various stakeholders involved, such as the official data community, media hubs,

---

[25] For the examples related to Eurostat's dissemination of microdata, see ec.europa.eu/eurostat/cros/content/public-use-files-eurostat-microdata-0_en; for the various possibilities set up at Statistics Canada to access micro data, see www.statcan.gc.ca/eng/microdata.

[26] Allowing for a probabilistic comparison of two groups of similar characteristics, with only one of them subject to the policy intervention.

etc. It can also be improved by streamlining data management processes, for instance, with a greater use of alternative, high-frequency indicators as a complement to conventional statistics. Another point is to translate policy requests into information needs, eg in terms of types of data and potential sources. This means that users should not be asked which data they want, but rather what their analytical needs are. This calls for a close interaction between official statisticians and public authorities, while still preserving independence that is key for ensuring public trust.

A **second issue is how NSS organisations should manage their communications with policy-makers**. Obviously, official statisticians are best placed for selecting and transforming raw data to produce high quality information. But they should also put themselves in the position of the users and focus on "what are the data sources that can answer that question?" instead of saying "here's a survey, what can it tell you about the world?" – to quote the former UK NSO Head (CSQ (2019)). Moreover, data experts should not just produce statistics that reflect the reality, but also present these indicators in an easy and attractive way. To this end, NSSs are developing in-house capacity, for instance by setting up dedicated communication units, investing in writing analytical reports and preparing simple statistical outputs that can be easier for use. A last communication aspect is to ensure that authorities are also well equipped to understand the data at stake. Indeed, one important paradox is that despite the increased push for conducting data-driven policies, many decisions continue to be taken without due consideration of statistical basics. To this end, the AU-related regional statistical training centres have emphasised the need to define capacity programmes dedicated not only to producers of statistics, but also to users including public servants.[27]

The last issue is **how to promote the exchange of experience among countries** since communication with policy makers is mostly organised at the country level and often reflects national idiosyncrasies. One suggestion is to provide more international platforms to provide successful examples of best practices that may not be widely known. For instance, by promoting the dialogue between users and producers of statistics. This can be done by organising "wiki-type based" sharing exercises showcasing successes, setting up adequate user surveys, and outlining the public value of official statistics (UNECE (2018)). Another initiative has been to communicate about statistical "champions". For instance, the Global Partnership for Sustainable Development Data has engaged a number of such champions to support the SDG data initiative and provides a platform for them to report on progress. One of these champions is Ghana, where the NSI is working across ministries, departments and agencies to improve the way administrative data is collected and connected to other sources so that disaggregated statistics can be derived for SDG monitoring (Global Partnership for Sustainable Development Data (2019a)).

---

[27] For instance, the Eastern Africa Statistical Training Centre has been releasing a series of courses devoted to narrowing the gap with policy makers in the fields of the production and application of statistical information for public concern, covering, in particular, data management issues.

6. Looking forward: data governance and the evolving role of the National Statistical System

The international statistical community has established a number of principles to ensure that the production of official statistics relies on strong ethical considerations. In particular, one of the **key objectives enshrined in international codes for public statisticians is to collect and analyse data of the "highest quality possible"**, with due respect for privacy and confidentiality, as emphasised by Stephen Penneck, ISI President-Elect.

A well-defined data governance framework can clarify how this "data quality" can be ensured at the level of individual institutions participating in the NSS. It would, in particular, clarify what "high level principles" are followed to secure confidence in data and public trust.[28] It would also document that adequate methods are applied to preserve confidentiality, clarify data ownership and responsibilities, provide a frame to balance the benefits against the risks posed by using confidential information, and help to promote debates on these issues in the profession and publicly.

Such an institution-level approach to data governance should be complemented by a broader, NSS-wide scheme focussing on the protection and dissemination of the key features of official statistics. This is important to:

(i)   ensure a coherent approach of the various bodies participating in the NSS that have to work in close cooperation, for instance in the context of the System of the National Accounts (European Commission et al (2009)); and

(ii)   deal with the new stakeholders brought by the data revolution, especially those data providers sitting outside the NSS, since public statisticians have to gain legal access rights or negotiate agreements with them; obtain data in an identifiable and secure way; and ensure that this information is used for well-defined statistical purposes, respects personal privacy, and is based on appropriate consent.

In view of this need for a holistic framework, **there is a case for clarifying responsibilities related to the evolution of the national data governance landscape as a whole** (British Academy and The Royal Society (2017)). In the United Kingdom, a dedicated and independent advisory body (the Centre for Data Ethics and Innovation) has been set up to steward the country's data governance framework and, in particular, connect policymakers, industry, civil society and the public to develop the right governance regime for data-driven technologies. This initiative has been complemented by the set-up of a National Statistician's Data Ethics Advisory Committee[29] to provide transparent and timely ethical advice to official statisticians

---

[28]   According to the UK British Academy and The Royal Society (2017) the "overarching principle is that systems of data governance should promote human flourishing". It is complemented by four high-level principles which are to: (i) protect individual and collective rights and interests; (ii) ensure that trade-offs affected by data management and data use are made transparently, accountably and inclusively; (iii) seek out good practices and learn from success and failure; and (iv) enhance existing democratic governance.

[29]   See uksa.statisticsauthority.gov.uk/the-authority-board/committees/national-statisticians-advisory-committees-and-panels/national-statisticians-data-ethics-advisory-committee/.

more specifically. The goal is to ensure that access, use and sharing of public data for research and statistical purposes is ethical and conducted for the public good. This body also reviews the appropriateness of the various projects and policy proposals for using novel alternative data. Turning to Norway, the groups of Norwegian National Research Ethics Committees (2020) have worked on identifying ethical challenges associated with big data tools and sources, especially as regards the issues of human dignity, responsibility, transparency, dissemination, uncertainty, privacy, fair data access and data quality.

At the international level, the Global Partnership for Sustainable Development Data has established an Ethics and Integrity Framework to ensure the application of ethical considerations in carrying out the work of the global network of data communities. This network is working together to ensure that the new opportunities of the data revolution are used to achieve the SDGs (The Global Partnership for Sustainable Development Data (2019b)). In addition, a data charter has been set up that emphasises key data governance principles, such as the need to cover all the population, the promotion of data for public use, the need for data granularity and for drawing from all available sources, the accountability of data collectors, and human and technical capacity requirements.

NSS organisations and in particular NSOs and central banks are playing a driving role in these developments to reinforce the value of official statistics at national, regional and global levels. **This suggests a fundamental evolution in their mission, given that their traditional function of "data collectors" is diminishing** (at least in relative terms) with the growing ability of alternative sources. As highlighted by Irena Krizman, former ISI Vice President and Head of the NSO of Slovenia, official statisticians have to adapt to a range of new challenges. These include competencies (impact of IT innovation, understanding of new big data sources and tools), legal frameworks (to get access to data but also coordinate the compilation of statistics and ensure quality assurance including for data collected from outside the NSS), and political support (to secure public commitments and the provision of adequate resources).

NSS bodies therefore have a key role to play as reference custodians of the quality of the data used by society. In other words, they need to become "data curators", by developing the use of (external) secondary data, providing managerial, methodological and technical standards and guidance, and ensuring trust in all the stakeholders involved in statistical compilation. One view is that they should get an expanded mandate to provide assurance on the quality of statistics derived from big data sources. In the absence of internationally accepted standards, a related certification process could, for instance, be established when designing national quality assurance frameworks. Whatever the specific institutional arrangements considered, the main objective should be to strengthen the national statistical infrastructure by enhancing the relationship between the NSS and external data compilers. Needless to say, establishing **sound data governance frameworks would be an essential element in developing this "data curator approach"**.

References

African Union Commission (2015): *Agenda 2063 – The Africa we want*, Background Note 1.

Apple (2021): *A day in the life of your data*, April.

Basel Committee on Banking Supervision (BCBS) (2013): *Principles for effective risk data aggregation and risk reporting*, January.

Bean, C (2016): *Independent review of UK economic statistics*, March.

Biancotti, C, A Rosolia, G Veronese, R Kirchner and F Mouriaux (2021): "Covid-19 and official statistics: a wakeup call?", *Bank of Italy Occasional Paper*, No 605, April.

British Academy and The Royal Society (2017): *Data management and use: Governance in the 21st century*, joint report, June.

Choi, S and Y Hashimoto (2017): "The effects of data transparency policy reforms on emerging market sovereign bond spreads", *IMF Working Paper*, WP/17/74, March.

Civil Service Quarterly (CSQ) (2019): "An interview with John Pullinger, former UK National Statistician", *Blog*, 30 October.

De Beer, B and B Tissot (2020): "Implications of Covid-19 for official statistics: a central banking perspective", *IFC Working Papers*, no 20, November.

Development Initiatives (2020): "Poverty in Uganda", *National and regional data and trends*, October.

Diokno, B (2019): *Speech at the data governance seminar*, Manila, 14 June.

Drozdova, A (2017): "Modern informational technologies for data analysis: from business analytics to data visualization", *IFC Bulletin*, no 43, March.

European Central Bank (ECB) (2008): *ECB Statistics Quality Framework (SQF)*, April.

European Commission (2020a): "A European strategy for data", *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions*, COM (2020) 66 final, February.

European Commission (2020b): *Data governance and data policies at the European Commission*, July.

European Commission, International Monetary Fund, Organisation for Economic Cooperation and Development, United Nations and World Bank (2009): *System of National Accounts 2008*.

Eurostat (2017): *European statistics code of practice*, November.

Financial Stability Board and International Monetary Fund (2009): *The financial crisis and information gaps*.

——— (2015): *The financial crisis and information gaps – sixth implementation progress report of the G20 Data Gaps Initiative*.

Fischer, S (2002): "Financial crises and reform of the international financial system", *NBER Working Paper*, no 9297, October.

G20 Italian Presidency (2021a): "Principled Data Access - Building Public-Private Data Partnerships for Better Official Statistics", mimeo.

——— (2021b): *Communiqué of the Second G20 Finance Ministers and Central Bank Governors meeting*, 7 April.

House of Commons Science and Technology Committee (2012): "The Census and social science", Third Report of Session 2012–13.

Hume McIntosh, S (2021): "Measuring aggregate housing wealth: new insights from automated valuation models", *IFC Bulletin*, no 53, April.

IFC (2016): "Central banks' use of the SDMX standard", *IFC Report*, March.

——— (2017): "Big data", *IFC Bulletin*, no 44, September.

——— (2018a): "The role of data in supporting financial inclusion policy", *IFC Bulletin*, no 47, May.

——— (2018b): "Central banks and trade repositories derivatives data", *IFC Report*, no 7, October.

——— (2019a): "INEXDA – the Granular Data Network", *IFC Working Paper*, no 18, October.

——— (2019b): "Business intelligence systems and central bank statistics", *IFC Report*, no 9, October.

——— (2020): "Computing platforms for big data analytics and artificial intelligence", *IFC Report*, no 11, April.

——— (2021a): "Use of big data sources and applications at central banks", *IFC Report*, no 13, February.

——— (2021b): "Micro data for the macro world", *IFC Bulletin*, no 53, April.

Institut national de la statistique et des études économiques (INSEE) (2020): "High-frequency data are especially useful for economic forecasting in periods of devastating crisis", *Point de Conjoncture*, June, pp 29–34.

Inter-Agency Group on Economic and Financial Statistics (IAG) (2017): *Update on the Data Gaps Initiative and the outcome of the Workshop on Data Sharing*, March.

International Monetary Fund (IMF) (2018): "Overarching strategy on data and statistics at the Fund in the digital age", *IMF Policy Paper*, March.

International Statistical Institute (2010): *Declaration on professional ethics*, July.

Lane P (2021): "Maximising the user value of statistics: lessons from globalisation and the pandemic", speech at the European Statistical Forum (virtual), 26 April.

Lazer, D, R Kennedy, G King, and A Vespignani (2014): "The parable of Google Flu: traps in big data analysis", *Science*, vol 343 (6176), pp 1203–05, 14 March.

Norwegian National Research Ethics Committees (2019): *A Guide to Internet Research Ethics*, May.

——— (2020): *Opinion on research ethics and artificial intelligence*, October.

Nuti, S, B Wayda, I Ranasinghe, S Wang, R Dreyer, S Chen and K Murugiah (2014): "The use of Google Trends in health care research: a systematic review", *Plos One*, October.

Organisation for Economic Co-operation and Development (OECD) (2002): *Quality Framework for OECD Statistics*, Paris, June.

——— (2019): "The path to becoming a data-driven public sector", *OECD Digital Government Studies*, OECD Publishing, Paris, November.

Prévost, J-G and J-P Beaud (2012): "Statistics, public debate and the state, 1800–1945: A social, political and intellectual history of numbers", *Studies for the International Society for Cultural History*, no 1, Pickering and Chatto (eds).

Statistical Journal of the IAOS (SJIAOS) (2020): "Official Statistics in the context of the Covid-19 crisis", statement prepared by W Radermacher.

The Economist (2017): "The world's most valuable resource is no longer oil, but data", 6 May edition.

The Global Partnership for Sustainable Development Data (2019a): *Five-year strategy 2019-2023*, August.

——— (2019b): *Ethics and Integrity Framework*, October.

Tissot (2017): "Using micro data to support evidence-based policy", paper presented at the 61st Congress of the International Statistical Institute, Marrakesh, July.

——— (2019): "Making the most of big data for financial stability purposes", in Strydom and M Strydom (eds), *Big Data Governance and Perspectives in Knowledge Management*, IGI Global, pp 1–24.

United Nations (UN) (2013): "Fundamental Principles of Official Statistics", resolution adopted by the Economic and Social Council, E/RES/2013/21, 28 October.

——— (2015): *Transforming our world: the 2030 agenda for sustainable development*, A/RES/70/1.

——— (2019): *United Nations National Quality Assurance Frameworks Manual for Official Statistics (UN NQAF Manual)*, March.

United Nations Economic Commission for Europe (UNECE) (2018): *Recommendations for promoting, measuring and communicating the value of official statistics*.

Wibisono, O, H D Ari, A Widjanarti, A Zulen and B Tissot (2019): "The use of big data analytics and artificial intelligence in central banking", *The Capco Institute Journal of Financial Transformation,* Data analytics, 50th edition, November.

Witt, E and J Blaschke (2019): "ECB data for analysis and decision-making: data governance and technology», *IFC bulletin*, no 49, January.

Zaidi, E and G De Simoni (2019): "Augmented data catalogs: now an enterprise must-have for data and analytics leaders", *Gartner Report*, 12 September.