

A fast-evolving landscape for Official Statistics: How to respond to the challenges?¹

Martine Durand^{a,*} and Jan Robert Suesser^b

^a*Member of the French Statistics Authority and Former OECD Chief Statistician*

^b*Member of IAOS Executive Committee and of ISI Advisory Board on Ethics*

Today, we live in a digitised world where data has become a global commodity that is produced, shared and traded by a myriad of actors around the world. Sometimes called the data revolution, this trend is facilitated by technological innovations such as the Internet of Things or Artificial Intelligence. It reflects a strong appetite for data-driven decisions by governments, businesses and citizens alike, and increased demand for, and use of data in the public debate.

In this complex new data ecosystem, competition for timely and more relevant data is strong. New actors, such as private producers or data scientists, are often perceived as agile and innovative. They can respond quickly to demands for more timely and more granular information by mobilizing and integrating different data sources and by using a range of communication and dissemination channels. As a result, Official Statistics are increasingly considered as only one actor among many in the data ecosystem, albeit an important one. It is fair to say that Official Statistics have lost their quasi-monopoly position in the production and dissemination of statistics that contribute to informing the public debate. Their authority is being contested and the once prevailing ‘statistics logic’ has been replaced by a ‘data logic’.²

In this new environment, not all players have the same approach to the notions of statistical independence, integrity, quality, transparency or privacy, however. In the field of dissemination, social media bubbles, alternative facts, emotions and beliefs often prevail over facts. All this poses great challenges for trust, governance and ethics in data and statistics. It also increases the risks of misinformation and misuse of data and statistics, with consequences for the functioning of democracies.

So, what should Official Statistics do in this fast-evolving environment? Sticking to their main mission of building knowledge for the common good aiming at informing public policies, the public debate and society, Official Statistics have a key role to play. They need to be bold, forward-looking, innovative and ahead of the curve. They should be at the forefront of actions aimed at addressing misuse, improving governance, fostering ethics and enhancing trust in relevant data and statistics. Official Statistics should identify the main issues at stake and tackle them in a holistic way to capture their interlinkages. Indeed, governance is key for trust and addressing misuse; misuse and data manipulation are detrimental to trust; professional ethics and adherence to deontological principles are essential for enhancing trust; and in turn trust in data and statistics are a prerequisite for data uptake, use and impact. Official statistics have to act on all these fronts if they are to contribute to making the data and statistics ecosystem a safe, relevant and trusted environment. To be effective in their endeavor they need to engage and partner throughout the whole data value chain with all the other relevant actors present in the data ecosystem.

¹The views expressed here are the authors’ sole responsibility.

*Corresponding author: Martine Durand, Member of the French Statistics Authority and Former OECD Chief Statistician. E-mail: martinedurand2905@gmail.com.

²See Radermacher Walter J, *Official Statistics 4.0, Verified Facts for People in the 21st Century*, <https://springer.com/gp/book/9783030314910>.

1. Building on strong assets

Official Statistics have strong assets to mobilize and build on to achieve this objective. They have an established reputation for the quality, reliability and transparency of the data produced; their activities are based on strong legal frameworks, strong methodological standards and recognised codes of ethics that ensure professional independence, objectivity, impartiality, relevance, privacy and confidentiality. Official Statistics should communicate more about these strengths that constitute their core values and underpin the key difference between well-designed statistics and any data.

2. Addressing weaknesses through innovation

But Official Statistics also have to address a number of weaknesses. As said earlier, Official Statistics suffer from a (perceived) lack of flexibility, agility and responsiveness. Typically, this is because applying sound methodologies and ensuring high quality necessarily requires time to validate statistics and ensure they are reliable and fit for purpose – something vital but often overlooked. In many instances however, the potential trade-off between timeliness and quality could be improved by exploring more innovative solutions. For example, methodologists in National Statistical Offices should continue to reflect on how to exploit new data sources and invest in data science techniques to turn big data into statistics. This will contribute to responding to policy-makers' and citizens' demands for more timely and granular information while respecting fundamental professional and ethical principles.

In order to avail of alternative data sources (administrative, commercial, web-scraped and text-mined, open data, geospatial, IoT, citizen-generated, etc.), long-term viable frameworks, regulations and partnerships for data access should be put in place with adequate rules and incentives. In the specific case of administrative data, National Statistical Offices (NSOs) should review and modernise the protocols and regulations to access and use administrative data for statistical purposes and get involved early on when new data collections are planned in public administrations. They should also invest in web scraping and crowdsourcing techniques for direct data sourcing. New investments are also necessary to ensure interoperability and build platforms for sourcing, managing, using, re-using and re-purposing all kinds of data, and in infrastructures that allow to integrate and combine established sources

with alternative data sources.³ These changes will require that NSOs review their skills strategies. They will need to invest more in data science skills, by hiring new cohorts of statisticians well-trained in new techniques, data analysts and data scientists, and by training and upskilling existing staff.⁴

Official statistics have demonstrated that they were up to the task during the COVID-19 crisis. Faced with disruptions in their standard collection methods, many NSOs have shown agility and responsiveness. They have quickly availed of alternative data sources to produce GDP growth statistics that have allowed to assess in a timely manner the consequences on economic activity of lockdowns and support measures taken by governments in response to COVID-19. Similarly, many NSOs have developed experimental statistics to inform on key health and social outcomes. Going forward, it is important that these initiatives do not remain one-off but on the contrary are embedded in long-term NSOs' strategies and practices.

3. Combatting misuse and misinformation

The same holds true for communication and dissemination. Official Statistics often suffer from poor communication strategies within the new, more modern means of data dissemination. In the face of an abundance of data (the so-called data deluge), some people don't hesitate to pick up (or make up) the data that suit their agenda to make authoritative statements or wrong inferences. Official Statistics are typically reluctant to address misuse and misinformation publicly as they often consider they lack the legitimacy to intervene in public debates, since this could in their view be in contradiction with their political independence.

In fact, Official Statisticians should have a duty to speak out if data quality is poor and thus misleading, if statistics are being misused or manipulated by public or private lobbies and vested interests. They should correct inaccurate facts and errors of interpretation. To prevent misuse and misinterpretation, Official Statistics should contribute to the statistical literacy of both data producers and users. In order to reach out to a wider

³See for example in the area of health data in France <https://www.health-data-hub.fr/> and in Germany <https://www.health-x.org/home?hsLang=de>.

⁴The OECD Smart Data Strategy provides an example of such a comprehensive approach. See <https://www.oecd.org/sdd/OECD-Smart-Data-Strategy-Vision-Statement.pdf>.

audience, they should also expand their social media culture and be more present where opinions are formed, as well as build partnerships and collaborations with other actors, in particular the media and fact-checkers – actions that a number of NSOs are already successfully pursuing.

As for data sourcing, a lot of innovation in data visualisation and communication tools has taken place in NSOs during the COVID-19 pandemic, related not only to established statistics but also to experimental ones. This increased presence in the media during the health crisis through blogs, data viz and articles has reinforced Official Statistics' reputation as a reliable, trusted source of information. Again, it is important to pursue these initiatives and invest in new ones over time.

4. Improving governance

The presence and active engagement of more and different actors in the data ecosystem require to review existing governance arrangements. Unlike for Official Statistics, which are typically governed by statistical laws that formally set out their mission, obligations, means and interactions with other stakeholders, at present, there are no formal institutional frameworks governing the data ecosystem at national and international levels, not even in domains where there is an obvious overlap in the data covered by Official Statistics, academia and the private sector. Over the past couple of years, some countries and international organisations have adopted a number of laws, regulations and guidance related to data governance. But these mainly deal with issues of personal data privacy, cross-border exchange of data, or data sharing among private producers and private and public actors. They generally do not cover statistics nor the inter-relation between statistics and data. This is a gap that will need to be filled and where Official Statistics should make their voice and concerns heard and addressed.

Discussions should be held with public and private actors to clarify the roles, responsibilities and accountability of the different participants in the data ecosystem. In terms of roles and responsibilities, even within the public sector, there are clear lines that Official Statistics cannot and should not cross. For example, while data scientists working in government departments may produce dashboards or other information that contain personal or very granular data for policy-makers, it is clear that official statisticians can only provide aggre-

gated information that preserve confidentiality. In other words, statistics have a different function than data used for direct operational purposes and this should be preserved.

NSOs typically play a weak role in the new data ecosystem. This may reflect insufficient financing for Official Statistics, especially in lower-income countries. They cannot compete with big tech companies that provide their data services to governments, academia and the general public. The need for adequate laws and regulations that recognise Official Statistics as a key player in the data ecosystem is particularly important in this case. More generally, NSOs could play an important stewardship role in areas where they have a lot to contribute, for example with respect to standards, classifications, methodologies or quality assurance. They should pro-actively engage and build partnerships with other public providers and private actors to offer their technical, methodological and legal services.

They should also engage in and lead co-investment, co-innovation and co-production with other actors. This may require a shift in the 'business model' of NSOs that would entail a much larger degree of co-investment in data science capabilities and a much broader sharing of data and knowledge. Within the national statistical system for instance, Official Statistics could create communities of practice that bring together statisticians and data scientists who can share and develop common knowledge around specific issues (e.g., data linkages, algorithms, geospatial information, etc.). Open-source collaboratives hold the promise of meeting more effectively the increasing users' demands by pooling resources and solutions, while reinforcing the NSOs' position in the data ecosystem.⁵

5. Enhancing trust

Organised collaboration, sharing of knowledge and best practices, rather than outright competition between Official Statistics and other data actors is likely to deliver more and better products and services to users and thereby enhance trust in data and statistics. Indeed, Official Statistics do not operate in isolation. We observe a general trend decline in trust in public institutions

⁵Examples of such collaboratives include the UNECE HLG-MOS <https://unece.org/statistics/networks-of-experts/high-level-group-modernisation-statistical-production-and-services>; the OECD-led SIS-CC <https://siscc.org/> or the World Bank-led Development Data Partnership <https://datapartnership.org/>.

and governments. Many people feel that they have little or no say in public decisions and that public policies do not address their concerns. While restoring people's trust in institutions requires more than data and statistics, providing relevant information about people's situations contributes to the quality of the public debate, a key part of the mission of Official Statistics.

Tapping into new data sources, including citizen-generated and geospatial data, investing in data science skills, data labs, experimental statistics and data viz, as mentioned earlier, combined with standard household surveys and administrative data will contribute to providing more timely and granular information on people's living conditions, on what matters most in their lives, with the potential to reinforce trust in Official Statistics. But for this to happen, this information has to be made readily available and easily accessible. Official Statistics have to make their data available, and more generally have to adopt open data policies by default.

In addition, as more granular information may be derived from individual administrative records or other personal data, it is essential that citizens and companies have absolute confidence in the system that has generated this information, that safeguards are in place to ensure data privacy and confidentiality, and that there is no interference and full independence from public authorities or lobbies. In these respects, any breach, or even perceived breach would greatly undermine trust.

Similarly, ensuring data quality in the new data ecosystem is essential for building and enhancing trust. The recourse to new data sources, including private ones, their linkages and integration with established sources implies reviewing the long-established quality frameworks that are commonly used in Official Statistics. Working with private and public actors, Official Statistics should take the lead in developing new or enhanced quality frameworks that take into account the nature of the new data used, the new processes put in place, as well as the new statistical products to offer. These new frameworks should thus encompass the whole data value chain, from the collection, to production, dissemination and use. There should be full transparency about potential caveats and limitations of the new statistical products and clear guidance on their interpretation.

As we have witnessed over the past decades, the data revolution is happening at a speed never seen before. NSOs and Official Statistics need to design strategies to anticipate potential sources of crises in the future and already develop the appropriate tools to respond and continue to deliver on their mission to inform policy-

making and public debates. Their readiness and capacity to react will, as we have seen during the COVID-19 pandemic, be a major factor to put Official Statistics at the centre of the data ecosystem and reinforce them as a trusted institution.

6. Fostering ethics

A very important characteristic of Official Statistics that highly contributes to the trust the general public attaches to them is their strong ethical underpinnings that have been collectively developed and updated over several decades. Existing sources of ethics for statistics are many at both national and international levels. These include *inter alia* national constitutions, national statistical laws, the UN Fundamental Principles for Official Statistics, the European Statistics Code of Practice, the OECD Council Recommendation on Good Statistical Practice, the Principles Governing International Statistical Activities, and the International Statistical Institute Declaration on Professional Ethics.

These sources of ethics for statistics are challenged by the new data environment. As already mentioned, some players (e.g., digital companies, data scientists, social media) may not be aware of, or may not feel concerned by the need to follow specific codes of ethics with respect to quality, independence, transparency, privacy or confidentiality, fairness and doing no harm, proportionality and necessity.⁶ Others who have shown interest in the issues (e.g., geospatial organisations) may be in the process of developing their own rules for their specific needs or clients. There is a risk that in the absence of involvement or coordination with Official Statistics, we may see an increasing number of private initiatives that simply ignore well-established statistical standards and codes of ethics, or the emergence of very disparate sets of principles that may be in contradiction with the ethical principles guiding Official Statistics.

Official Statistics should again be proactive to foster the adherence to strong ethical principles for data and statistics. They should speak up about, and advocate the existing codes of ethics for Official Statistics. A possible way forward would consist in initiating work with other relevant actors, including users, towards the creation of an International Forum to address issues around ethics

⁶For examples of initiatives that create federated data infrastructures involving private and public data providers, see: <https://www.gaia-x.eu/> and <https://beta.nsf.gov/science-matters/americas-datahub-consortium-seeing-and-understanding-entire-elephant>.

but also misuse, governance and trust. The aim of such a Forum would be to develop and promote the effective use of a consistent set of core values and of universal ethical data principles, based on existing frameworks and recent initiatives, to be endorsed and followed by public and private actors as well as academia. These principles could cover regulations and arrangements encompassing professional independence, responsibilities and accountability, methodological excellence, protection of personal data, data accessibility, data sharing and exchange, data interoperability, fairness and doing no harm principles, as well as prevention of abuse and misuse of data.

At the same time, to make such a Forum more effective, some people have made the potentially controversial proposal to create an independent enforcement mechanism that ensures that all participants in the International Forum are in compliance with the agreed principles. In their view, this enforcement mechanism could operate through regular assessments containing recommendations for improvements, similar to the Peer Reviews of the European Statistical System and the OECD Recommendation on Good Statistical Practice. The findings of the regular assessments would be made public. The Forum and its enforcement mechanism would permit to create alliances among actors to combat bad practices and promote good ones.

7. Conclusions

Official Statistics have a large agenda ahead if they are to continue to play a key role in the fast-evolving data ecosystem. Almost every day we hear about initiatives by private actors that put at risk various aspects of the fundamental mission of Official Statistics. Official Statistics have a lot of strengths to build on and mobilize but they also need to address a number of weaknesses. As they have demonstrated during the COVID-19 crisis, Official Statistics are able to act to innovate in both their processes and products, combat misuse, improve governance, enhance trust and foster ethics.

They should not act alone in this endeavor, but rather build partnerships with other actors in the data ecosystem, and where appropriate take leadership and act as the data steward. To further discuss the issues developed in this editorial – some more controversial than others – propose scenarios and advance proposals, the IAOS has launched a new Reflection Group at the 2022 Conference held in Krakow last April. In this way, the IAOS hopes to make a key contribution toward fostering integrity and trust in an evolving data ecosystem that truly works for the public good.