The contribution of UN data to the 2030 Agenda and UN Reform

Innovating now for better information in the future

Committee of Chief Statisticians of the UN System
September 2019

Introduction

The 2030 Agenda for Sustainable Development has presented an enormous challenge for the UN statistical system. So much so that Morgens Lykketoft, President of the seventieth session of the UN General Assembly, described it as an 'unprecedented statistical challenge'. The reform of the UNDS[1], which aims to make the UN system stronger, with a better-defined collective identity and with greater cohesion and accountability, presents challenges and opportunities for the UN statistical system, at national, regional and global level.[2] The 2030 Agenda has increased the visibility of data and statistics. The chief statisticians of UN agencies and departments would like to capitalize on this visibility and use it to contribute positively to the wider reform agenda.

To meet this unprecedented statistical challenge and exploit the opportunities it presents, the UN needs to modernize the way that data are collected, processed, integrated, disseminated and communicated. Modernization means changing the ways that we produce global and regional information, drawing on a broader range of data sources and analytical tools that can help to deliver more timely and nuanced information and utilizing modern dissemination platforms that can facilitate communication to a broader audience. Modernization also means rethinking the ways in which the UN, in cooperation with other International Organizations, promotes and delivers capacity development for work on data and statistics.

The High-Level Committee on Programmes (HLCP) has already highlighted the need to improve the quality and availability of data and statistics and has called, for example, for greater availability of disaggregated data, more user-friendly and policy-relevant data, and strengthened capacities for collecting and analysing data. Many of the recent documents adopted by the UN system (on migration, urbanization, disability, artificial intelligence,

[1] General Assembly of the United Nations in Resolution A/RES/72/279 of 31 May 2018.
[2] At regional level, for example, the Secretary General has called for the establishment of a coherent regional data ecosystem (A/74/73–E/2019/14).

education, drug policy, climate change to name a few) have highlighted the need for better and more timely data.

Also, the Independent Expert Advisory Group on a Data Revolution for Sustainable Development, in their 2014 report '*A World That Counts*', recommended that 'New institutions, new actors, new ideas and new partnerships are needed'. Following the UN World Data Forum in 2018, the *Dubai Declaration* acknowledged that 'the data demands for the 2030 Agenda require urgent new solutions that leverage the power of new data sources and technologies through partnerships between national statistical authorities and the private sector, civil society, and the academia and other research institutions.'

In reaction to these calls for action, the Committee of Chief Statisticians of the UN System (CCS-UN), the group mandated by the UN Statistical Commission to coordinate statistical activities across UN entities[3], have in this paper proposed a suite of innovations that would address at least some of the issues being raised across the UN system. The proposals in this paper are tentative – responsibilities have not been assigned, nor have timelines or budgets been detailed. This is deliberate as the nature of the paper is explorative. The submission of this paper to the HLCP aims at expanding the discussion on modernizing the production of statistical information so that the statistics produced by the UN genuinely respond to policy needs and satisfy people's entitlement to public information. It is hoped that the HLCP can shape a vision or road map for modernizing UN statistics that could be adopted by the CEB. With this in mind, the paper offers suggestions for concrete actions that the UN can take as part of this road map.

The note is presented in 3 short sections: (1) new data solutions; (2) innovative capacity development; and (3) responding to policy needs. In each section, a brief context is presented, followed by a set of practical and feasible actions. Some current activities are also summarised to highlight positive steps already being taken by the UN statistical system – though far from an exhaustive list, it is illustrative of the concrete coordination already underway. The paper finishes with a short conclusion.

Message 1. The UN system should integrate complimentary data sources to create new data solutions

1.1 Context

The bulk of data and statistics compiled and disseminated by UN entities are derived from statistics officially submitted by Member States. These official national statistics are compiled mainly from observations collected for statistical purposes by institutions that comprise national statistical systems. Despite innovations employed by national statistical institutions, the main data sources of official national statistics remain the traditional ones: population censuses, administrative records and registries, and sample surveys. New data sources will not replace traditional sources for the foreseeable future, but they can nevertheless offer

[3] The membership of the CCS-UN is listed in Appendix 1.

insights and more frequent measurements to complement and enhance traditional sources. They can also be used as the basis to fill data gaps.

The volume of digital information is growing with startling speed. The unprecedented richness of information available, combined with an equally large and growing appetite for statistics and data from users, is providing new opportunities for national and international statisticians[4]. The challenge for UN entities is to understand whether and how they can access, use and integrate these new sources, in particular to meet the data requirements of the Sustainable Development Goals.

To use the new data sources requires developing new mind-sets and capabilities in national and international statistical organizations. The capabilities needed are multi-dimensional, requiring investment in areas such as technology, methods, processes, information management, standards and frameworks, whilst not forgetting the need to develop the necessary human skills and institutional arrangements (legal, policy and organizational).

The use by the UN of new data sources to complement statistics regularly submitted by Member States ('national official statistics') can raise political challenges with Member States. The Committee on Coordination of Statistical Activities (CCSA)[5] has produced a paper on best practices for using non-official statistics to support the adoption of new sources. This paper emphasizes the critical role of communication between agencies and Member States and the importance of adopting strict quality standards.

There are also technical challenges in using new data sources, including availability or access, representativity of the data, suitability of the information, lack of appropriate skills for analysis, and lack of adequate computing power given the volume of data to be analyzed[6].

The innovations proposed in this section are designed to improve the UN's ability to integrate complimentary data sources, create new data solutions and benefit from efficiencies of scale.

1.2 Practical steps

I. The UN system could use the UN Global Platform[7] as a single point of entry for partnerships with international providers of data sources (including a unified approach to negotiations), and as a cost-effective mechanism to engage in common data projects, bringing together and sharing global data sets and new methods, tools and services.
II. Building on the work led by UN Global Pulse and the Office of Information and Communications Technology, UN agencies could adopt a common set of ethical and

[4] See Appendix 2 for a brief typology of the digital data sources.
[5] This coordinating mechanism has similar functions from the CCS-UN but with a broader membership as it includes the chief statisticians of non-UN agencies such as OECD, Eurostat, the World Bank and regional development banks, and other global or regional organizations.
[6] A fuller description of this topic is provided in Appendix 3.
[7] The UN Global Platform is a "collaborative research and development environment for the global statistical community."

quality standards for the use of "Big Data". While quality standards for 'traditional' statistics are well established[8], new data sources have opened-up a range of questions regarding confidentiality, quality assessment, and considerations for the protection of individual privacy and human rights.

III. Through existing mechanisms, such as, the UN Global Platform, the CCS-UN and the CCSA, the UN system could develop a multi-sectoral UN expert network on the use of new data sources. The network would harness existing expertise within UN entities to exchange experiences and benefit from a multi-disciplinary approach to using new data sources. The network would include UN staff from various disciplines, including statistics, data science, development research, information technology, remote sensing, geography, demography and economics.

IV. Develop a strategic alliance between the geospatial information and statistical communities in the UN to reinforce the existing synergies and to encourage both communities to develop innovative approaches to data collection, visualisation and dissemination.

1.3 Some on-going initiatives already underway

I. "Big data". The UN Statistical Commission created the *UN Global Working Group (GWG) on Big Data for Official Statistics* in 2014. The GWG provides strategic vision, direction and coordination for global action on the use of "big data" for official statistics, including, the compilation of SDG indicators for the 2030 Agenda for Sustainable Development. The use of big data is an important part of the modernization agenda of national statistical institutions. The GWG currently consists of 28 UN member states and 16 international agencies. In 2018, the GWG launched the UN Global Platform, a collaborative research and development environment for the global statistical community. Using networking and marketplace principles, the platform facilitates the exchange, development and sharing of data, methods, tools and expertise with the aim of accelerating data innovation. The platform will also be used to coordinate capacity-building activities dealing with the development of data science skills and the use of big data and machine learning techniques.

II. Big data and ethics. A guidance document by the UNDG, *Data Privacy, Ethics and Protection: a guidance note on Big Data for the achievement of the 2030 Agenda*[9], sets out general guidance on data privacy, data protection and data ethics concerning the use of big data. The note establishes common principles for UN entities on the use of new data sources for measuring progress towards the Sustainable Development Goals; serves as a risk-management tool with particular reference to human rights; and sets principles for the acquisition, retention, use and quality control of data coming from UN Global Pulse, the UN Innovation Network or the private sector.

III.

IV. Geospatial information systems. The newly established UN Geospatial Network can be a useful resource for supporting the adoption of new data sources by the UN system. Geospatial tools are already been used by many UN agencies to disseminate

[8] See the UN Statistics Quality Assurance Framework - https://unstats.un.org/unsd/unsystem/documents/UNSQAF-2018.pdf

[9] https://undg.org/wp-content/uploads/2017/11/UNDG_BigData_final_web.pdf

and visualize statistics, but few agencies use them for data collection. The UN Geospatial Network can offer opportunities to improve the data collected for monitoring the SDGs. Satellite images and remote sensing data, for example, continue to evolve, and the information they provide is becoming more accessible and accurate, offering major opportunities to the UN statistical system. Also, the geo-referencing of individual observations in data sources used to derive measures of human development and well-being may offer the best opportunity available for obtaining reliable disaggregated data that can be used to identify populations at risk of being left behind.

V. Open SDG hub for reporting on SDG indicators. In 2018, the Statistical Commission agreed to create a federated system of national and global data hubs for the SDGs to facilitate integration of different data sources, promote data interoperability[10], foster collaboration among partners from different stakeholder groups, including the geospatial community, and improve data flows and global reporting of the SDGs. The open SDG hub makes data of the global SDG Indicators database readily available as geospatial data web services, suitable for the production of maps and other data visualizations and analyses, and easy to download in multiple formats.

Message 2. The UN should encourage national statistical systems to use new data sources and should adopt innovative and coordinated ways to deliver technical assistance in statistics, and increase partnership among agencies in delivering capacity development programmes

2.1 Context

A challenge facing all NSOs and national statistical systems is how to balance the considerably increased demands of Agenda 2030 with the multitude of other national data priorities that exist. As countries progress along the development spectrum or find themselves on different slopes of the business cycle, their data and information requirements, will be different and will change. The 2030 Agenda, like the Millennium Development Goals before it, has a fifteen-year lifespan. In the context of history, 15 years is the blink of an eye. Nevertheless, we can safely predict that, for many individual countries in various parts of the world, during the next 10 years, many unanticipated events will unfold and some of these will have a dramatic and profound impact on national priorities and development plans. Natural disasters, such as, tsunamis, earthquakes and famines; and political or economic events, such as, wars, the unravelling or creation of political unions, the emergence of new countries or new financial or economic crises may all play their part in shifting international or national priorities. Thus, we can anticipate that national and global priorities will change over the next 10 years.

The challenge therefore for countries, and for the UN, is how to build affordable statistical systems that are sufficiently flexible and responsive to these evolving priorities but can at the same time satisfy the enormous appetite of the SDG indicator framework. This poses a dilemma - in designing statistical capacity building programmes, how can a global indicator

[10] See http://www.data4sdgs.org/index.php/resources/interoperability-practitioners-guide-joining-data-development-sector

framework (with urgent data demands) be maintained and populated so that it does not suffocate or overwhelm national and regional priorities (with current and long-term data needs) but rather feeds off them? While pondering this dilemma, it is worth recalling that capacity development is most effective when it is home-grown, long-term in perspective and managed collectively by those who stand to benefit. The key therefore must be to develop capacity that contributes to the development of national statistical systems, their management and leadership and promotes a culture of open data while also embracing a culture of statistical innovation with the use of new data sources complementing traditional sources to better serve the need of policy makers.

The capacity development innovations proposed in this section are designed to contribute to the strengthening of national statistical systems, while simultaneously supporting the SDG global indicator framework. The focus is on innovative ways to provide technical assistance and ensure coordination at field level.

The innovations proposed in this section are designed to improve the UN's ability to deliver a more efficient and coherent capacity development programme to member states.

2.2 Practical steps

I. UN country teams could adopt a united and sustainable strategic vision for national statistics through a *'One UN' National Statistical System Capacity Development Programme*. Within the CCA and UNSDG mechanisms, UN country teams assess capacity and needs of national statistical systems and articulate a coordinated UN statistical assistance programme which promotes sustainability, open data and the use of new data sources where relevant and feasible. Existing bodies and mechanisms such as the CCS-UN, Paris21, and the UNSDG Working Group 1 Task Team 3 (dealing with country level data and reporting) can provide the tools to support UN country teams in this process. The revised *UN Handbook of Statistical Organisation* can also serve as a guide.

II. Partner with major regional statistical training institutes around the world to train UN personnel and members of national statistical systems on how to implement the *'One UN' National Statistical System Capacity Development Programme*. This could build on the work of the existing global network of institutions for statistical training[11].

III. Design a 'One UN' *Statistical Leadership Programme*, where Director Generals of National Statistical Offices are trained. This could be designed in conjunction with universities around the world and would involve senior short visiting programmes with UN agencies and other National Statistical Offices. This program would focus on management and leadership rather than on statistical techniques. The programme could eventually become the embryo of master programmes in selected universities.

IV. A 'One UN' Statistical *Capacity Development Calendar* could be developed. This would allow all agencies and UN country teams to understand the scale and timing of bespoke statistical capacity development planned around the world. This could be used to better coordinate capacity development activities, as a more coherent offer to member states.

[11] See https://unstats.un.org/GIST

V. Following the call to create 'knowledge hubs' in each region, these hubs should incorporate the notion of *'Statistical Innovation Hubs'* and modernization. These hubs would build on pioneering work already begun by UNECE and other national innovations. They could also identify and establish a pool of statistical and data science experts in each region who can be mobilized to provide specialized advice and assistance on the range of data sources that can be harnessed and other innovative statistical practices. The UN Global Working Group on Big Data for official statistics is in the process of establishing UN Big Data hubs in Brazil, China, Rwanda, UAE and the UK.

VI. The *UN World Data Forum* could give dedicated space to showcase innovative solutions already developed and pioneered by the UN statistical system. Presenting both entity and common or 'One UN' developments would help member states to see statistical coordination in action.

2.3 Some on-going initiatives already underway

I. The High-level Group for Partnership, Coordination and Capacity Building, established by the UN Statistical Commission in 2015, provides strategic guidance to the global statistical system for the production and use of data for the SDGs and the implementation of the data revolution. The group spearheaded the Cape Town Global Action Plan for Sustainable Development Data, which provides a framework for planning and implementing statistical capacity building efforts necessary to achieve the scope and intent of the 2030 Agenda, including through innovation and modernization of national statistical systems and the application of new technologies and new data sources into mainstream statistical activities.

II. *'One UN' National Statistical System Capacity Development Programme* can build on the existing SDG Monitoring and Reporting Toolkit for UN Country Teams. This toolkit is developed to support national governments in the monitoring and reporting on the Sustainable Development Goals. It brings together the collective experience of United Nations entities—the funds, programmes, specialized agencies, regional commissions and other parts of the UN Development System—in a single, easy to access page. This toolkit is a 'live' document that will be updated continuously as new resources become available.

III. UN capacity development calendar. UNSD launched an initiative in September 2019 to compile a statistical capacity development calendar for all UN entities. This calendar will be constructed over the coming months and thereafter updated on a regular basis.

IV. UNSDS Task Team. With a view to supporting UN country teams to design statistical programmes, the UNSDG Working Group 1, Task Team 3 has developed a repository of UN reference material on the different thematic areas of statistics related to SDG measurement and reporting. The next task will be to mainstream this repository into the work of the UN Country Teams so they can support development of the National Statistical System.

V. Centres of Excellence. UN agencies have begun to adopt a new model for providing technical assistance in statistics at regional and global level in partnership with national statistical offices. The UNODC-INEGI *Centre of Excellence* in Mexico on statistical information on governance, public security, victimization and justice and the UNWomen-INEGI Center of Excellence in gender statistics together with the UNODC-

KOSTAT Center of Excellence on crime statistics in the Republic of Korea are examples of this approach. Partnership with national statistical offices offers a framework whereby assistance benefits from operational support provided by the national statistical office in combination with assistance in meeting international standards provided by the UN.

Message 3. Ensure that the UN statistical system responds to emerging policy needs

3.1 Context

The continued relevance, efficiency and effectiveness of global policy development and operational implementation rely heavily on the availability and use of timely and quality statistics. Data and statistics produced by the UN play a critical role in supporting evidence-informed policy making, but only if they are available when needed and if they address the right questions. In a world of 'Fake News' the UN, as a compiler of quality statistics, plays a vital and legitimate role as a custodian of knowledge and a protector of deliberative public space.

The UN statistical system as a producer of normative standards and as a compiler of international official statistics strives towards making high quality information available, with a particular emphasis on accuracy. This emphasis often means that other dimensions of quality, most especially timeliness or relevance, are sometimes sacrificed to secure precision and preserve time series. The result of this trade-off has not always best served political expectations or needs.

Ensuring policy relevance, for example, entails faster release of data and the engagement in areas where statistical standards may be weak or in earlier stages of development, such as, in the field of human rights and governance.[12] Timeliness of UN data can sometimes lag behind, not capturing the most up to date situation, thereby limiting its relevance to explaining the past rather than helping to address current challenges. Modernizing the UN statistical system requires more attention be given to timely and relevant data.

The Independent Expert Advisory Group on a Data Revolution for Sustainable Development notes in its report '*A World That Counts*' that 'If data is to be useful and support good decision making, it has to be ready at the time when decisions are being made or where the opportunity for influencing the outcomes is there.' Unfortunately, the alignment of data availability with policy decision making cycles is an ongoing challenge for all statistical compilers. However, increases in computing power and methodological development allied with the greater variety and volumes of data now available, offer some tantalizing opportunities to provide real-time or close to real-time statistics. As noted above, a variety of secondary data sources (administrative, geospatial, commercial, social and machine generated) now exist and if they can be accessed offer unique opportunities to disseminate very timely statistics.

[12] See the Praia City Group on Governance Statistics (the Praia Group) created in March 2015 at the forty-sixth session of the United Nations Statistical Commission.

One approach is to mainstream the use of 'nowcasting' or 'contemporaneous forecasting.' Some UN agencies are already using this modelling technique to produce real-time estimates based on changes to other relevant and more timely indicators. These forecasts are not based on macro-economic or exogenous scenario models but rather on the correlations between variables.

Presenting UN statistics in a context where policy makers can clearly see their value is an important step to ensure these are truly used as the basis for the evidence which underpins policy making. Some entities have been active in encouraging national statistical offices to develop story lines around the crude reporting of numbers, but the same approach has not always been mainstreamed within the UN system itself. This is true also for understanding the needs of policy makers. A lot has been done to inject within the UN statistical system a culture of responsiveness to policy needs, but more needs to be done in terms of presentation of data and coverage of policy priorities. For good reasons the UN statistical system has often prioritized its own professional independence (to ensure the impartiality of UN data), but sometimes at the expense of partnerships with policy makers, preventing a systematic dialogue on what is needed versus what can be produced. There is a fine line between professional independence and policy relevance and the UN should discuss more how to ensure both without one undermining the other. Beyond mere policy relevance, there is also a need to consider further statistics for their relevance in fulfilling people's right to public information, in accordance with the first principle of the Fundamental Principles of Official Statistics as well as related internationally agreed human rights standards about access to information. [13]

Being proactive in maximizing the use of UN data is also a goal of the UN statistical system. Statistics are only useful if they are actually used. The Committee on Coordination of Statistical Activities (see footnote 5) has recently gone through an exercise of evaluating the extent to which its members disseminate data in an open fashion ('open data'). None of the UN agencies participating in the evaluation achieved the full score of data openness. Open data is a value that the UN needs to embrace more broadly not only to serve as an example for national institutions, but also for projecting the idea of data as a public good, ensuring that all efforts made by the UN to collect and process data translate into the data being fully used.

The innovations proposed in this section are designed to improve the UN's ability to respond to emerging policy needs.

3.2 Practical steps

I. Systematically nowcast and forecast key indicators used to monitor the SDGs. To facilitate this the UN statistical system would: (1) develop common nowcasting and forecasting methodologies and standards; and (2) organize training and development to roll-out implementation and produce guidelines; (3) alert policy makers on the

[13] Elaboration on the linkage between Principle 1 of the Fundamental Principles of Official Statistics and international human rights standards can be found in OHCHR guidance note on Human Rights-Based Approach to Data: https://www.ohchr.org/Documents/Issues/HRIndicators/GuidanceNoteonApproachtoData.pdf

likely trajectories of SDG indicators through mechanisms such as the "traffic light" system.

II. Develop a 'One UN' Statistics brand. This would include raising awareness across UN entities on the importance of statistics, including for human rights, and provide training on how to visualize and communicate statistics to non-specialist users.

III. All UN entities to achieve full open data by end of 2021.[14]

3.3 Some on-going initiatives already underway

I. Nowcasting or short-term contemporaneous forecasting is a technique for producing a real-time evaluation of an indicator. The Global Partnership for Sustainable Development Data, the World Bank, the DESA United Nations Statistics Division, and SDSN TReNDS have launched a 'Data For Now' initiative which seeks to increase the use of robust methods and tools that improve the timeliness, coverage, and quality of SDG data through collaboration and partnership, technical and capacity support, and information sharing. A number of UN entities, such as, UNCTAD and UNIDO already use this approach for a selection of indicators to respond to an increased demand for up-to-date information to support policy needs. In September 2019, the CCS UN established a working group to share methodologies and begin work toward a set of guidelines.

II. Open Data. In 2018, the United Nations Statistical Commission formally included the issue of open data in its deliberations around the Fundamental Principles of Official Statistics. One year later, at its 50th session, the Commission established a working group on Open Data, tasked with developing guidance for the assessment and practical application of open data in the context of official statistics. As part of these efforts, DESA Statistics Division is collaborating with Member States and partners from Civil Society to develop training materials and conducting capacity building activities aimed to mainstream open data best practices in all stages of the official statistical business process while protecting privacy and confidentiality. Open data principles and practices are also being addressed in the new handbook for national statistical systems.

III. Open data is the idea that some data should be freely available to everyone to use and republish as they wish, without restrictions from copyright, patents or other mechanisms of control. Similar in ethos to the open-source movement, open data is based on the principle that data are a non-rival, public good. Since 2018, the World Bank on behalf of the CCSA has been undertaking an audit of the openness of international organisations. Discussions are now underway to consider how to develop a common 'UN Open Data License'.

[14] To meet all requirements as set up by the work of the CCSA.

Conclusion

We live in an information age. However, for a variety of reasons the information space is cluttered, confused and chaotic, with a multitude of sources publishing a wide range of data and statistics, all claiming to be authoritative. In this chaos, UN data and statistics have never been more vital to support evidence-based decision making and impartial public discourse. In a world where facts and truth are not always valued, the agencies of the UN system, by compiling and publishing statistics recognized for their high quality, play a vital and legitimate role as guardians of the public trust and stewards of information for the 21st century, safeguarding data and statistics as public goods and helping to fulfil people's right to information.

A coherent modernization of data and statistical systems across UN entities will help to enhance the global role of the UN as a credible and relevant source of information on the economic, social and environmental aspects of development, human security, governance and human rights. Any modernization should enhance the reputation of the UN's work on data and statistics, upholding the *Principles Governing International Statistical Activities*[15] and the *UN Statistical Quality Assurance Framework*[16] as the foundation of such activities.

The innovative but practical steps outlined in this note, proposed by the Committee of Chief Statisticians of the UN System, can be seen as first steps towards creating a UN data ecosystem and a 'One UN' statistics brand, which can build on and integrate the work of existing networks. The proposals will improve the position of the UN statistical system to meet current and future demands. Moreover, they will ensure that the UN system is better equipped to integrate new data sources, to create new data solutions, to provide more innovative and coherent capacity-development programmes, and to respond more effectively to emerging policy needs.

The High-level Committee on Programmes may wish to examine the desirability and feasibility of developing a road map / system-wide approach for modernizing UN data, and consider inviting the Committee of Chief Statisticians of the UN System to prepare, in collaboration with interested HLCP members, a proposed draft, based on the practical steps identified in this paper, for HLCP's consideration at its 39th session.

[15] These principles were endorsed by the CCSA in 2005. See - https://unstats.un.org/unsd/ccsa/principles_stat_activities/

[16] This framework was adopted by the Committee of Chief Statisticians of the UN System in 2018. See - https://unstats.un.org/unsd/unsystem/documents/UNSQAF-2018.pdf

Appendix 1: Members of the Committee of Chief Statisticians of the UN system (CCS-UN)

Food and Agriculture Organization of the United Nations (FAO)
International Atomic Energy Agency (IAEA)
International Civil Aviation Organization (ICAO)
International Labour Organization (ILO)
International Organization for Migration (IOM)
International Telecommunication Union (ITU)
Office of the United Nations High Commissioner for Human Rights (OHCHR)
United Nations Children's Fund (UNICEF)
United Nations Conference on Trade and Development (UNCTAD)
United Nations Development Programme (UNDP)
United Nations Economic Commission for Africa (UNECA)
United Nations Economic Commission for Europe (UNECE)
United Nations Economic Commission for Latin America and the Caribbean (UNECLAC)
United Nations Economic and Social Commission for Asia and the Pacific (UNESCAP)
United Nations Economic and Social Commission for Western Asia (UNESCWA)
United Nations Educational, Scientific and Cultural Organization (UNESCO)
United Nations Entity for Gender Equality and the Empowerment of Women (UN Women)
United Nations Environment Programme (UNEP)
United Nations High Commissioner for Refugees (UNHCR)
United Nations Human Settlements Programme (UN-Habitat)
United Nations Industrial Development Organization (UNIDO)
United Nations Office for the Coordination of Humanitarian Affairs (UNOCHA)
United Nations Office on Drugs and Crime (UNODC)
United Nations Population Division (UNPD)
United Nations Population Fund (UNFPA)
United Nations Statistics Division (UNSD)
Universal Postal Union (UPU)
Word Health Organization (WHO)
World Tourism Organization (UNWTO)

Appendix 2: Typology of New Digital Data Sources

- Social networks (human-sourced information): this information is the record of human experiences. Human-sourced information is now almost entirely digitized and stored everywhere from personal computers to social networks. Data are loosely structured and often ungoverned, such as social networks (Facebook, Twitter or Instagram), blogs and comments, pictures (Flickr, Pinterest), videos (You-Tube) or Internet searches.

- Traditional business systems (process-mediated data): these processes record and monitor business events of interest, such as registering a customer, manufacturing a product, taking an order, etc. The process-mediated data thus collected is highly structured and includes transactions, reference tables and relationships, as well as the metadata that sets its context. Examples are data produced by public or semi-public agencies (Medical records) or by businesses (commercial transactions, banking/stock records, e-commerce, credit cards).

- Internet of Things (machine-generated data): derived from the growth in the number of sensors and machines used to measure and record the events and situations in the physical world. The output of these sensors is machine-generated data, and from simple sensor records to complex computer logs, it is well structured. As sensors proliferate and data volumes grow, it is becoming an increasingly important component of the information stored and processed by many businesses. Its well-structured nature is suitable for computer processing, but its size and speed are beyond traditional approaches. Examples include data from fixed sensors (home automation, weather sensors, traffic sensors/webcam, scientific sensors, security/surveillance videos/images) or mobile sensors (mobile phone data, GPS, satellite images, AIS vessel tracking) or data from computer systems (logs or web logs).

Source: Conference of European Statisticians

Appendix 3: Selection of challenges for using 'new' data sources

Negotiated common access to and use of proprietary new data sources

The statistical community has been successful in accessing satellite data for estimation of agriculture and environment statistics and related SDG indicators, in getting access to vessel tracking data for measuring transport and economic statistics and in obtaining scanner data from large retail chains for the calculation of consumer price indices. The UN Global Platform already contains access to several new data sources, such as satellite imagery in many spectral bands and real-time shipping and flight data with hundreds of millions of data points. The data service provided by the Platform is designed to handle additional datasets from platform partners. When these are combined with the cloud, they provide a uniquely powerful toolset.

However, getting access to other kinds of new data sources can be much more difficult. Many of these sources are proprietary i.e. data that are commercially or privately-owned and are not publicly available. For example, data generated from using credit cards, search engines, social media, mobile phones and store loyalty cards are all proprietary and may not be available for use by the UN system. The current proprietary status of some data may change in the future as people around the world realise that their data are being used and traded. But for the moment many datasets are not currently accessible by the statistical community, either because costs are prohibitive, data protection legislation prevents it or proprietary ownership makes it impossible.

Collectively the UN agencies have the possibility to negotiate a much better deal with private data owners than what each organization could do individually. The kind of arrangements could be long-term access to frequent and high-resolution satellite imagery data, access to feeds of social media and data from search engines, or an umbrella arrangement which stipulates pre-negotiated terms for access to mobile phone or credit card data from globally operating data providers, where specific contracts will be further worked out at national level.

Ethical and quality standards in access to and use of new data sources

The UN system must be extremely careful not to damage its reputation and the public trust it enjoys. A UN agency must ensure it does not break the law or stray too far outside the culturally acceptable boundaries or norms of the UN. The UN must decide whether it is legally permissible, ethical or culturally acceptable to access and use big data. These are not always easy questions to answer. When it comes to accessing new sources of digital data, the legal, ethical and cultural boundaries are not always clear-cut. This poses a challenge as public trust and reputation are fragile; hard won but easily lost. For example, from a technical, statistical perspective the most logical and cost-effective method of deriving international travel and tourism statistics might be to use mobile phone data, but from a data protection and public opinion perspective using these types of data may not be acceptable.

This tension or trade-off does not appear to be well understood and is certainly not well reflected in many national and international policy documents. In an increasingly complex data protection environment, there is a growing but discernible mismatch between potential and actual, between expectations and reality. The United Nations Economic Commission for Europe (2016) reflecting on its experiences, notes 'High initial expectations about the

opportunities of Big Data had to face the complexity of reality. The fact that data are produced in large amounts does not mean they are immediately and easily available for producing statistics… Data from mobile phones represent a notable example in this sense. It has been proved that such data can be exploited for a wide range of purposes, but they are still largely outside the reach of the majority of statistical organizations, due to the high sensitivity of the data.' Moreover, the United Nations should be leading on support to Governments for ethical use of new data sources, to uphold human rights in every country.

Developing a multi-sectorial UN expert network on the use of new data sources

A multifaceted transformation of statistical systems is needed to meet the new data innovation challenges and to reap benefits from using new data sources. Existing UN statistical programmes should be broadened to include developing the infrastructures and skills to exploit new data sources and redesign statistical products and services. Transforming technology architectures would facilitate the shift from physical information technology equipment on-site towards the introduction of a cloud-computing environment along with the adoption of common services and application architecture for data collection, registers, metadata and data management, analysis and dissemination.

This approach would need to be accompanied by a changing workforce which can support the progressive diversification of the required new skill sets, ranging from data scientists and data engineers using new multisource data and modern technology, to human rights lawyers strengthening the legal environment, to managers leading the change in corporate culture with a continuously improving quality standard and much broader partnerships. Those new capabilities will allow for the adoption of a new corporate business architecture that is more flexible and adaptable to emerging demands and would be process-based rather than product-based, with an increasing use of new data sources for multiple statistical outputs. In addition, dissemination and communication strategies should be upgraded and made adaptable to target different segments of users by applying a diverse set of data dissemination techniques, including mobile device applications and data visualization of key findings.

Develop a network of data-scientists across the UN system who can complement current statistical and IT skills

A new generation of tools and technologies is being used to exploit the growth and availability of new data sources and innovative methods to provide rich informed measurement and analyses on the economy, the global environment and wider society. The statistical community will also have to work at the frontier of data science and Artificial Intelligence – building skills and applying tools, methods and practices – to create new understanding and improve decision-making for public good.

Data science is a concept to unify substantive expertise, statistics, data analysis, machine learning and their related methods in order to understand and analyze all aspects related to the SDG. It sits at the intersection of mathematics, statistics, computer science, machine learning, traditional research and domain expertise. Partnerships and knowledge exchange with the private sector, academia and the wider research community are crucial elements for a successful application of data science.