

Emerging and systemic issues in Official Statistics

Access to SDG information, data revolution and big data, challenges of using administrative data for statistical purposes, developments in UN GGIM and GGIM:Europe

Michael Nagy, 10 December 2015



UNECE

Questions addressed by the presentation

- Access to environmental information in light of the relevant SDGs, their targets and indicators (e.g. targets 12.8 and 16.10)
- Data revolution and big data
- Challenges of using administrative data for statistical purposes
- Developments in UN GGIM and GGIM:Europe initiative

Role of National Statistical Offices in measuring and monitoring the Sustainable Development Goals



Declaration, adopted by the 63rd plenary session of the Conference of European Statisticians.

Main roles:

- Measuring SDGs
- Coordinate with data producers
- Leadership in dissemination and communication of SDG data
- Advise in the interpretation of data
- Develop statistical capacity in countries

SDG targets addressing access to information



- **12.8 - By 2030, ensure that people everywhere have the relevant information and awareness for sustainable development and lifestyles in harmony with nature**
 - Proposed indicator: *Percentage of educational institutions with formal and informal education curricula...* (grey – more in-depth discussion still needed)
- **16.10 - Ensure public access to information and protect fundamental freedoms, in accordance with national legislation and international agreements**
 - Proposed indicator: *Number of verified cases of killing, kidnapping, enforced disappearance, arbitrary detention and torture of journalists, associated media personnel, trade unionists and human rights advocates in the previous 12 months* (grey – more in-depth discussion still needed)

Fundamental Principles of Official Statistics honour citizens' entitlement to public information

Statistics and dissemination is fundamental for ALL SDGs!

Fundamental Principles of Official Statistics

Providing high quality statistical data and ensuring confidentiality at the same time

Developed by UNECE in 1991 and adopted by the UN General Assembly in January 2014

- Principle 1: *Official statistics provide an **indispensable element in the information system of a democratic society, serving the Government, the economy and the public** with data about the economic, demographic, social and environmental situation. To this end, official statistics that meet the test of practical utility are to be compiled and made available on an impartial basis by official statistical agencies **to honour citizens' entitlement to public information.***
- Principle 6: ***Individual data** collected by statistical agencies for statistical compilation, whether they refer to natural or legal persons, are to be **strictly confidential** and used exclusively for statistical purposes.*

Confidentiality of individual data

From the disseminated information it should not be possible to identify the individual who it refers to.

- **No contradiction with Aarhus Convention:**
 - Data for statistical and administrative purposes should be kept separately
 - Duplication in data collection avoided by use of administrative data for statistical purposes (e.g. most of Environment Statistics is from administrative sources)
- Different solutions used by NSOs to serve specific information requests (but always protecting the privacy of respondents)
- Example: Information about emissions covered by PRTRs

Open Data Strategies at NSOs

- Often stimulated by right to information acts.
- Several countries in UNECE region have open data policies at either their statistical office or for the government as a whole.
- Open data policies are often led by NSOs and national statistical systems are the largest data providers.
- HLG Modernisation Committee on Products and Sources intends to support national statistics offices in implementing an open data strategy.

Data Revolution (≠ Big Data!)

The data revolution is characterised by:

- An explosion in the **volume** of data, **production speed**, **number of producers**, **dissemination**, and the **range of things** on which there is data, coming from **new technologies**
- A **growing demand for data** from all parts of society.

The data revolution for sustainable development is:

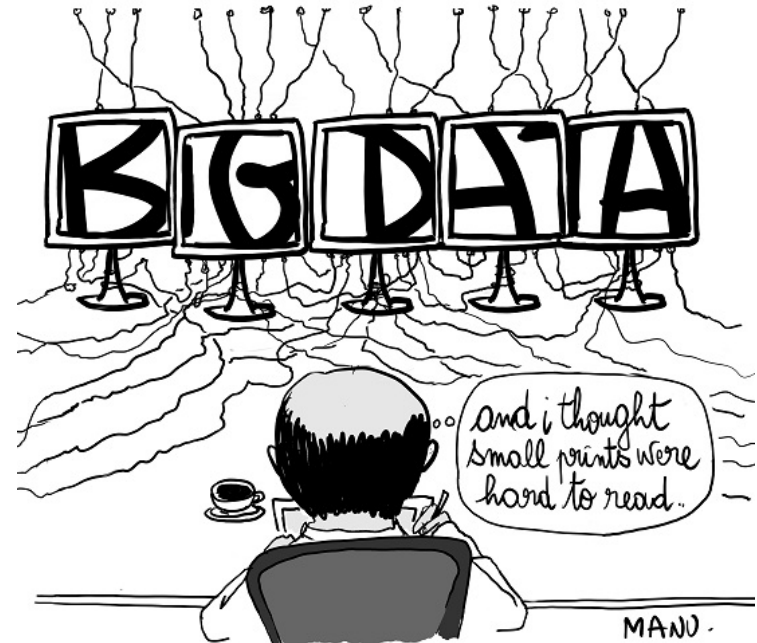
- Integration of these new data with traditional data to produce high-quality **information that is more detailed, timely and relevant** for many purposes and users, especially to foster and monitor sustainable development;
- Increase in the usefulness of data through a much **greater degree of openness and transparency**, avoiding invasion of privacy and abuse of human rights from misuse of data on individuals and groups, and minimising inequality in production, access to and use of data;
- Ultimately, **more empowered people, better policies, better decisions and greater participation and accountability**, leading to better outcomes for people and the planet.

Data Revolution and Official Statistics

- In August 2014 the UN Secretary General appointed the **Independent Expert Advisory Group on the Data Revolution for Sustainable Development**
 - Advise the SG on measures required to close the data gaps and strengthen national statistical capacities.
 - Assess new opportunities linked to innovation, technical progress and the surge of new public and private data providers to support and complement conventional statistical systems and strengthen accountability at the global, regional and national level.
- In UNECE Region: **High-Level Group for the Modernisation of Official Statistics** (under Bureau of CES, since 2010)
 - Standards based production of statistics
 - Increased use of administrative data
 - Potentially use Big Data solutions and non official sources

Big Data: Characterised by the 3 V's

- Data sets of increasing volume, velocity and variety
- Often largely unstructured (no pre-defined data model and/or do not fit well into conventional relational databases)
- Examples:
 - Scanners
 - Satellite or aerial imagery
 - Smart meters
 - Road sensors
 - Mobile phones
 -



- **Challenges for official statistics:**

- Legislative: access to data
- Privacy: managing public trust and acceptance of data re-use
- Financial: potential costs versus benefits
- Management: e.g. policies and directives
- Methodological: data quality and suitability of statistical methods
- Technological: issues related to information technology

Big Data and SDGs

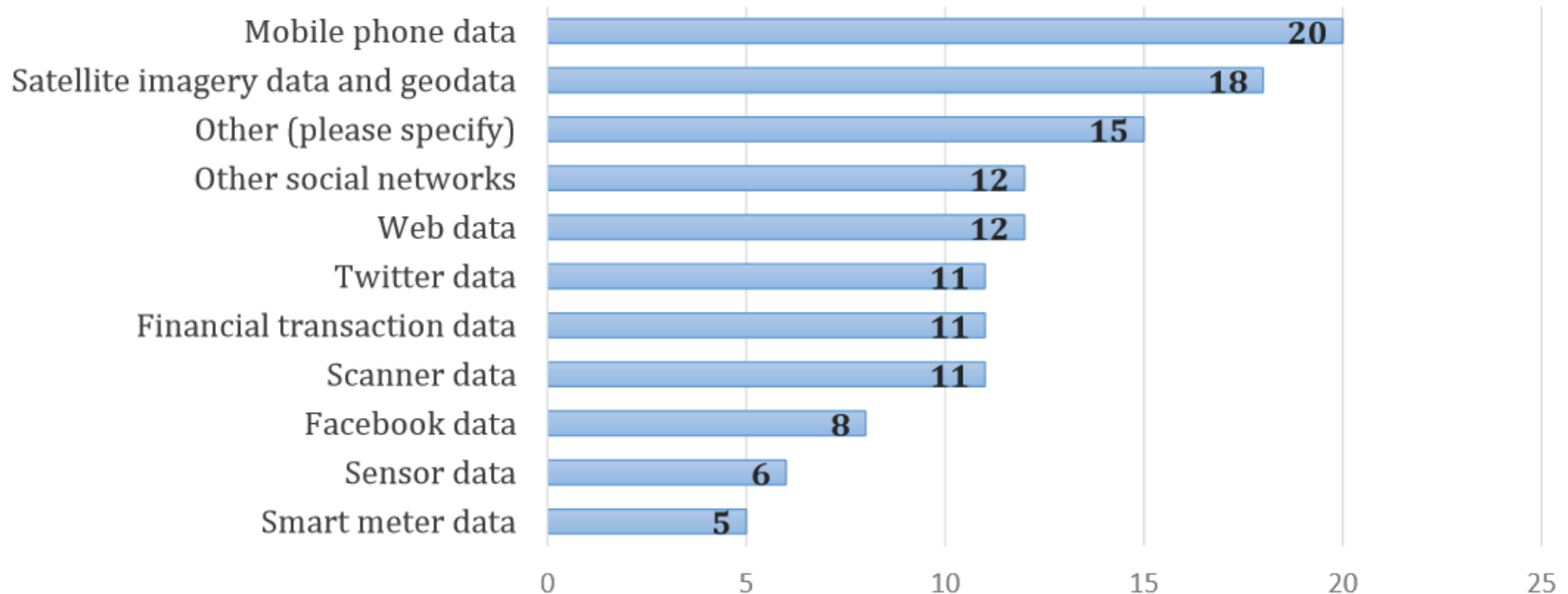
- Statistical Commission created in 2014 the **Global Working Group on Big Data for Official Statistics**
 - Identifying big data as potential source for monitoring SDG
 - Big Data will be used as one of the sources, but will take time to develop methodology and to train staff in all countries
 - UNECEs HLG Sandbox can potentially be used to develop indicators

- Big Data (not yet) source for official statistics nor for SDGs
 - Standards for Official Statistics (regarding coverage, quality, representativeness, confidentiality, sustainability) higher than for producers of non official statistics
 - Further development is needed
 - UNECE supports and facilitates countries in collaboratively getting necessary skills (HLG Big Data project and HLG Big Data Sandbox)

Big Data and SDGs

- Survey carried out by the Big Data Task Team in 2015: 58 responses
- Report available by end of December

Type of data source

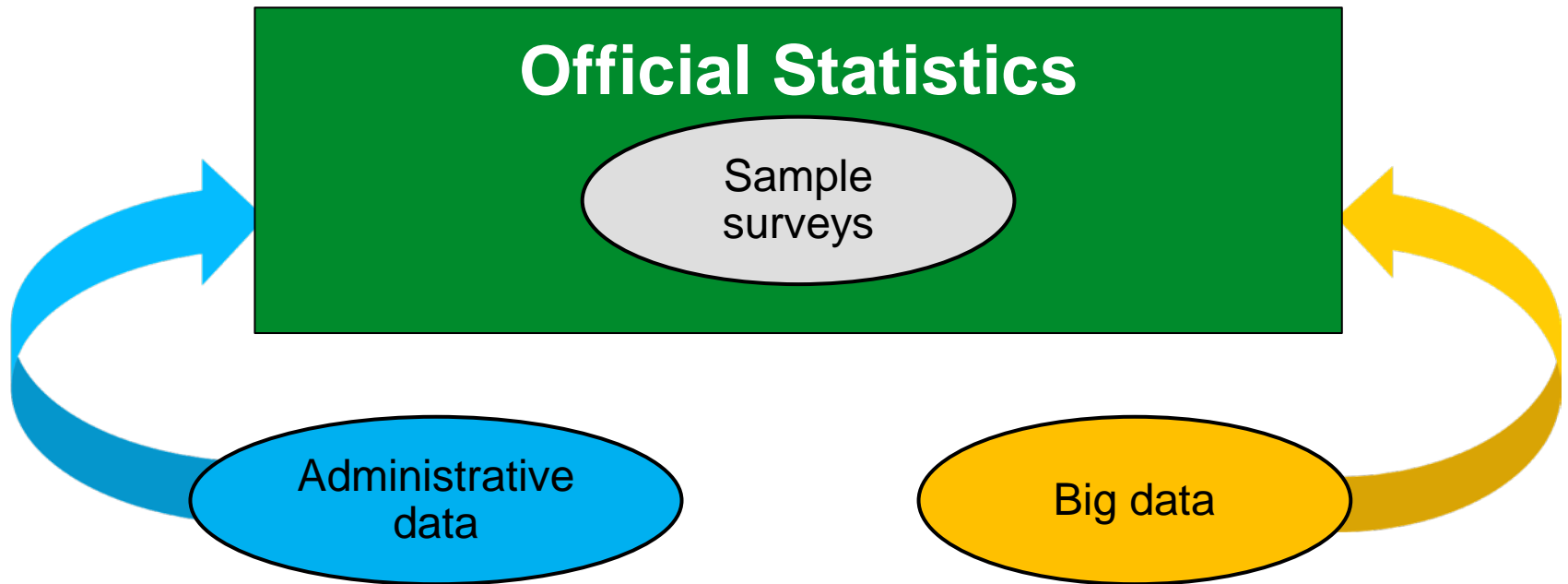


Big Data and SDGs - Inventory

Title	Organization	Sector	Overview	Indicator/topic [potential indicator]	Potential SDG/Target (G.T) [potential Goal/Target]
2. Use of web activity data for the production of flash estimates		General/ Cross-cutting		Unemployment data [other socio-economic indicators]	8 8.5
3. Harvest statistics based on satellite images counts of individuals held on a major commercial marketing database comparison to Census data estimates	UK Office for National Statistics	Agricultural Statistics Demographic and social statistics (including subjective well-being)	Counts of individuals by age band and sex were obtained from the data provider Experian. The counts were based on their commercial marketing database - a foundation of edited electoral roll plus various other data sources including large scale continuous surveys fielded by Experian. The counts were compared with Census data	Agricultural productivity [Food security/availability & location]	2 [14.4]
4. Smart meter type data for household structure/size and occupancy	UK Office for National Statistics	Demographic and social statistics (including subjective well-being)	This is exploratory research, commissioned out to academia, into the potential of electricity smart meter type data to identify household structure and size. A second objective is to research models to see if probability of occupancy by time of day might be derived. Smart meter data will be collected on all households in England by 2020. The minimum specification is energy usage every 30 minutes per meter. Data will be centralised and might be available for research (details/legislation still to be formally agreed). This research is being conducted on data from trials of energy use.	Household structure/types, occupancy by time	
5. Smart meter data potential for detecting unoccupied dwellings	UK Office for National Statistics	Demographic and social statistics (including	Very much exploratory research. ONS has acquired electricity smart meter data from trials of energy usage. This data has various potential uses within official statistics - the focus for our work is currently on occupancy. Another objective is to familiarise ONS with the	See 11 [energy use/efficiency if combined with housing	7.3

Challenges of using administrative data for statistical purposes

- Traditionally two main sources for statistics:
 - Sample surveys
 - Administrative data sources (including registers)
 - (Use of Big Data as a potential future source)



- E.g. environment statistics uses lots of administrative data

Use of administrative sources – common problems and possible solutions

- One-way flow of sensitive data
 - Code of practice following international standards needed
- Administrative units may be different to statistical units
 - Conversion needed
- Different classifications uses
 - Tools are needed to convert codes
 - Stress advantages of using a common classification
 - Give early notice of classification changes and implement them across government
- Missing data
 - Impute where possible
- Timeliness
 - Look for ways to reduce lags where possible
- Different time periods
 - Statistical corrections or estimations
- Data from different sources may not agree
 - Establish priority rules
- Sudden changes
 - Legal provisions or contractual agreements; regular contacts, anticipation of changes etc.

UN initiative on Global Geospatial Information Management (UN-GGIM)



5th Session of the **UN Committee of Experts on UN-GGIM** 3-7 August 2015 in New York:

- Focus on finance, governance, common principles, policies, methods, mechanisms and standards
- Aiming to finalize guidelines to assist Member States in implementing and adopting international geospatial standards and best practices
- Work closely related to SDG and post-2015 agenda

2nd Plenary of **UN-GGIM:Europe** 7 October 2015 in Belgrade; approval of work plan 2015-2018:

• **Core Data**

- 1. Specifications of core data (End of 2016)
- 2. Economic model for production & distribution of core data (End 2017)
- 3. Existing political & financial frameworks supporting core data availability (Mid-2018)

• **Data Integration**

- 1. Definition of the priority user needs for data combinations (accomplished)
- 2. Recommendation for implementing prioritized combinations of data (Mid-2016)
- 3. Recommendation how to manage side-effects induced by data combinations (Mid-2016)

• **Tasks for the Secretariat**

- 1. Cooperation with UN-GGIM, Regional Committees & geospatial community (On-going)
- 2. Compile and maintain knowledge base of the European NMCAs and NSIs (End 2015)
- 3. Compile and maintain knowledge base of the European NMCAs and NSIs (End 2017)

More information available on <http://ggim.un.org>

Conclusions

- **Official Statistics play a central role in measuring SDGs** and the dissemination and interpretation of SDG indicators.
- **Serving the government, the economy and the public with data** about the economic, demographic, social and environmental situation **AND maintain statistical confidentiality** are fundamental for official statistics.
- **Data revolution is a great opportunity with lots of challenges** to improve quantity, quality and the speed of production of official statistics. The UN Secretary General appointed an independent expert advisory group.
- **Big data are being explored** as potential data source for some SDG indicators by the Global Working Group (GWG) on Big Data for Official Statistics and the UNECE High-Level Group for the Modernisation of Official Statistics
- **UNECE High-Level Group for the Modernisation of Official Statistics** also explores the increased use of administrative data
- **UN-GGIM** aims at adoption and implementation of standards by the global geospatial information community and supporting SDGs