



Breaking through the Big Data Barrier



Steve MacFeely
Head of Statistics and Information

Q1: What is the role of unofficial data/statistics ... in compiling official statistics and SDGs?



Q2: How can we work together to strengthen connections between citizen science initiatives and UN offices, beyond the odd example?

I am particularly interested in exploring how the groups can support each other and if there is a role for the UN to supporting increase of project scalability if methods are valid for indicators.



Q3: International Organizations and NGOs mostly use national surveys in their studies. Those surveys are only produced every 4 years, they imply high costs, and the standardization is difficult across countries since different methodologies are used. Can big data or data gathered through social media platforms be a better alternative in producing faster results, at lower costs and using similar data collection methods across countries?



Q4: Should there be Fundamental Principles for big data signed by all relevant stakeholders including the private sector and a forum to discuss big data governance that goes beyond inter-governmental institutions?

Q5: How is the involvement and regulation of private companies providing the data analysis infrastructures envisioned? What kinds of ethical and data privacy issues are relevant to consider?

Q6: Developing big data learning, international collaboration and private sectors, conflicting with the protection of privacy and confidentiality - suggestions and thoughts of mitigating....

Q7: How can we generate big data without encroaching on privacy issues for ongoing and future pandemics management?

We urgently need a Global Data Convention. Here's why. WEF Blog, Nov 2020.

<https://www.weforum.org/agenda/2020/11/global-data-convention/>

A call for a Global Data Convention. WSD Blog, October 2020.

<https://worldstatisticsday.org/2020/blog/Global-Data-Convention.html>



Q9: What are the most likely data sources to contribute to measuring SDGs? (Part A)

Table 1. Big data sources and project topics registered by National and International Organisations on the UN big data project inventory

Data Source	National	International	Project topic	National	International
Web scraping	22	4	Prices	22	4
Scanner	20	1	Population/migration	10	4
Mobile phone/CDR	14	18	Transport/mobility	9	11
Social media	8	23	Geographical/spatial	8	7
Satellite imagery	6	7	Labour market	7	2
Smart meter	5	1	Agriculture/Land use	6	4
Credit card	3	1	Tourism	5	1
Road sensor	5	–	Health/disease	4	7
Health records	5	2	Energy/Environment	4	6
Ship identification	2	–	Crime/Corruption	2	4
Criminal records	1	2	Poverty/inequality	1	9
Other	20	31	Disaster risk reduction	–	8
			Other	31	24
Total	111	90	Total	109	91

Source: Authors own calculations derived from UN Big Data Project Inventory <https://unstats.un.org/bigdata/inventory/> [examined on 27 April, 2018]¹³.

Taken from: MacFeely, S. (2019). The Big (data) Bang: Opportunities and Challenges for Compiling SDG Indicators. *Global Policy*. Vol. 10, No.1, pp.121 – 133.

<https://onlinelibrary.wiley.com/doi/epdf/10.1111/1758-5899.12595>



Some UNCTAD examples - Nowcasting

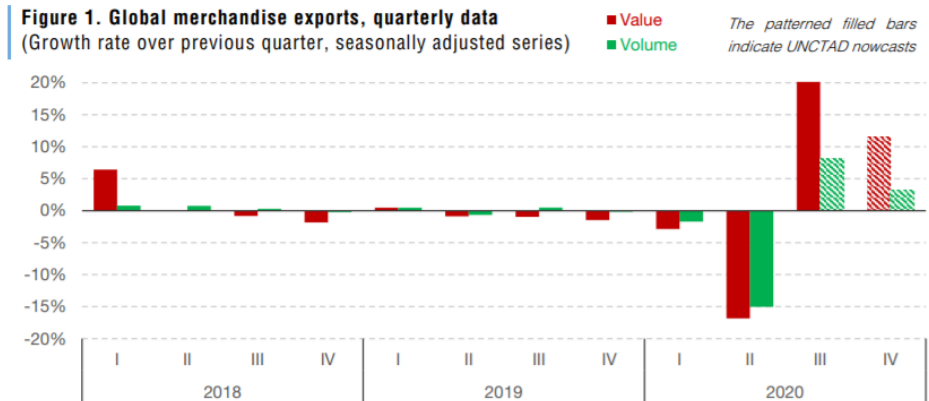
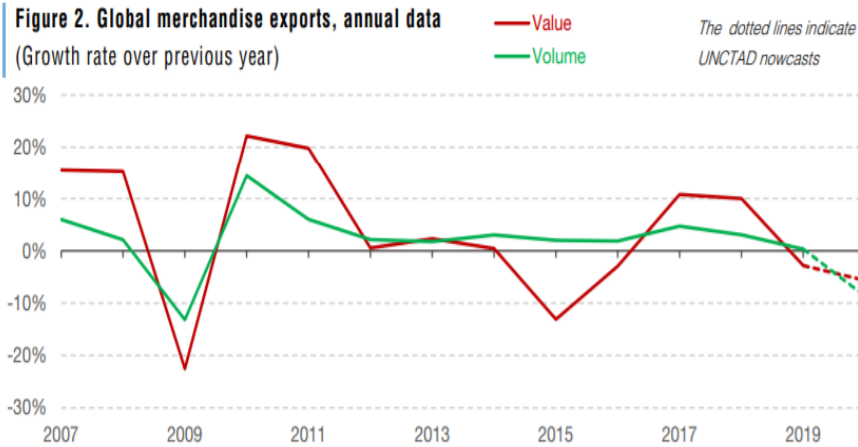
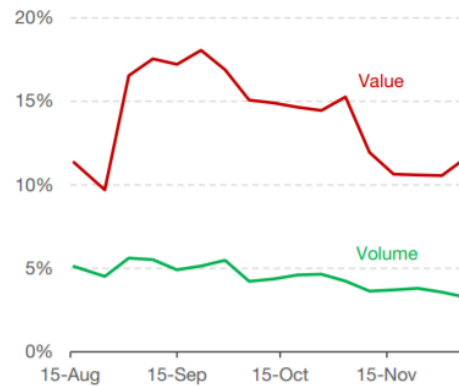


Figure 3. Evolution of global merchandise trade nowcast, 4th quarter of 2020
(Growth rate over previous quarter, seasonally adjusted series)



UNCTAD Global Trade Nowcast
<https://unctad.org/official-documents/search?f%5B0%5D=product%3A1419>



Some UNCTAD examples

AIS – automatic identification system



Port Performance Statistics
Port Liner Connectivity Index

<https://unctadstat.unctad.org/wds/ReportFolders/reportFolders.aspx>
<https://unctadstat.unctad.org/wds/ReportFolders/reportFolders.aspx>

Some UNCTAD (and UNEP) examples SDG 12.6.1 – Sustainability Reporting

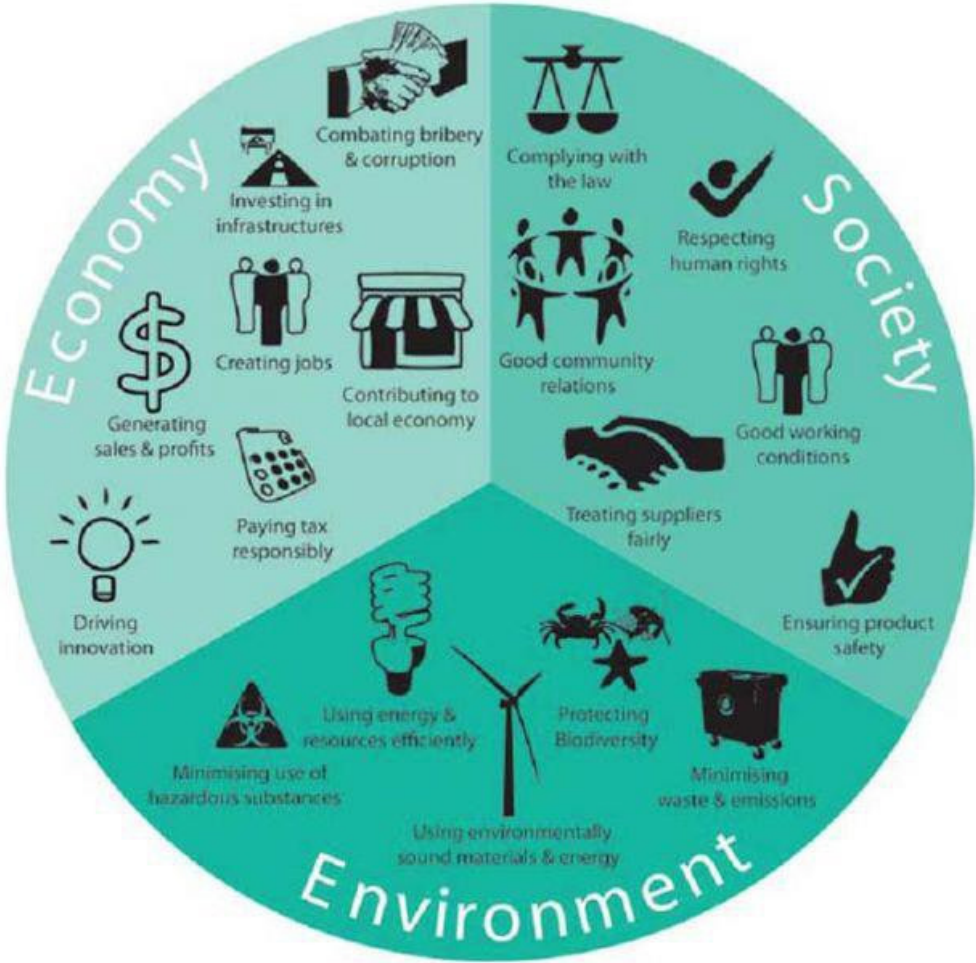
Working with *Global Algorithmic Institute*

Using advanced Natural Language Processing (NLP) and Artificial Intelligence (AI) analytics

Preliminary results but a number of methodological issues to be resolved



One final thought...



Some further discussion

MacFeely, S. (2020). Measuring the Sustainable Development Goal Indicators: An Unprecedented Statistical Challenge. *Journal of Official Statistics*. Vol. 36, No.2, pp. 361 – 378.

<https://content.sciendo.com/view/journals/jos/36/2/article-p361.xml?language=en>

MacFeely, S. (2020). In search of the Data Revolution: has the official statistics paradigm shifted? *Statistical Journal of the International Association of Official Statistics*, Vol. 36, No. 4, pp. 1075 – 1094.

https://www.researchgate.net/publication/346945320_In_search_of_the_data_revolution_Has_the_official_statistics_paradigm_shifted

Nastav, B. and MacFeely, S. (2020). You say you want a (data) revolution: reflections one year on. *Statistical Journal of the International Association of Official Statistics*, Vol. 36, No. 4, pp. 1299 – 1306.

<https://content.iospress.com/download/statistical-journal-of-the-iaos/sji200722?id=statistical-journal-of-the-iaos%2Fsji200722>

MacFeely, S. and Nastav, B. (2019). You say you want a (data) revolution? A proposal to use unofficial statistics for the SDG Global Indicator Framework. *Statistical Journal of the International Association of Official Statistics*, Vol. 35, No. 3, pp. 309 – 327.

<https://content.iospress.com/download/statistical-journal-of-the-iaos/sji180486?id=statistical-journal-of-the-iaos/sji180486>

MacFeely, S. (2019). The Big (data) Bang: Opportunities and Challenges for Compiling SDG Indicators. *Global Policy*. Vol. 10, No.1, pp.121 – 133.

<https://onlinelibrary.wiley.com/doi/full/10.1111/1758-5899.12595>



Thank you

steve.macfeely@un.org

