Big Data for Official Statistics

Herman Smith

United Nations Statistics Division

I. Official Statistics

II. Big Data

III. Conclusions

United Nations Statistical Commission

- **Functional Body under ECOSOC**
- **Chief Statisticians of all UN Member States** and international organizations (WHO, UNICEF, ILO, UNESCO, World Bank, IMF)
- **Agree on Statistical Standards, best practices in the compilation of statistics, and capacity building**

- **Fundamental Principles of Official Statistics**
  - ✓ GA Resolution 68/261 – 19 January 2014
- **Indicator Framework for the Monitoring of Progress on SDGs**
  - ✓ GA Resolution 71/313 – 6 July 2017

SDGs: 17 Goals and 169 Targets

Fundamental Principles

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.

Fundamental Principles

Principle 5.

Data for statistical purposes may be drawn from all types of sources, be they statistical surveys or administrative records. Statistical agencies are to choose the source with regard to quality, timeliness, costs and the burden on respondents.

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**.

Why are Big Data important?

- ✓ Big Data can keep official statistics **relevant** – private sector moves fast
- ✓ Big Data are part of **modernization of statistical systems** – new production processes and partnerships
- ✓ Big Data can help meeting the **data demand of the 2030 agenda** – monitoring policies – "leave no one behind"
- ✓ Big Data are needed for **agile statistics** – for emergency issues

Classification of Big Data sources

Social Networks (Human-sourced information)

Mostly digitized data, loosely structured and often ungoverned

- **Social Networks: Facebook, Twitter, Tumblr etc.**
- **Blogs and comments**
- **Personal documents**
- **Pictures: Instagram, Flickr, Picasa etc.**
- **Videos: You-tube etc.**
- **Internet searches**
- **Mobile data content: text messages**
- **User-generated maps**
- **E-Mail**

Classification of Big Data sources

Business Systems (Process-mediated data)

Recording and monitoring of business events; highly structured

- **Data produced by Public Agencies**
  - Medical records
- **Data produced by businesses**
  - Commercial transactions
  - Banking/stock records
  - E-commerce
  - Credit cards

Classification of Big Data sources

Internet of Things (Machine-generated data)

Sensors and machines used to measure and record events and situations in the physical world

- Fixed sensors
  - ✓ **Home automation; Weather/ pollution sensors; Traffic sensors / webcam; Scientific sensors; Security / Surveillance cameras**
- Mobile sensors
  - ✓ **Mobile Phone Location; Cars; Satellite images**
- Data from computer systems
  - ✓ **Logs / Web logs**

GWG on Big Data for official statistics

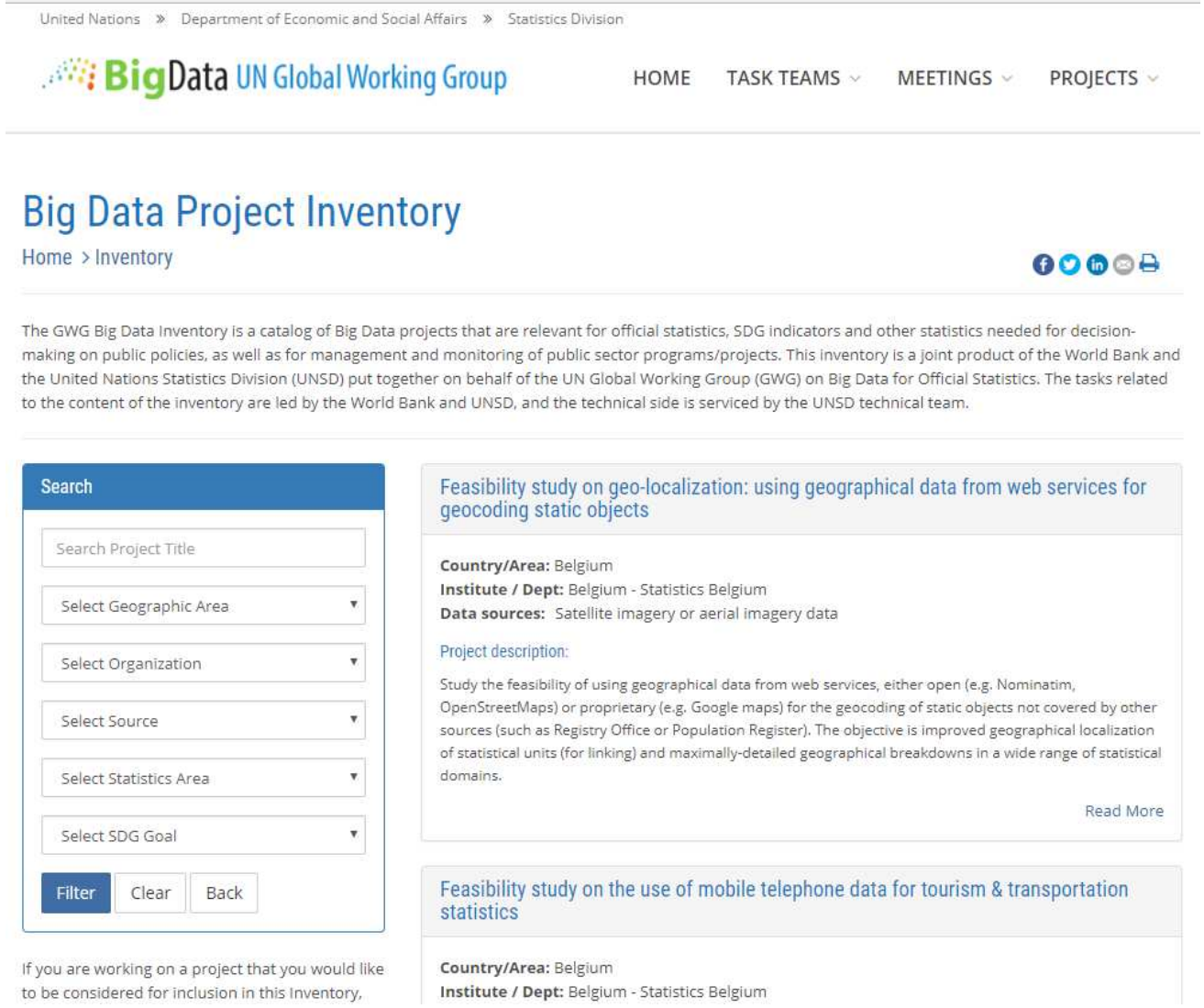
https://unstats.un.org/bigdata/

UN Global Working Group (GWG) on Big Data for Official Statistics

- Created in March 2014 by the UN Statistical Commission
- Mandated to give direction to the use of Big Data for Official Statistics (on issue related to methodology, quality, technology, data access, legislation, privacy, management and finance, etc.)
- Consisting of 28 countries and 16 international organizations

Tasks teams

- ✓ Big Data and the Sustainable Development Goals
- ✓ Mobile Phone Data
- ✓ Satellite Imagery and Geo-Spatial Data
- ✓ Scanner Data
- ✓ Social Media Data
- ✓ Training, Skills and Capacity-building
- ✓ Committee on Global Platform for Data, Services and Applications


United Nations » Department of Economic and Social Affairs » Statistics Division
BigData UN Global Working Group
HOME
TASK TEAMS
MEETINGS
PROJECTS
Big Data Project Inventory
Home > Inventory
The GWG Big Data Inventory is a catalog of Big Data projects that are relevant for official statistics, SDG indicators and other statistics needed for decision-making on public policies, as well as for management and monitoring of public sector programs/projects. This inventory is a joint product of the World Bank and the United Nations Statistics Division (UNSD) put together on behalf of the UN Global Working Group (GWG) on Big Data for Official Statistics. The tasks related to the content of the inventory are led by the World Bank and UNSD, and the technical side is serviced by the UNSD technical team.
Search
Search Project Title
Select Geographic Area
Select Organization
Select Source
Select Statistics Area
Select SDG Goal
Filter
Clear
Back
If you are working on a project that you would like to be considered for inclusion in this Inventory,
Feasibility study on geo-localization: using geographical data from web services for geocoding static objects
Country/Area: Belgium
Institute / Dept: Belgium - Statistics Belgium
Data sources: Satellite imagery or aerial imagery data
Project description:
Study the feasibility of using geographical data from web services, either open (e.g. Nominatim, OpenStreetMaps) or proprietary (e.g. Google maps) for the geocoding of static objects not covered by other sources (such as Registry Office or Population Register). The objective is improved geographical localization of statistical units (for linking) and maximally-detailed geographical breakdowns in a wide range of statistical domains.
Read More
Feasibility study on the use of mobile telephone data for tourism & transportation statistics
Country/Area: Belgium
Institute / Dept: Belgium - Statistics Belgium

Global platform for data, services and applications

- The GWG is facilitating the development of the UN Global Platform with involvement of large tech companies
- Building on the best practices of public and private Big Data initiatives, and offering the technology infrastructure and a network for data innovation to the official statistical community,

Partnerships with private sector companies

- **Google** – Earth Observation data for measuring Fresh Water Extent (SDG 6)
- **Microsoft** – Earth Observation data for estimating Agricultural Crop production
- **Nielsen** – Price data for calculating the Consumer Price Index
- **Twitter/ Facebook** – Consumer Sentiment Index

Example of use of big data in economic statistics/

- Construction of a social media index based on the sentiment analysis of social media messages (facebook, twitter, etc.)
- Credit cards data for the estimation of consumer expenditures
- use of mobile telephone data for tourism & transportation statistics
- Use of scanner data for consumer prices
- …

Final remarks

- **Countries are exploring Big data sources for the compilation of official statistics. Important to share experiences.**
- **The use of alternative and Big data source combined with the latest technology can be done and will indeed improve the quality and relevance of official statistics.**
- **Privacy and Confidentiality are important issues. Collaborating on sensitive information is difficult, but not impossible.**
- **Global platforms facilitate the collaboration among many players from the statistical community, academia, research and private sector.**

Thank you