Big Data and Official Statistics
The UN Global Working Group

Dr. Ronald Jansen
Chief, International Trade Statistics
United Nations Statistics Division
jansen1@un.org
Overview

What is Big Data?

What is Big Data for Official Statistics?

What are the Challenges?

UN Global Working Group on Big Data for Official Statistics
What is Big Data?

Big Data characteristics
- Highly distributed
- Loosely structured
- Large in volume

Big Data benefits
- Automatically generated data
- Timely data
- Potentially relevant

Big Data sources:
- Mobile Devices
- Digital Transactions
- Sensors
- Social Networking
What is Big Data?

Processing and Analyzing of Big Data sources (IT concerns):

- Parallel processing techniques (HADOOP)
- Non-relational data storage capabilities
- Advanced analytics and data visualization technology
Big Data for Development

UN Global Pulse
- Early warning
- Real-time awareness
- Real-time feedback
Big Data for Development

UN GLOBAL PULSE

- Research & Development
- Big Data Partnerships
- Pulse Lab Network
  - New York
  - Jakarta
  - Kampala
What is Big Data for Official Statistics?

**Big Data sources / Statistics**

- Mobile Phone Positioning Data / Tourism statistics
- **Supermarket scanner data** / Price statistics
- **Vehicle tracking device** / Transport statistics
- **Twitter** / Consumer confidence statistics
- Satellite imagery / Agriculture statistics
What are the Challenges?

Big Data Challenges

• Methodology
• Privacy
• IT infrastructure
• Human skills
• Partnerships
Methodological Challenges

Methodology

- Representativeness ?
- Volatility (no time series?)
- Standardization ?
- Modelling
Big Data partnerships opportunities

- National Statistical Offices
  - Standards and methodology
- Academia and scientific communities
  - Analytical capacity and modelling
- Big Data providers
- Data Sources
- Research institutes and technology providers
  - Upstream processing
  - IT infrastructure,
Partnership arrangements

- Financial commitments
- Legislative issues
- Privacy and confidentiality
- Responsibilities and ownership
- IT infrastructure/technical platforms

Three examples of partnership arrangements for Big Data based on mobile phone positioning data
Mobile Phone Positioning Data

- Call Detail Record (CDR)
- Data Detail Record (DDR) (Internet Protocol Detail Record)
- Location Updates (LA)
- Radio coverage updates (Abis data)
Example 1: Estonia

- Mobile phone operator
- Provides raw data to research institute

Shared responsibility for data processing and analysis

Research institute
- Provides IT infrastructure
- Preparation of data
- Ensuring privacy and confidentiality

Academia
- Providing modeling and analytics

National Statistical Office
- Receives pre-processed and aggregated data
- Prepares and disseminates statistics
Example 2: Netherlands

Responsibility for data

Data source/provider:
- Vodafone
- Makes data available to research institute

Research institute:
- Intermediate commercial group
- Provides IT platform for data filtering and privacy

National Statistical Office:
- Receives pre-processed Big Data
- Analyze, prepare and disseminate statistics
Example 3: Belgium

Shared responsibility for preparation and analysis of Big data

Data provider/source
- Belgacom
- Provides IT platform
- Processes data

National Statistical Office
- Prepares and disseminates statistics
Example 4: Tanzania

Data provider/source

- Orange?
- Provides IT platform?

Regional Office?

- Processes data

National Statistical Office

- Prepares and disseminates statistics
UN Global Working Group on Big Data

Mandate

Based on Decision 45/110, the mandate of the working group is formulated as follows:

1. **Provide strategic vision**, direction and coordination of a global programme on Big Data for official statistics;

2. **Promote practical use of sources of Big Data** for official statistics, while building on the existing precedents in Big Data and finding solutions for
   i. **Methodological** issues, covering quality concerns and data analytics;
   ii. **Legal** issues of access to data sources;
   iii. **Privacy** issues, in particular those relevant to the use and reuse of data, data linking and re-identification
   iv. **Security** and management of data, including assessment of cloud computing and storage;
   v. **IT/technological** expertise, including cost benefit analysis
UN Global Working Group on Big Data

Mandate
(continue)

3. Promote capacity building, training and sharing of experience;

4. Foster communication and advocacy of use of Big Data for policy applications;

5. Build public trust in the use of private sector Big Data for official statistics including issues of confidentiality and privacy.
## Composition of the group

<table>
<thead>
<tr>
<th>Developed countries</th>
<th>Developing countries</th>
<th>International Organizations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>Bangladesh</td>
<td>ITU</td>
</tr>
<tr>
<td>Denmark</td>
<td>China</td>
<td>OECD</td>
</tr>
<tr>
<td>Italy</td>
<td>Colombia</td>
<td>UN Global Pulse</td>
</tr>
<tr>
<td>Mexico</td>
<td>Morocco</td>
<td>UNECE</td>
</tr>
<tr>
<td>Netherlands</td>
<td>Philippines</td>
<td>UNESCAP</td>
</tr>
<tr>
<td>USA</td>
<td>Tanzania</td>
<td>UNSD</td>
</tr>
<tr>
<td></td>
<td></td>
<td>World Bank</td>
</tr>
</tbody>
</table>
## Deliverables

<table>
<thead>
<tr>
<th>Guidelines / Handbooks</th>
<th>Big Data examples</th>
<th>Pilot Projects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methodology</td>
<td>* Mobile Phone Positioning Data</td>
<td></td>
</tr>
<tr>
<td>Privacy</td>
<td>* Scanner data</td>
<td></td>
</tr>
<tr>
<td>IT infrastructure</td>
<td>* Tracking device</td>
<td></td>
</tr>
<tr>
<td>Human skills</td>
<td>* Twitter</td>
<td></td>
</tr>
<tr>
<td>Partnerships</td>
<td>* Satellite imagery</td>
<td></td>
</tr>
</tbody>
</table>
Thank you

Ronald Jansen
United Nations Statistics Division
jansen1@un.org