Using **Big Data** for the **Sustainable Development Goals**

Presented by: Amparo Ballivian
Objective:

To provide concrete examples of the use of Big Data for monitoring the indicators associated with the Sustainable Development Goals.
Task Team Members

- **Chaired by The World Bank and INEGI**
- 7 international agencies and companies
  - WEF, Orange, ODI, Data-Pop Alliance, NASA, Paris 21, Positium
- 6 United Nations agencies
  - UNSD, UNECE, UNESCAP, ITU, Global Pulse, UN Department of Economic and Social Affairs
- 3 universities
  - University of Pennsylvania, MIT, Harvard
- Colombia’s National Administrative Department of Statistics
Plan of Actions:

1. Survey to identify which of the 169 SDG targets could use Big Data, as well as proposals of Big Data-specific indicators related to the SDG targets (which may be different to the current set of indicators based on traditional sources of data).

2. Make an inventory of past and ongoing research work on Big Data and identify those that could be used to calculate one or more SDG targets.

3. Pilot research in 1-2 countries on calculating 2-3 SDG indicators using Big Data.

4. Presentation at the Big Data Conference of UAE.

5. Write report of the Working Group.
Summary status:

1. Research survey of Big Data Initiatives for Sustainable Development Goals (SDGs) conducted and results analysed
2. Consolidated inventory of Big Data projects has been created; linking each project to an SDG target is ongoing
3. Pilot projects: no financing for new projects, selection of projects for task team report ongoing
4. Presentation at Abu Dhabi conference: done
5. Report of the Task Team: Due in December 2015
1. **Survey of SDG-related Big Data projects**

**Purposes of the survey**

- Identify characteristics of Big Data projects that can be used to monitor and achieve the SDGs
- Learn about SDG areas, data sources, partners and objectives of the Big Data projects
- Understand scope for project replication elsewhere
- Assess the feasibility of proposed projects
1. Survey of SDG-related Big Data projects

Sample and representativeness of responses

Survey population:

- World Bank (47)
- UNECE Big Data Project (52)
- Heads of Statistical Departments of IOs (46)
- Heads of International Relations Departments (74)
- DGs of NSOs (141)

Responses by Role:

- 74.14% NSO
- 12.07% Academia
- 10.34% IO
- 3.45% Private Sector

58 Responses out of 360 Surveyed
1. Survey of SDG-related Big Data projects

Sample Representativeness

**Representation of NSOs by Income Group**

- 90% of High Income Countries replied, from which 50% had at least one big data for SDG project and the rest did not.
- Response rate of other countries is lower but HIC is where innovation is to happen anyway.
1. Survey of SDG-related Big Data projects

World map of countries with reported projects
1. Survey of SDG-related Big Data projects

Descriptives: Frequency of SDG targets

- Affordable and Clean Energy
- Clean Water and Sanitation
- Climate Action
- Decent Work and Economic Growth
- Gender Equality
- Good Health and Well-Being
- Industry, Innovation, and Infrastructure
- Life below Water
- Life on Land
- No Poverty
- Partnerships for the Goals
- Peace, Justice, and Strong Institutions
- Quality Education
- Reduced Inequalities
- Responsible Consumption and Production
- Sustainable Cities and Communities
- Zero Hunger
1. Survey of SDG-related Big Data projects

<table>
<thead>
<tr>
<th>Type of data source</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mobile phone data</td>
<td>20</td>
</tr>
<tr>
<td>Satellite imagery data and geodata</td>
<td>18</td>
</tr>
<tr>
<td>Other (please specify)</td>
<td>15</td>
</tr>
<tr>
<td>Other social networks</td>
<td>12</td>
</tr>
<tr>
<td>Web data</td>
<td>12</td>
</tr>
<tr>
<td>Twitter data</td>
<td>11</td>
</tr>
<tr>
<td>Financial transaction data</td>
<td>11</td>
</tr>
<tr>
<td>Scanner data</td>
<td>11</td>
</tr>
<tr>
<td>Facebook data</td>
<td>8</td>
</tr>
<tr>
<td>Sensor data</td>
<td>6</td>
</tr>
<tr>
<td>Smart meter data</td>
<td>5</td>
</tr>
</tbody>
</table>

- Mobile phones (20), satellite imagery (18) and social media (11+8) are the most prominent sources
- Otherwise, wide range of sources
1. **Survey of SDG-related Big Data projects**

<table>
<thead>
<tr>
<th>Partners</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Government agency</td>
<td>28</td>
</tr>
<tr>
<td>Academic/Research institute</td>
<td>17</td>
</tr>
<tr>
<td>Satellite imagery provider</td>
<td>13</td>
</tr>
<tr>
<td>Mobile Phone operator</td>
<td>12</td>
</tr>
<tr>
<td>Other (please specify)</td>
<td>11</td>
</tr>
<tr>
<td>IT services company</td>
<td>7</td>
</tr>
<tr>
<td>Intermediary Big Data provider</td>
<td>7</td>
</tr>
<tr>
<td>Cloud server provider</td>
<td>3</td>
</tr>
</tbody>
</table>

- Most projects partner with the public sector
- Among private sector partnerships, mobile phone operators and IT services dominate
1. Survey of SDG-related Big Data projects

Descriptives: Big Data projects’ objectives

- The main objectives are improvements in data frequency and disaggregation
- This is consistent with the demands of the SDGs
2. Inventory

- We have a consolidated inventory of projects from: UNECE, Global Pulse, World Bank, Sandbox and Positium.

- We have started identifying projects that are linked to one or more of the SDG targets. Work is concluded for UNECE projects, but needs to be completed for other projects.

- Next steps:
  - Conclude the mapping of projects to SDG targets
  - Keep including other projects into the consolidated inventory
  - Coordinate/join forces with the inventory proposed by the TT on Cross-Cutting issues
## A. UNECE

<table>
<thead>
<tr>
<th>Title</th>
<th>Organization</th>
<th>Sector</th>
<th>Overview</th>
<th>Indicator/topic</th>
<th>Potential SDG/Target (G.T)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Developing a Curriculum and Training Modules on Using Big Data for Official Statistics</td>
<td>ESCAP</td>
<td>General/ Cross-cutting</td>
<td>To use &quot;big data&quot; in producing official statistics, statisticians and managers in national and local statistical systems need to closely examine and understand the potential use of various types of big data as well as the issues and limitations of their usage for producing official statistics and indicators. Statisticians and managers also need to gain and/or improve their knowledge and skills for working with such data and integrating such work in the standard statistical business processes. This project aims to develop a training curriculum to address capacity-building requirements on understanding and assessing the potentials for utilizing big data for official statistics, particularly in developing statistical systems of Asia and the Pacific, based on improvement of knowledge and skills level of the officials.</td>
<td>NA</td>
<td>[potential indicator]</td>
</tr>
<tr>
<td>Project Title</td>
<td>Description</td>
<td>Source</td>
<td>Key Outcome(s)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------</td>
<td>-------------</td>
<td>--------</td>
<td>----------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Automatization data recovering and integration</td>
<td>General/ Cross-cutting</td>
<td>Unemployment data [other socio-economic indicators]</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of web activity data for the production of flash estimates</td>
<td>General/ Cross-cutting</td>
<td></td>
<td>8.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Harvest statistics based on satellite images</td>
<td>Agricultural Statistics</td>
<td>Agricultural productivity [Food security/availability &amp; location]</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Demographic and social statistics (including subjective wellbeing)</td>
<td></td>
<td>[14.4]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smart meter type data for household structure/size and occupancy</td>
<td>Demographic and social statistics (including subjective wellbeing)</td>
<td>Household structure/types, occupancy by time</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smart meter data potential for detecting unoccupied dwellings</td>
<td>Demographic and social statistics (including</td>
<td>See 11 [energy use/efficiency if combined with housing</td>
<td>7.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3. Pilot Projects

- We intended to do new pilot research in 1-2 countries on calculating 2-3 SDG indicators using Big Data. The funding requests did not materialize.
- As an alternative, we have selected 10 big data projects that are closely related to SDG targets (out of which 2 are new pilots to be started by Postium in Estonia).
- These projects were selected based on their close links to SDGs and are either completed or will have substantive results that can reported by December 2015.
- Next steps:
  - Collect details for each of the pilots to be included in the report.
  - Consult with other TTs to determine if they have identified other pilots that could be included in our report.