Collaborative Data Innovation Fund

Alla Morrison
World Bank Group

4th International Conference on Big Data for Official Statistics Bogota, Colombia
November 8, 2017
Agenda

• Collaborative Data Innovation
• Data Innovation Fund Overview
• Insights from Rounds One & Two
• Looking ahead
Collaborative Data Innovation

Need for financing scalable and replicable innovation in technology and approaches, such as partnerships, to improve data on the ground.

Strong emphasis on knowledge-sharing.
Collaborative Data Innovation

Two tiers of collaboration, at program and activities levels:

I. World Bank and the Secretariat of the Global Partnership for Sustainable Development Data (GPSDD)

II. Every funded activity – collaboration among public agencies, private companies, local communities, academia, NGOs, etc.
Data Innovation Fund Overview

• $5 million allocated by the World Bank’s pilot window of TFSCB in 2016-17
• Over 600 proposals received in two rounds
• 15 projects now in early stages in 20 countries across Africa, the Middle East and Asia
• 10+ more will be selected by January 2018

• Examples of projects:
  • Improving vital registration of Syrian refugees in Lebanon
  • Helping health workers predict patient behavior in Africa
  • Using low-orbit satellites to detect illegal fishing in Southeast Asia
Round One - 2016

- Call for Proposals in August 2016: Data Innovation for SDGs
- Over 400 proposals received
- 372 proposals reviewed requested over $70 million and offered over $45 million in matching funds
- 15 projects in mostly early stages in 20 countries across Africa, the Middle East and Asia
Round One: Regions

What are the most cited regions?
Each bar shows how many times a specific region was cited across proposals

- Sub-Saharan Africa: 244
- East Asia & Pacific: 79
- Latin America & Caribbean: 74
- South Asia: 62
- Europe & Central Asia: 32
- Middle East & North Africa: 16
- North America: 5
Round One: Target SDGs

What are the most cited Sustainable Development Goals?
Each bar shows how many proposals mentioned a specific SDG

17 - Partnerships for the Goals
13 - Climate Action
3 - Good Health and Well-being
16 - Peace, Justice and Strong Institutions
15 - Life on Land
11 - Sustainable Cities and Communities
1 - No Poverty
2 - Zero Hunger
4 - Quality Education
6 - Clean Water and Sanitation
5 - Gender Equality
9 - Industry, Innovation and Infrastructure
8 - Decent Work and Economic Growth
12 - Responsible Consumption and Production
10 - Reduced Inequalities
14 - Life Below Water
7 - Affordable and Clean Energy

91 - 81 - 79 - 75 - 75 - 75 - 75 - 73 - 70 - 68 - 67 - 63 - 56 - 44 - 34 - 28 - 26
Round One: Collaborators

What are the most cited collaborators?
Each bar shows how many proposals cited a specific collaborator

- Government: 191
- Private sector: 134
- Academic & research institutions: 127
- Local partnerships: 118
- Civil society: 108
- UN agency: 104
- Sector experts: 95
- NGOs: 95
- Stakeholders: 73
- National statistics office: 70
- Multilateral Development Aid Agencies: 51
- Foundations: 30
- National Development Aid Agencies: 20
- Press and Media: 17
Round One: Data Products

What are the most common data products?
Each bar shows how many proposals mentioned a specific data product

- Web platform: 149
- Database: 125
- Report: 93
- Methodology: 61
- Data visualization: 52
- Mobile app: 51
- Capacity building: 25
- Survey / Data collection: 20
- Software: 16
- Manual: 10
- Algorithm: 9
- Collaboration: 2
Round One: Data Sources

What are the most cited data sources?
Each bar shows how many proposals mentioned a specific data source

<table>
<thead>
<tr>
<th>Data Source</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geospatial data</td>
<td>111</td>
</tr>
<tr>
<td>Administrative data &amp; Official statistics</td>
<td>106</td>
</tr>
<tr>
<td>Survey data</td>
<td>75</td>
</tr>
<tr>
<td>Mobile data</td>
<td>43</td>
</tr>
<tr>
<td>Crowd sourcing</td>
<td>40</td>
</tr>
<tr>
<td>Satellite data</td>
<td>30</td>
</tr>
<tr>
<td>Health records</td>
<td>29</td>
</tr>
<tr>
<td>Open &amp; web data</td>
<td>26</td>
</tr>
<tr>
<td>Remote sensing</td>
<td>17</td>
</tr>
<tr>
<td>Qualitative data</td>
<td>8</td>
</tr>
<tr>
<td>Photo and video data</td>
<td>6</td>
</tr>
<tr>
<td>Unstructured data</td>
<td>3</td>
</tr>
</tbody>
</table>
Round One: Data Sources by Regions

What are the most cited data sources for each region?

Each bar shows how many proposals mentioned a specific combination of region and data source. The top left bar reads as follows: 46% of proposals mentioning East Asia also mentioned Geospatial data as a data source.

- **East Asia & Pacific**
  - Geospatial data: 46%
  - Administrative data & Official statistics: 23%
  - Survey data: 19%
  - Crowd sourcing: 27%
  - Satellite data: 35%
  - Mobile data: 15%
  - Remote sensing: 12%
  - Health records: 11%
  - Open & web data: 8%
  - Unstructured data: 3%
  - Qualitative data: 4%
  - Photo and video data: 4%

- **Europe & Central Asia**
  - Geospatial data: 56%
  - Administrative data & Official statistics: 33%
  - Survey data: 22%
  - Crowd sourcing: 19%
  - Satellite data: 11%
  - Mobile data: 11%
  - Remote sensing: 9%
  - Health records: 6%
  - Open & web data: 3%
  - Unstructured data: 3%
  - Qualitative data: 3%
  - Photo and video data: 3%

- **Latin America & Caribbean**
  - Geospatial data: 50%
  - Administrative data & Official statistics: 22%
  - Survey data: 19%
  - Crowd sourcing: 28%
  - Satellite data: 9%
  - Mobile data: 16%
  - Remote sensing: 6%
  - Health records: 3%
  - Open & web data: 6%
  - Unstructured data: 6%
  - Qualitative data: 6%
  - Photo and video data: 6%

- **South Asia**
  - Geospatial data: 44%
  - Administrative data & Official statistics: 32%
  - Survey data: 28%
  - Crowd sourcing: 20%
  - Satellite data: 16%
  - Mobile data: 16%
  - Remote sensing: 12%
  - Health records: 8%
  - Open & web data: 19%
  - Unstructured data: 11%
  - Qualitative data: 7%
  - Photo and video data: 7%

- **Sub-Saharan Africa**
  - Geospatial data: 29%
  - Administrative data & Official statistics: 34%
  - Survey data: 22%
  - Crowd sourcing: 16%
  - Satellite data: 8%
  - Mobile data: 19%
  - Remote sensing: 7%
  - Health records: 11%
  - Open & web data: 7%
  - Unstructured data: 1%
  - Qualitative data: 1%
  - Photo and video data: 1%
Round One: Data Sources by SDGs

What are the most cited data sources for each SDG?

Each bar shows how many proposals mentioned a specific combination of SDG and data source. The top left bar reads as follows: 36% of proposals mentioning SDG #1 also mentioned Administrative data and Official statistics as a data source.

Administrative data & Official statistics
- Geospatial data
- Survey data
- Mobile data
- Satellite data
- Crowd sourcing
- Health records
- Remote sensing
- Open & web data
- Photo and video data
- Qualitative data
- Unstructured data

1 - no poverty
- Administrative data & Official statistics 36%
- Geospatial data 33%
- Survey data 33%
- Mobile data 21%
- Satellite data 14%
- Crowd sourcing 10%
- Health records 7%
- Remote sensing 2%
- Open & web data 2%
- Photo and video data 2%
- Qualitative data 2%
- Unstructured data 2%

2 - zero hunger
- Administrative data & Official statistics 32%
- Geospatial data 37%
- Survey data 32%
- Mobile data 9%
- Satellite data 12%
- Crowd sourcing 12%
- Health records 10%
- Remote sensing 5%
- Open & web data 1%
- Photo and video data 1%
- Qualitative data 0%
- Unstructured data 0%

3 - good health
- Administrative data & Official statistics 35%
- Geospatial data 37%
- Survey data 25%
- Mobile data 13%
- Satellite data 10%
- Crowd sourcing 4%
- Health records 4%
- Remote sensing 10%
- Open & web data 2%
- Photo and video data 2%
- Qualitative data 0%
- Unstructured data 0%

4 - quality education
- Administrative data & Official statistics 37%
- Geospatial data 23%
- Survey data 23%
- Mobile data 9%
- Satellite data 12%
- Crowd sourcing 14%
- Health records 5%
- Remote sensing 6%
- Open & web data 1%
- Photo and video data 2%
- Qualitative data 0%
- Unstructured data 0%

5 - gender equality
- Administrative data & Official statistics 38%
- Geospatial data 34%
- Survey data 34%
- Mobile data 12%
- Satellite data 14%
- Crowd sourcing 11%
- Health records 9%
- Remote sensing 3%
- Open & web data 3%
- Photo and video data 3%
- Qualitative data 3%
- Unstructured data 3%

6 - clean water and sanitation
- Administrative data & Official statistics 23%
- Geospatial data 34%
- Survey data 34%
- Mobile data 26%
- Satellite data 17%
- Crowd sourcing 17%
- Health records 11%
- Remote sensing 11%
- Open & web data 9%
- Photo and video data 3%
- Qualitative data 3%
- Unstructured data 3%

7 - affordable and clean energy
- Administrative data & Official statistics 42%
- Geospatial data 33%
- Survey data 19%
- Mobile data 19%
- Satellite data 15%
- Crowd sourcing 15%
- Health records 7%
- Remote sensing 4%
- Open & web data 8%
- Photo and video data 8%
- Qualitative data 8%
- Unstructured data 8%

8 - decent work and econ. growth
- Administrative data & Official statistics 41%
- Geospatial data 36%
- Survey data 22%
- Mobile data 17%
- Satellite data 11%
- Crowd sourcing 4%
- Health records 7%
- Remote sensing 4%
- Open & web data 8%
- Photo and video data 8%
- Qualitative data 8%
- Unstructured data 8%

9 - industry and infrastructure
- Administrative data & Official statistics 32%
- Geospatial data 26%
- Survey data 21%
- Mobile data 16%
- Satellite data 10%
- Crowd sourcing 8%
- Health records 7%
- Remote sensing 4%
- Open & web data 8%
- Photo and video data 8%
- Qualitative data 8%
- Unstructured data 8%

10 - reduced inequalities
- Administrative data & Official statistics 39%
- Geospatial data 21%
- Survey data 19%
- Mobile data 6%
- Satellite data 4%
- Crowd sourcing 2%
- Health records 5%
- Remote sensing 6%
- Open & web data 5%
- Photo and video data 4%
- Qualitative data 4%
- Unstructured data 4%

11 - sustainable cities
- Administrative data & Official statistics 21%
- Geospatial data 43%
- Survey data 31%
- Mobile data 19%
- Satellite data 13%
- Crowd sourcing 6%
- Health records 6%
- Remote sensing 2%
- Open & web data 5%
- Photo and video data 5%
- Qualitative data 5%
- Unstructured data 5%

12 - responsible consumption
- Administrative data & Official statistics 39%
- Geospatial data 39%
- Survey data 39%
- Mobile data 39%
- Satellite data 39%
- Crowd sourcing 39%
- Health records 39%
- Remote sensing 39%
- Open & web data 39%
- Photo and video data 39%
- Qualitative data 39%
- Unstructured data 39%

13 - climate action
- Administrative data & Official statistics 17%
- Geospatial data 45%
- Survey data 15%
- Mobile data 15%
- Satellite data 15%
- Crowd sourcing 15%
- Health records 15%
- Remote sensing 15%
- Open & web data 15%
- Photo and video data 15%
- Qualitative data 15%
- Unstructured data 15%

14 - life below water
- Administrative data & Official statistics 15%
- Geospatial data 46%
- Survey data 38%
- Mobile data 38%
- Satellite data 38%
- Crowd sourcing 38%
- Health records 38%
- Remote sensing 38%
- Open & web data 38%
- Photo and video data 38%
- Qualitative data 38%
- Unstructured data 38%

15 - life on land
- Administrative data & Official statistics 26%
- Geospatial data 46%
- Survey data 31%
- Mobile data 10%
- Satellite data 10%
- Crowd sourcing 10%
- Health records 10%
- Remote sensing 10%
- Open & web data 10%
- Photo and video data 10%
- Qualitative data 10%
- Unstructured data 10%

16 - justice and strong institutions
- Administrative data & Official statistics 42%
- Geospatial data 24%
- Survey data 26%
- Mobile data 5%
- Satellite data 3%
- Crowd sourcing 3%
- Health records 3%
- Remote sensing 3%
- Open & web data 3%
- Photo and video data 3%
- Qualitative data 3%
- Unstructured data 3%

17 - partnerships for the goals
- Administrative data & Official statistics 32%
- Geospatial data 29%
- Survey data 15%
- Mobile data 10%
- Satellite data 5%
- Crowd sourcing 5%
- Health records 5%
- Remote sensing 5%
- Open & web data 5%
- Photo and video data 5%
- Qualitative data 5%
- Unstructured data 5%
Round Two - 2017

• Call for Proposals in August 2017: Data Innovation for Environment and Leave No One Behind
• Over 200 proposals received
• Selection process underway - 24 reviewers representing GPSDD Secretariat, Technical Advisory Group, DFID and World Bank
• Expect to announce winners in January
Round Two: Proposals by Regions

- Sub-Saharan Africa: 126 proposals
- South Asia: 56 proposals
- East Asia & Pacific: 38 proposals
- Latin America & Caribbean: 35 proposals
- North America: 28 proposals
- Europe & Central Asia: 24 proposals
- Middle East & North Africa: 12 proposals

Number of Proposals
Round Two: Proposals by Focus Area

- People Who Live Outside Traditional Households: 103 proposals
- Climate Change: 83 proposals
- Urban Resilience: 59 proposals
- People with Disabilities: 47 proposals
Round Two: Proposals by Themes

- Health: 70
- Climate Change: 56
- Environment: 56
- Housing and Land Reforms: 48
- Water and Sanitation: 40
- Gender Equality: 37
- Equality and Rights: 36
- Agriculture: 31
- Food Security: 30
- Disaster Management: 26
- Employment and Jobs: 26
- Infrastructure: 25
- Education: 25
- Accessibility for the Disabled: 22
- Refugees: 17
- Energy: 15

Number of Proposals
Round Two: Proposals by Product

- Reports: 105
- Data Collections and Surveys: 88
- Datasets and Databases: 86
- Maps and Geospatial Products: 65
- Web Platforms: 56
- Monitoring Tools: 50
- Workshops: 39
- Mobile Apps: 36
- Mathematical Models: 29
- Code Repositories: 22
- Journal/Conference Publications: 22
- Policy Briefs: 18

Number of Proposals
Round Two: Proposals by Data

- Survey data: 136 proposals
- Official statistics: 118 proposals
- Open data: 112 proposals
- Geospatial data: 110 proposals
- Qualitative data: 107 proposals
- Administrative data: 98 proposals
- Satellite imagery: 79 proposals
- Photo or video data: 65 proposals
- Private sector data: 53 proposals
- Unstructured data: 53 proposals
- Web or social media data: 46 proposals
- Other (please specify): 41 proposals
- Sensor data: 40 proposals
- Metadata (e.g. call detail records): 36 proposals
- Commercially licensed data: 17 proposals
Looking ahead

• Continue support for tackling development challenges with data innovation
• Learning from successes and failures
• Sharing new tools with the global community
• Scaling effective approaches through World Bank’s investment and advisory operations, and through GPSDD