

Accessing corporate data: are PPP's the solution for closing SDG data gaps?

Johannes Jütting

Manager, PARIS21 Secretariat



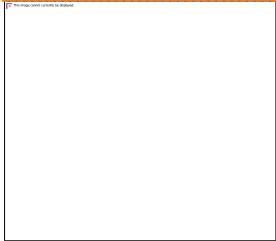
PARIS21

Partnership in Statistics for
Development in the 21st Century

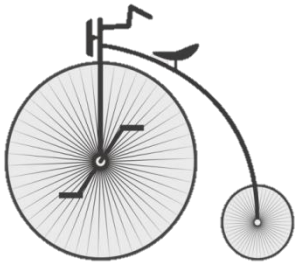
3 mistaken assumptions about PARIS21



We are in the 21st arrondissement of Paris



We are connected to COP21



We were founded in 1921

Who we are

A global partnership of institutions and countries which **promotes the better use and production of statistics** in developing countries

Outline

1. Introduction – PPPs for data
2. The pros and cons
3. Business models to access corporate data
4. Emerging solutions to overcome incentive problems
5. Conclusion and the way forward

Motivation

SDG implementation and data gaps:

- Important data gaps – quality, timelines, granularity and interoperability
- Partnerships as a solution ...but
- ... access as big problem...

Question:

Can PPP's help to facilitate access to private data?

Emerging literature on PPPs for data

- Robin, Klein and Jütting (2016): generic types of PPPs
- Ballivian and Hoffman (2015): taxonomy of risks and benefits of data sharing
- OECD/PARIS21 (2017): Access to new data sources for statistics (forthcoming)
- Events/Reports specific to telecom data
 - Eurostat (2014) Feasibility Study on the Use of Mobile Positioning Data for Tourism Statistics
 - UNECA 2015 conference on the “Use of mobile technology for statistical processes”
 - Meersman et al. (2016) on “win-win” partnership between MNO Proximus and Statistics Belgium

What are PPP's in statistics?

- Public-Private Partnerships for Statistics:
 - Voluntary, collaborative agreement
 - aimed at increasing an NSS' capacity to provide new or better statistics.
- Distinguishing features:
 1. Long-term agreement that defines concrete roles, responsibilities & rights
 2. Central role of proprietary and privacy risks
 3. Can cover any stage of “data value chain”

2. The pros and cons

Corporate Data from an NSO perspective

Web crawling,
web scraping/
web search analysis



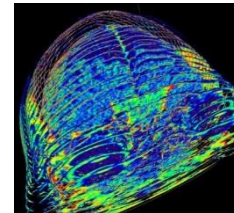
Social
media



Telecom data



Sensor and
geospatial data



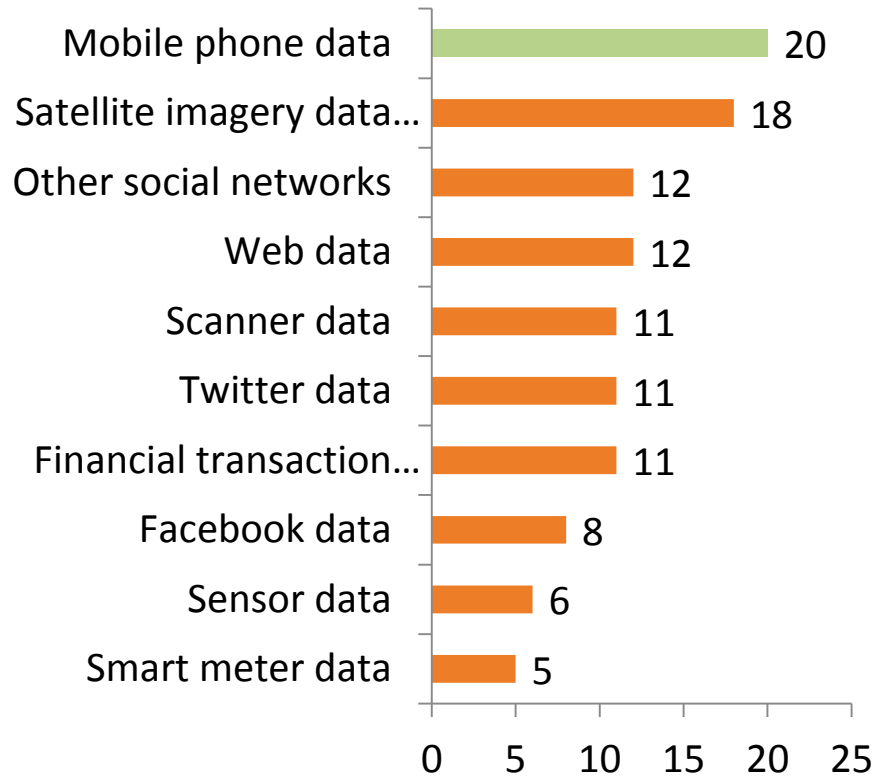
Commercial
transactions
(scanner data,
credit card data)



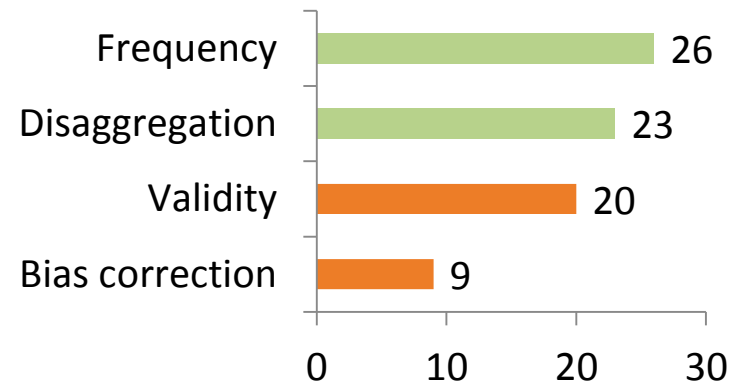
And combinations of these, also with established source such as censuses, surveys, administrative records

Corporate data for SDGs

Projects by type of data source



Feature that indicator improves on



Source: PARIS21 et al. (2015). Global Survey on Big Data projects for SDGs.

Benefits & Complementarities

- For existing statistics
 - Cost effectiveness
 - Timeliness
 - Granularity
- In new areas
 - Data in new areas
 - Increased responsiveness e.g crisis situation

Risks & Challenges

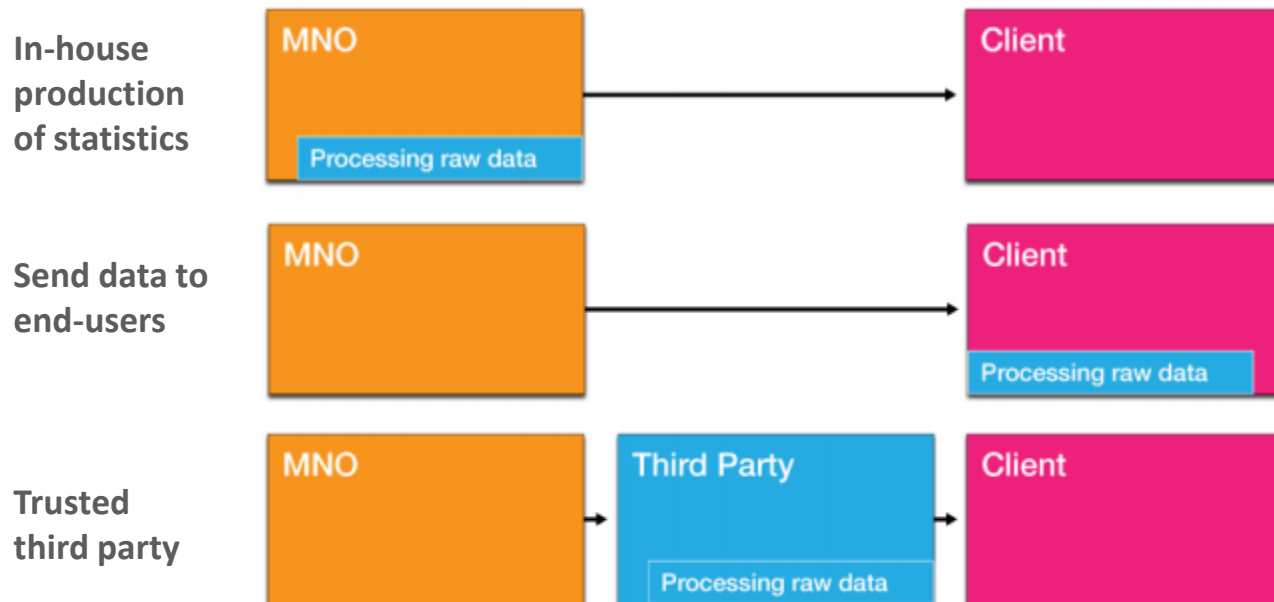
- Access
- Incentives and sustainability
- Privacy and ethics
- Technical and statistical challenges

3. Business Models

	Description and key characteristics	Long term view	Examples of stakeholders
In-house production of statistics	<ul style="list-style-type: none">• Telecom operators compute and “sell” key aggregates based on own algorithms and data	<ul style="list-style-type: none">• Limited scalability because of need to understand each end-user’s requirements	<ul style="list-style-type: none">• Orange, Telefonica, Proximus
Send data to end-users	<ul style="list-style-type: none">• Telecom operators send data to end-users.	<ul style="list-style-type: none">• Difficult to scale up as risks are too high	<ul style="list-style-type: none">• Ad-hoc analyses in case of natural disaster, research projects (e.g. Orange with D4D)
Trusted third party	<ul style="list-style-type: none">• Private or public party hosting aggregated data• Requires setting up the governance, e.g. standard data format, access	<ul style="list-style-type: none">• Allow broad access to aggregate data	<ul style="list-style-type: none">• Some players (e.g. Positium) go into that direction

Positium

- Positium as third-party aggregator/distributor
 - MNO has commercial contract with a third party aggregator responsible for distribution of the data
 - Fixed price / rev-sharing agreements



Business Models

Description and key characteristics

Long term view

Examples of stakeholders

In-house production of statistics

- Telecom operators compute and “sell” key aggregates based on own algorithms and data

- Limited scalability because of need to understand each end-user’s requirements

- Orange, Telefonica, Proximus

Send data to end-users

- Telecom operators send data to end-users.

- Difficult to scale up as risks are too high

- Ad-hoc analyses in case of natural disaster, research projects (e.g. Orange with D4D)

Trusted third party

- Private or public party hosting aggregated data
- Requires setting up the governance, e.g. standard data format, access

- **Allow broad access to aggregate data**

- Some players (e.g. Positium) go into that direction

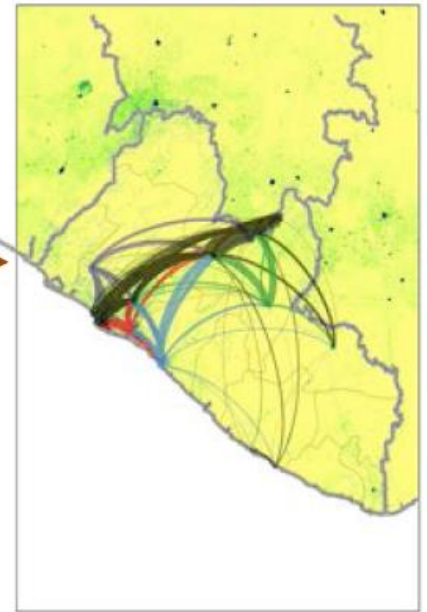
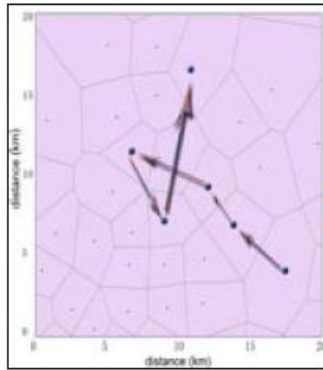
Remote access

- Users develop their algorithms within the corporate systems

- **Allow broad access to individual data**

- RIA, Flowminder

Flowminder



Operators Call Detail Records' (CDRs) including low-resolution location data (nearest tower location) anonymized on separate server hosted by operator.

Flowminder researchers conduct analyses under operator supervision, de-identified raw data always behind operator firewall

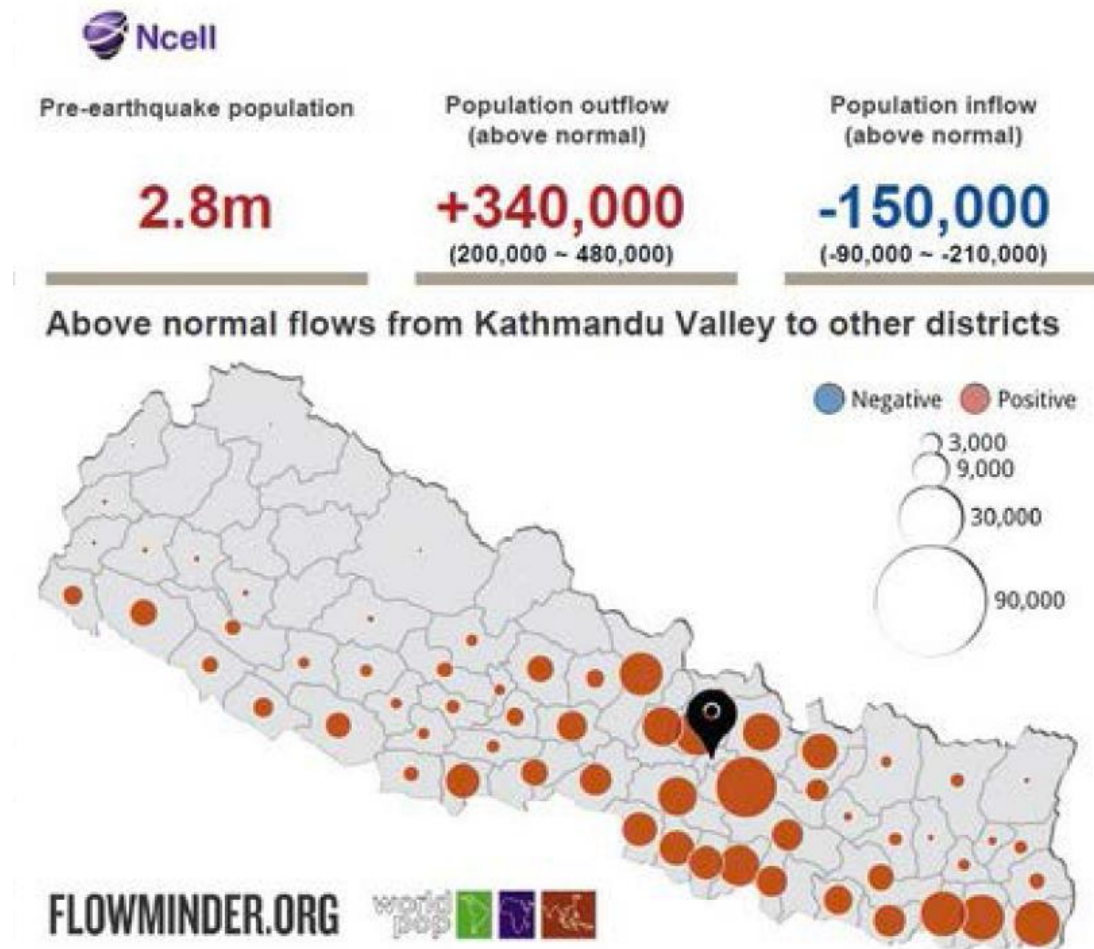
Mobile operator firewall

Non-sensitive aggregated mobility estimates are exported,

Flowminder (cont'd)

Nepal 2015 earthquake:

- Data access/analysis within 14 days
- Information on above normal population flows
- Life-saving information for disaster response



Business Models

Description and key characteristics

Long term view

Examples of stakeholders

In-house production of statistics

- Telecom operators compute and “sell” key aggregates based on own algorithms and data

- Limited scalability because of need to understand each end-user’s requirements

- Orange, Telefonica, Proximus

Send data to end-users

- Telecom operators send data to end-users.

- Difficult to scale up as risks are too high

- Ad-hoc analyses in case of natural disaster, research projects (e.g. Orange with D4D)

Trusted third party

- Private or public party hosting aggregated data
- Requires setting up the governance, e.g. standard data format, access

- **Allow broad access to aggregate data**

- Some players (e.g. Positium) go into that direction

Remote access

- Users develop their algorithms within the corporate systems

- **Allow broad access to individual data**

- RIA, Flowminder

Move algorithms

- Users develop publicly available algorithms and extract results from corporate system

- **Allow broad access to 1st layer algorithms**

- OPAL, e.g. Orange, universities/research institutes

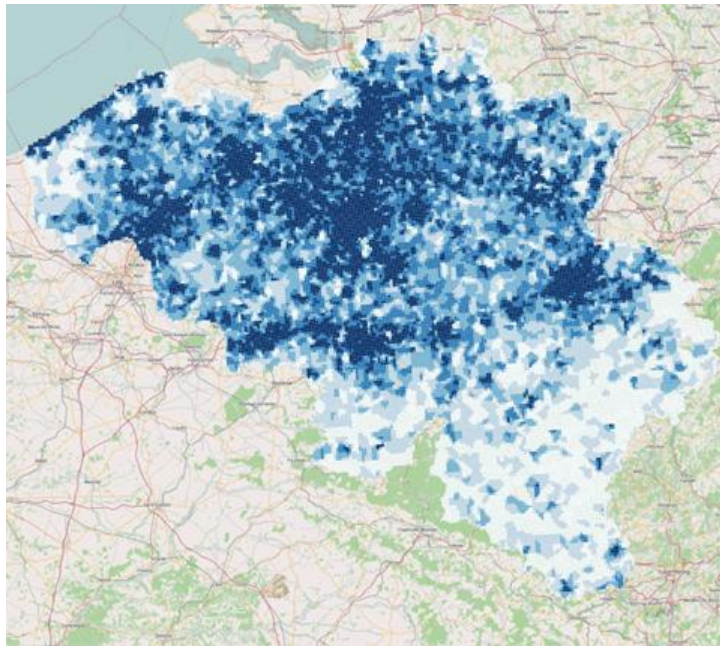
4. Corporate sector incentives

- What does it take for “win-win” partnerships?
- NSO incentives
 - data access to produce statistics
 - operators can provide: technical expertise, data storage, processing infrastructure and use cases
- MNO incentives
 - increase commercial value of MNO data from collaboration with NSO and geocoded NSO data
 - statistical and domain expertise of NSOs
 - Corporate social responsibility (public good)

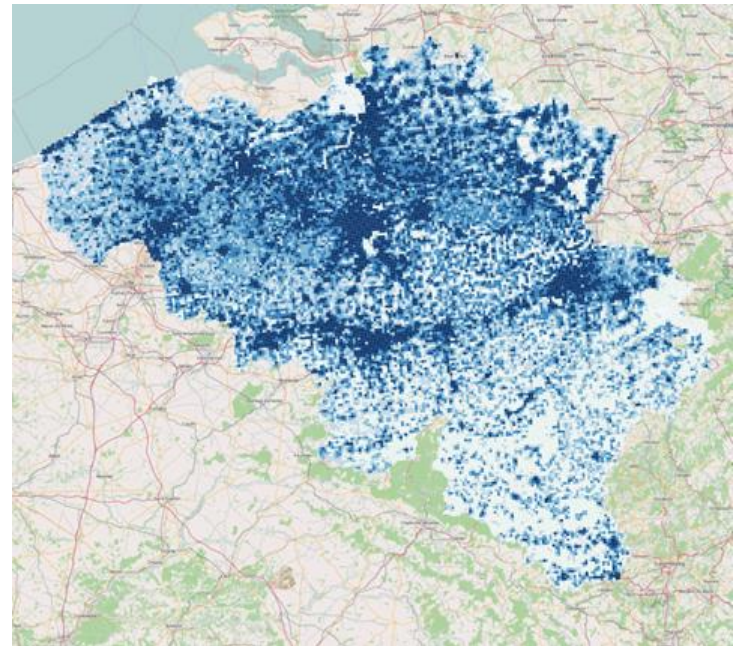
Proximus and Statistics Belgium

Estimates of population density per km² for Belgium

Rho = 0.85



Based on mobile phone data



Based on 2011 Census

5. Conclusion

- Establishing PPPs holds promises and caveats – no miracles to be expected
- Critical question: PPP for what? – profit vs. CSR vs. true “win-win” partnership
- Our survey shows: In developing countries, most business models rely on **ad-hoc data exchange**
- To be scalable for official statistics, need to create a standardised **safe environment** for sharing data

The way forward

- Define a **decision tree** to inform the choice of business models dependent on context, intended use and data type
- Corporate data access is facilitated by “**data stewards**” that act as a first point of contact
- Harnessing the potential and engaging with other actors requires new skills from all actors involved – “**capacity development 4.0**”





PARIS21 Secretariat

OECD/STD

46 Quai Alphonse le Gallo,
92100 Boulogne-Billancourt,
France

contact@paris21.org

www.paris21.org

-  @ContactPARIS21
-  ContactPARIS21
-  www.soundcloud.com/paris21-datataalk
-  www.youtube.com/user/PARIS21OECD

