



SOLEP

Société Luxembourgeoise de l'Evaluation  
et de la Prospective

# The use of big data analytics for policy insight

- ▶ Typology, benefits and challenges



1

## Concepts

- Evaluation and public decision-making
- Data continuum

2

## Examples

- Surveys and census data
- Geospatial imagery and sensors
- New insights from big data analytics

3

## Risks and challenges

- Panel discussion

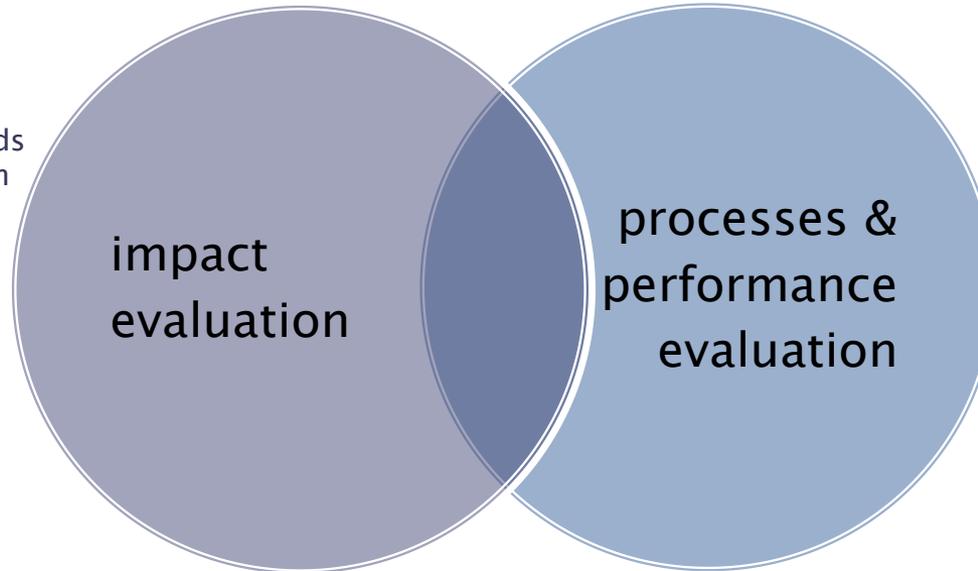
# Evaluation and public decision-making

---

**! Not mutually exclusive !**

Based on **quantitative** methods  
Quasi-experimental design  
→ **counterfactual**

Ask what is the impact on  
beneficiaries? = attribution



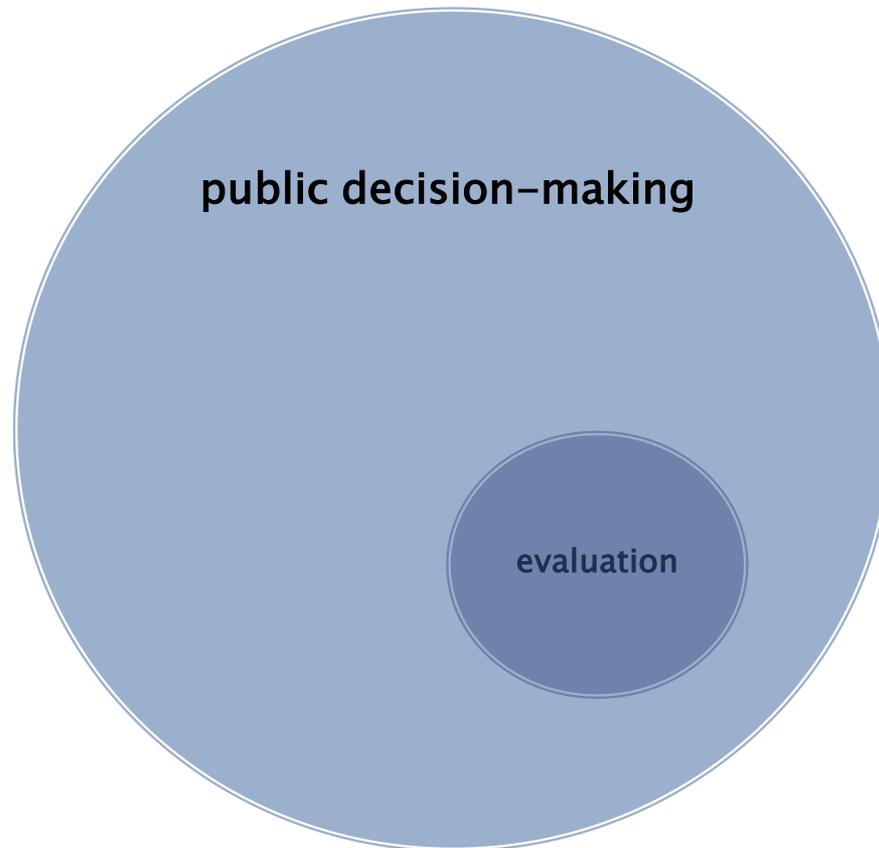
**Qualitative** approach &  
assessment of contribution

OECD evaluation criteria  
*relevance*  
*effectiveness*  
*efficiency*  
*sustainability*

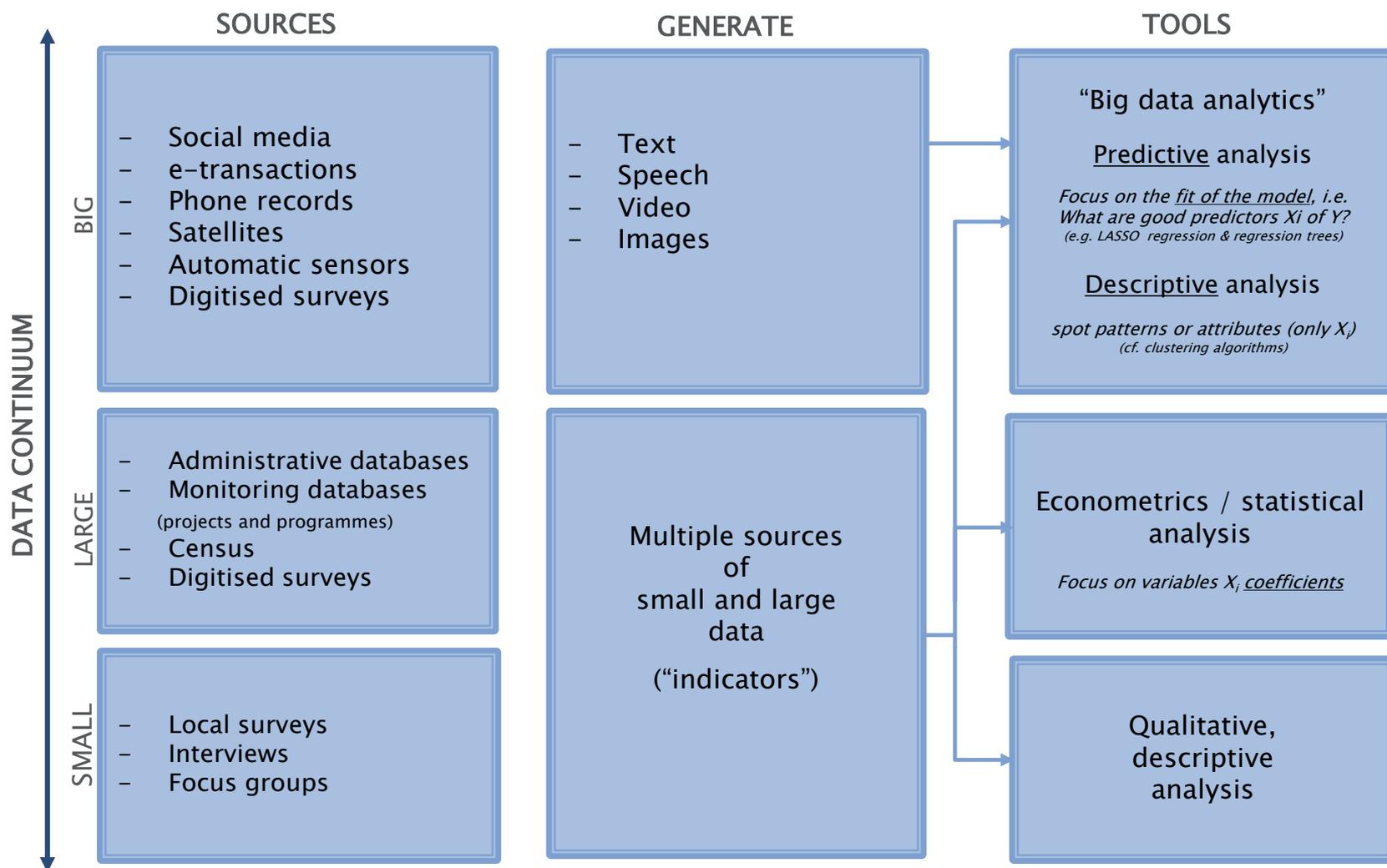
UNDERLYING INTERVENTION

# Evaluation and public decision-making

---



# Data





# SOLEP

Société Luxembourgeoise de l'Évaluation  
et de la Prospective

---

1

## Concepts

- Evaluation and public decision-making
- Data continuum

2

## Examples

- Surveys and census data
- Geospatial imagery and sensors
- New insights from big data analytics

3

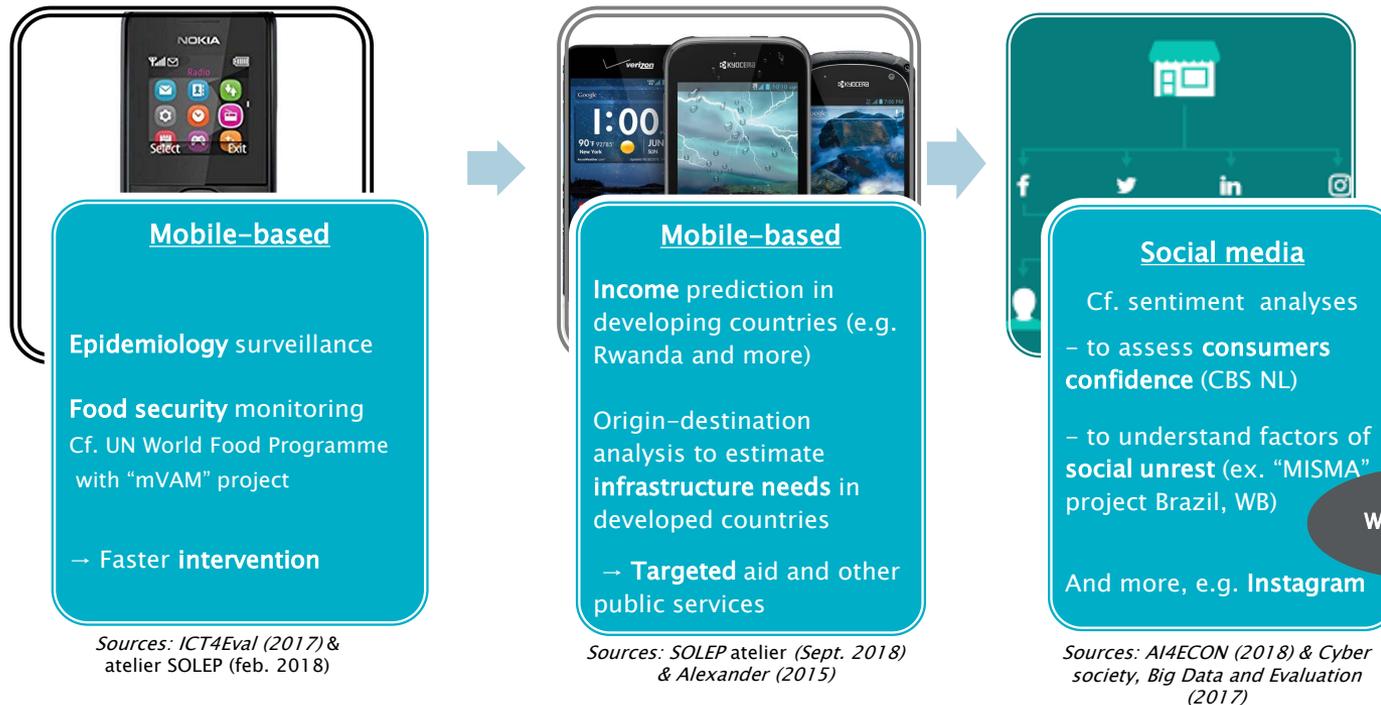
## Risks and challenges

- Panel discussion

# (1) Surveys and census data

ACTIVELY GENERATED DATA

PASSIVELY GENERATED DATA



Workshop 3

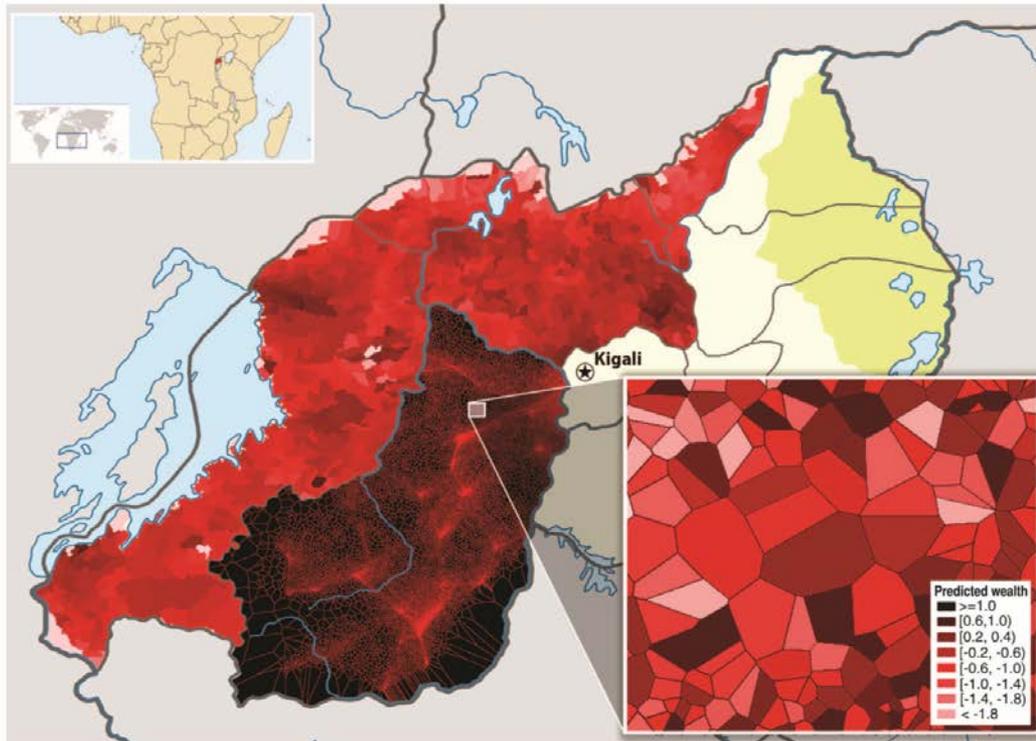
(near) real time information

A-Party-ID	B-Party-ID	Date	Time	Duration	A-Party-Cell	...
979ae8cd	97939b87	2014-01-04	22:00:11	42	2837	...

*Source: Blumenstock (Dec. 2017)*

**PREDICTIVE ANALYSIS in Rwanda**

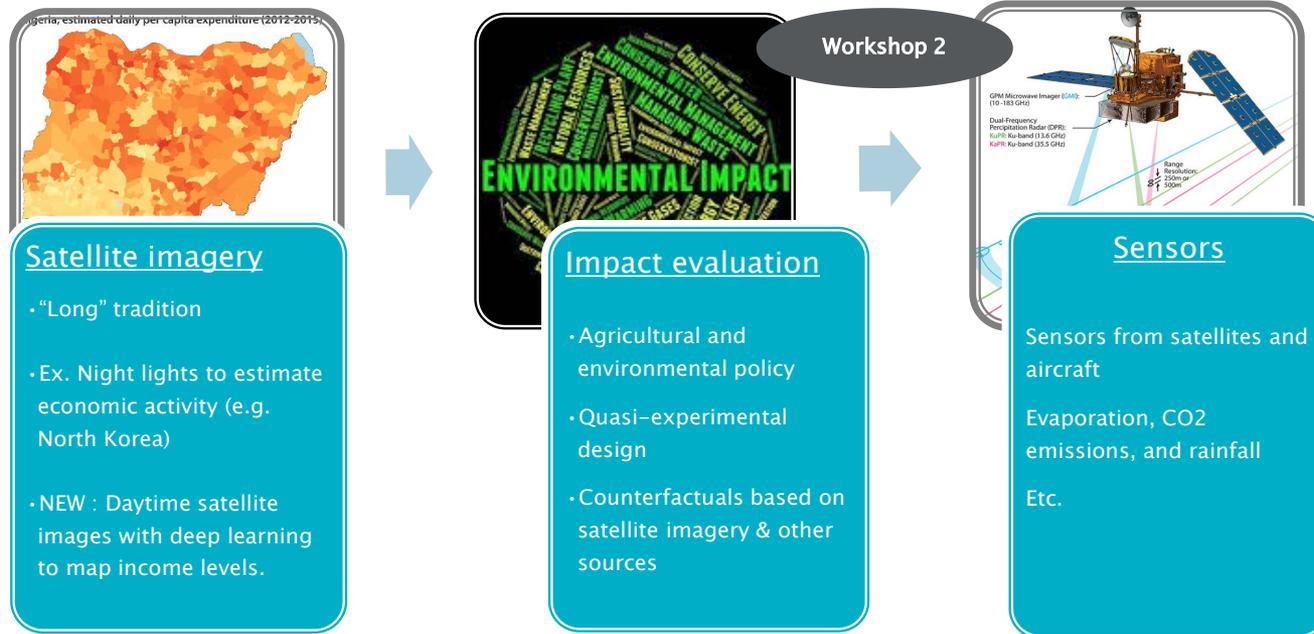
- (1) **Train subsample:** learn about  $X_i$  &  $Y_i$  relationship  
2,200 household (HH) phone survey
- (2) **Predict out-of-sample:** predict  $Y_i$  from  $X_i$   
approx. 1,500,000 mobile phone subscribers
- (3) **Cross-validate**  
with info from National census: 15,000 HH field survey



**Construction of high-resolution maps of poverty and wealth from call records.** Information derived from the call records of 1.5 million subscribers is overlaid on a map of Rwanda. The northern and western provinces are divided into cells (the smallest administrative unit of the country), and the cell is shaded according to the average (predicted) wealth of all mobile subscribers in that cell. The southern province is overlaid with a Voronoi division that uses geographic identifiers in the call data to segment the region into several hundred thousand small partitions. **(Bottom right inset)** Enlargement of a 1-km<sup>2</sup> region near Kiyonza, with Voronoi cells shaded by the predicted wealth of small groups (5 to 15 subscribers) who live in each region.

Source: Blumenstock et al (2016)

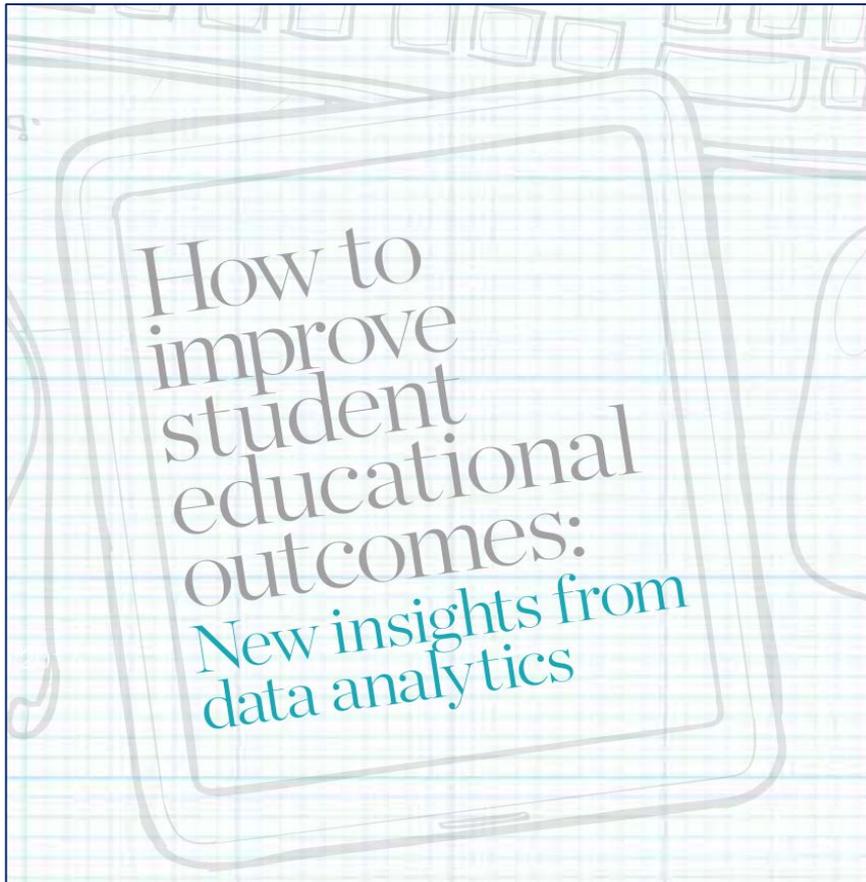
# (2) Geospatial imagery and sensors



IMPROVED IMAGE RESOLUTION  
& OPEN SOURCES

# (3) New insights from big data analytics

---



Source: McKinsey (2017)

- ▶ Rapidly evolving field
  - Training and education
  - Labour policy and more
- ▶ Large and small data
  - Surveys (OECD PISA database e.g.)
  - Administrative databases
  - Projects/programmes-specific databases
- ▶ Aim to uncover patterns & attributes
  - Enrich evaluation analysis.
- ▶ Emerging literature on machine learning and impact evaluation
  - See Athey (2019), AI4ECON (2018) ICT4EVAL conferences (2017)

Workshop 1

Source: SOLEP ateliers (2018)



# SOLEP

Société Luxembourgeoise de l'Évaluation  
et de la Prospective

---

1

## Concepts

- Evaluation and public decision-making
- Data continuum

2

## Examples

- Surveys and census data
- Geospatial imagery and sensors
- New insights from big data analytics

3

## Risks and challenges

- Panel discussion

# Risks and challenges

---

- ▶ No panacea
  - Ground in theory and triangulate methods
    - See afternoon workshops
- ▶ Algorithms weaknesses and ethical concerns
  - Remain critical and be open about these
    - See session 2 and afternoon workshops
- ▶ Culture shock
  - Cf. UN Global Pulse: *“You cannot just put a social scientist and a data scientist in a room and assume that magic will ensue (...). What statisticians, demographers and economists need to realise is that data science is not just a fad. And what computer scientists need to acknowledge is that they cannot solve global poverty by crunching numbers alone”.*
    - How to work together?