

Organische data als bron voor openbare statistieken?

Dries Verlet & Marc Callens

Statistics Flanders

Overzicht

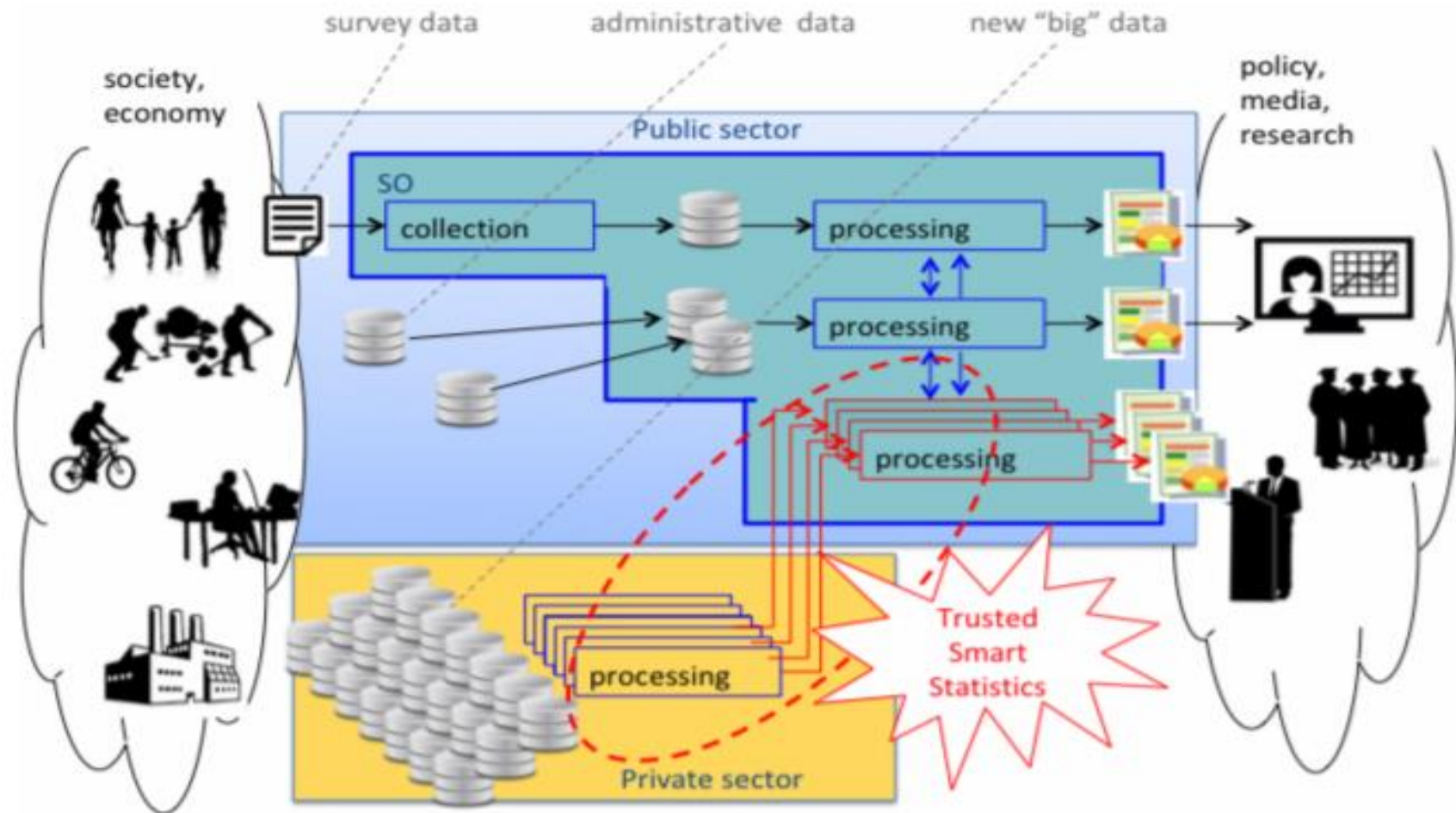
1. Methoden voor gegevensverzameling binnen de wereld van de officiële statistiek
2. Big data/organische data van naderbij bekeken
3. Enkele beschouwingen

→ Eigenlijk niet zozeer over het specifieke gebruik van Twitter data, maar wel over de context waarin we die gebruiken

Inleiding

- Cf. Kotzeva (2018): *'official statistics measures life, and when life is changing, official statistics changes as well'*
 - Daarnaast dienen zich ook nieuw uitdagingen aan:
 - Toenemende nood aan meer gedetailleerde, actuele en instant informatie, snelle ontwikkelingen van nieuwe technieken, traditionele survey onder druk, minder budgetten voor onderzoek, ...
 - Beschikbaarheid van (grote hoeveelheden) data als een bijproduct bij het functioneren van organisaties, registraties allerhande en door onze communicatie via het internet
- De analyse van dat soort data kan op zich ook veel zeggen over hoe gesteld is met een samenleving, bv. Wat is het sentiment?

Drie types methoden om data te verzamelen



Voornaamste methoden (gegevensverzameling)

1. Survey

- Bevestigingen op grond van steekproeven
- Tot nu toe de meest gebruikte en in veel gevallen ook de meest betrouwbare methode (indien we dit op een goede manier uitwerken).

2. Administratieve data

- Data die verzameld worden in het kader van administratieve dienstverlening (statistische of wetenschappelijke gebruik is secundair)
- Vereist vaak het linken van diverse grote databanken
- Gegevensbescherming

Voornaamste methoden

3. Big data/organische data

- Tot nu toe maar weinig gebruikt om openbare statistieken te voeden
- Beschikbaarheid van data; data vaak in handen van private actoren

4. Combinatie van methoden

Bv. Survey in combinatie met administratieve data

Bv. Administratieve data van ondernemingen aanvullen met online beschikbare data (webscraping)

Organic data

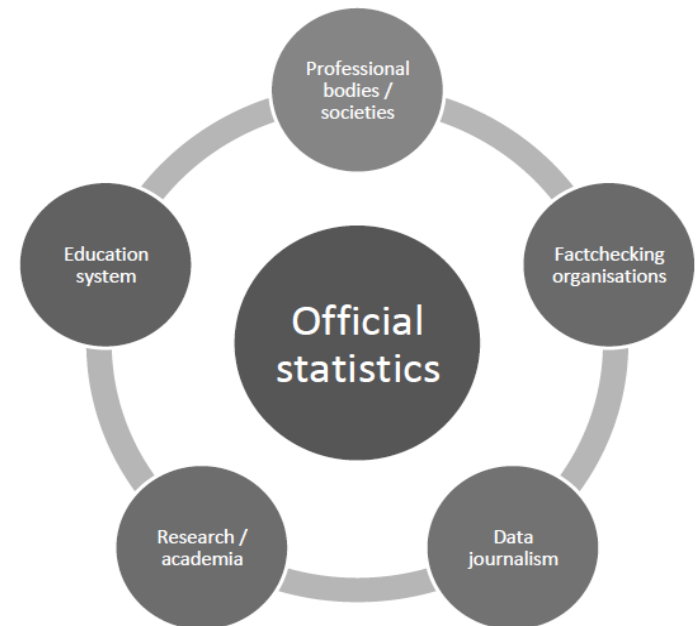
- Is een “bijproduct”, namelijk dergelijke data worden niet verzameld met het oog op administratieve of statistische doelstellingen (cf. organic data)
- Van waar komen dergelijke data?
 - Sociale interactie, publicatie en conversaties via sociale netwerken zoals Twitter, Facebook etc.
 - Sensoren en mechanische metingen
 - Bedrijfsmatige transacties
 - Elektronische bestanden
 - Broadcast media

Organic data: tal van uitdagingen dienen zich aan

- Omtrent de validiteit en betrouwbaarheid van de data:
 - In welke mate kan dergelijke data gemanipuleerd worden?
 - Beschikbaarheid van data? cf. aanlevering van data die onzeker is/kan zijn, regelgevend kader, instemming?
 - Selectiviteit van de dataverzameling en bereik populatie
 - ...
- Technische uitdagingen: zware rekenkracht, ontwikkeling van algoritmes, analytische tools, integratie van data...
- Context

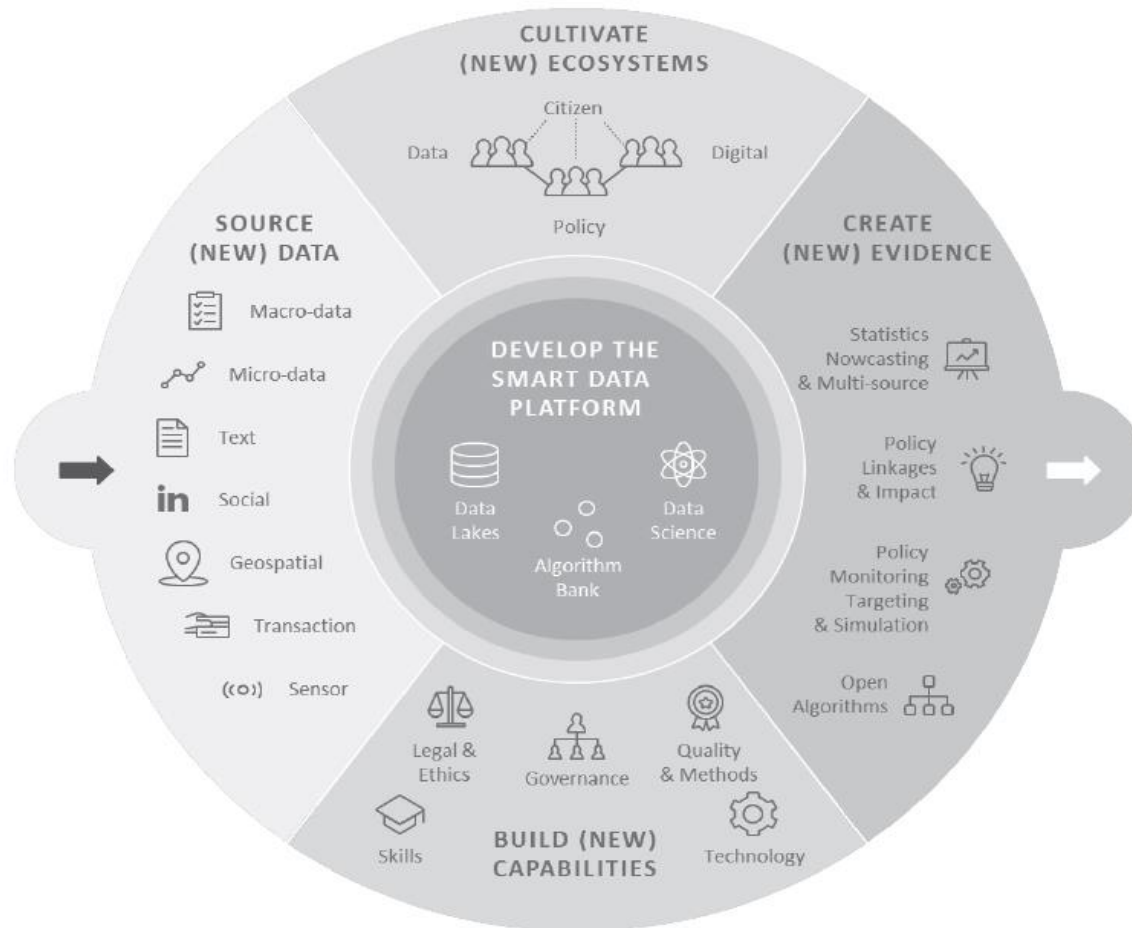
Organic data: tal van uitdagingen dienen zich aan

- Context:
 - Subtieler, maar daarom niet minder belangrijk: veranderingen in de percepties, verwachtingen, gedragingen en relaties tussen de verschillende actoren/stakeholders, cf. de nood aan een ecosysteem van data
 - Cf. discussie status experimenteel dan wel officieel karakter van statistieken
 - Cf. Smart data framework als raamwerk voor smart statistics
bv. Eurostat, OECD



Bron: Beerten 2018: 68

Cf. nood aan een “smart data framework”

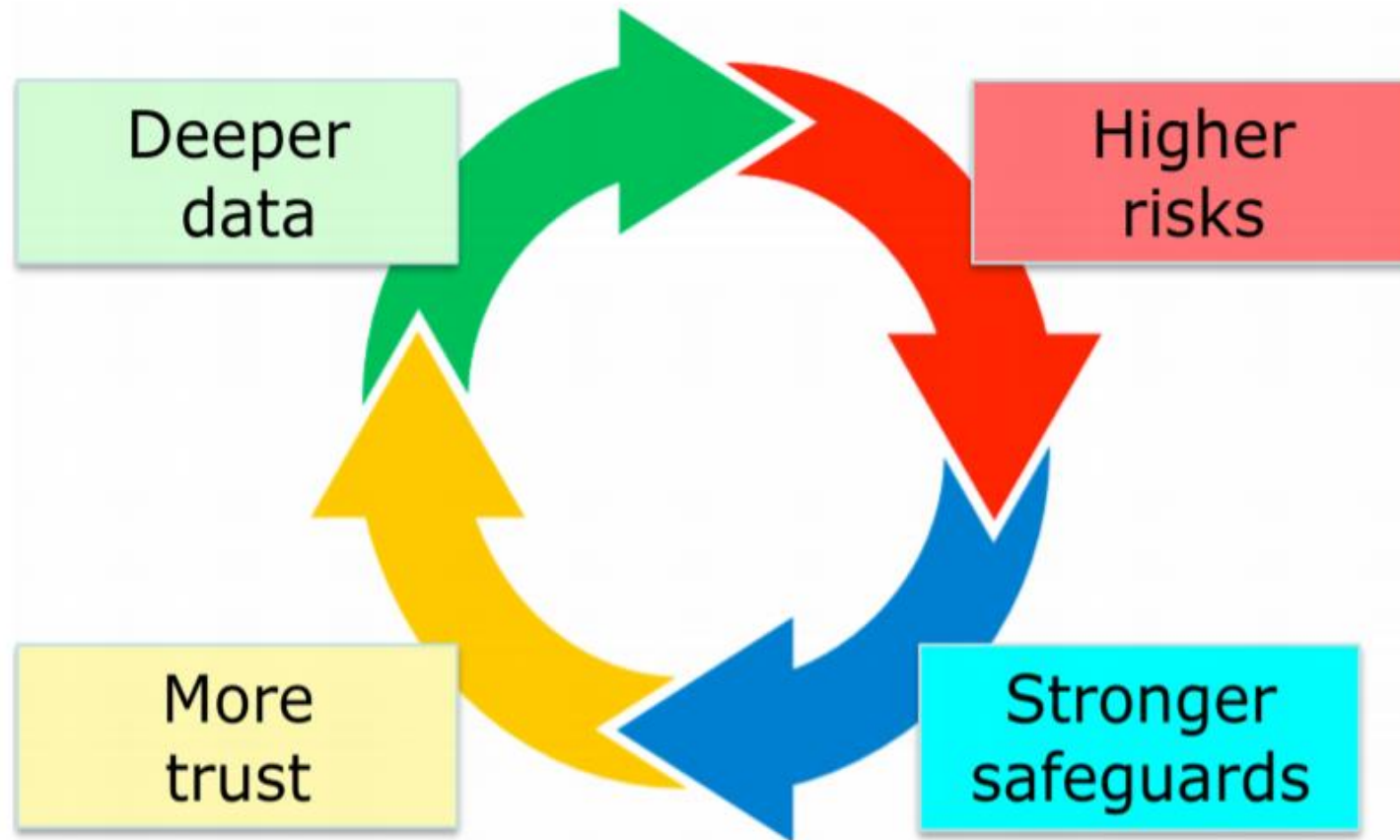


Bron: OESO 2018: 6

Smart Statistics

- Samenspel tussen “slim” en “vertrouwen” als twee aspecten die elkaar mogelijk maken en versterken
- Cf. Ricciato e.a., : *This system should be ‘more extrovert towards external stakeholders (citizens, private companies, public authorities) with whom Statistical Offices will be sharing computation, control, code, logs and of course final statistics, without necessarily sharing the raw input data.’*

Wisselwerking tussen “slimme” en “vertrouwde” data



Enkele beschouwingen ter afronding

- Toegenomen nood aan meer gedetailleerde, actuele en “instant” informatie, snelle ontwikkelingen van nieuwe technieken, andere actoren in een nieuwe setting,
 - Nieuwe (grote) uitdagingen
- Nieuwe mogelijkheden dienen zich aan, bv. data van sociale media om data uit meer traditionele databronnen aan te vullen
- Niet alleen een technisch gegeven, moeten ook bouwen aan een omgeving waar stakeholders vertrouwen in hebben
 - Ook hier is een belangrijke rol weggelegd voor statistiekinstellingen

Nog vragen?

Contact:

Dries.Verlet@vlaanderen.be

Organische data als bron voor openbare statistieken?

Dries Verlet & Marc Callens

Statistics Flanders