

Nya datakällor – nya möjligheter

Lilli Japec, Dr

Surveyföreningens kvalitetsseminarium

190207



Inledning

- Förändringar i omvärlden och statistikbranschen
- Strategier för att hantera utmaningarna
- Exempel från bl.a. BigSurv18 om hur nya datakällor kan användas i statistikproduktionen

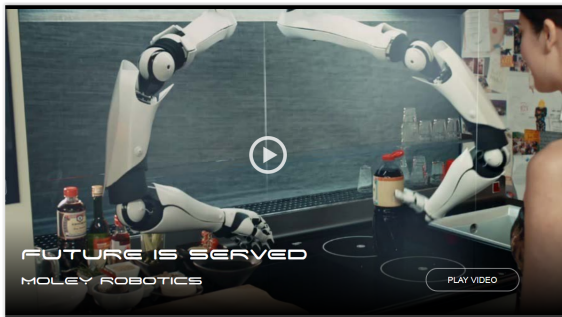
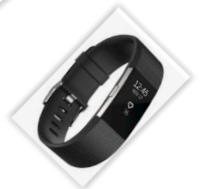
Omvärlden förändras



How the IoT is creating 'precision farming'
precisionagricultu.re



Building A Smart City? 10 Big ...
forbes.com



FUTURE IS SERVED
MOLEY ROBOTICS

<http://www.moley.com/>



Smart Homes Will Change Our Way of Life ...
medium.com



Källa: Smart Slab. SapienStone



Exempel på nya datakällor

1. Sociala nätverk t.ex. Facebook, Twitter, bloggar, foton, videofilmer, Youtube och internetsökningar.
2. Traditionella företagssystem t.ex. transaktioner, kundregister, kreditkortsdata, medical records
3. Internet of Things d.v.s. olika typer av sensordata t.ex. väder, trafik, övervakning, mobiltelefoner, bilar och satellitbilder

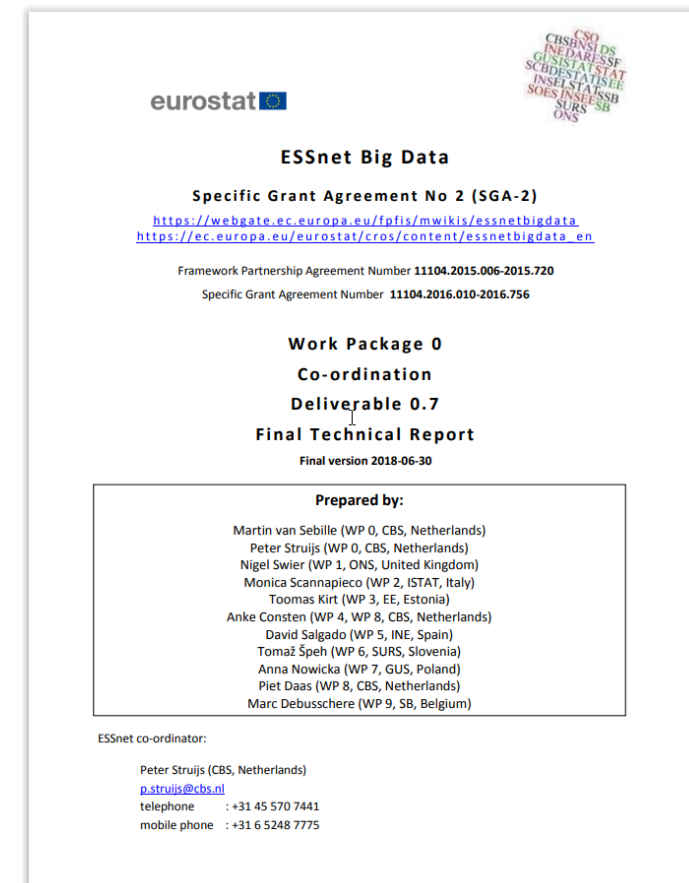
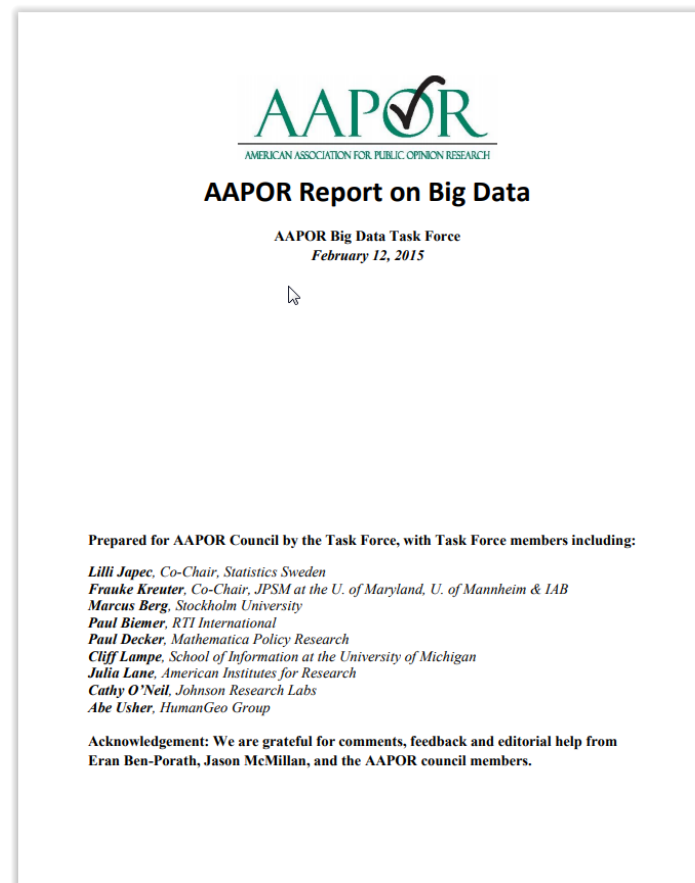
Källa: UNECE classification of Big Data

Statistikbranschen förändras

- Bortfallet i undersökningar ökar –speciellt svårt i undersökningar som kräver omfattande insatser från uppgiftslämnare
- Dyrt att samla in data
- Med ökat bortfall ökar risken för fel i skattningarna
- Nya datakällor – hur kan dessa användas i statistikproduktion?

Olika initiativ och strategier för att möta utmaningar ...

- AAPOR rapport om Big Data
- ESSNet on Big Data



...hackathons

- team med t.ex. "computer scientist", programmerare och ämneskompetens
- en begränsad uppgift som ska lösas under en kort begränsad tidsperiod
- uppgiften går ofta ut på att hitta en kreativ lösning t.ex. ta fram ett verktyg eller att visualisera data på nytt sätt

 **Hack for Sweden** @HackForSweden · Feb 1
We've announced the categories for which the hackathon will focus on this year – the six most data-mature sectors!

-  Mobility
-  Environment, nature and geography
-  Education and Science,
-  Labor market
-  Business sector
-  Health

The application will open really soon! 🗓️



The European Big Data **Hackathon** is an event organised by the European Commission (**Eurostat**) which gathers teams from all over Europe to compete for the best data product combining official statistics and big data to support policy makers in one pressing policy question or statistical challenge facing Europe.



European Big Data Hackathon 2019 - European Commission

https://ec.europa.eu/eurostat/cros/EU-BD-Hackathon_en ✓



Hackathon

The Green City Hackathon consisted of six teams who addressed bike use in the city of Barcelona using data published in the [Open Data BCN Portal](#).

The winning presentation was by Angela Ulrich, Camila González, Ramón Ferri García, and Ian Thomas (Team 6) on "Barcelona. Bike Use and Accident Overview." Congratulations to the winning team and all participants!

All hackathon presentations can be found here at <http://opendata-ajuntament.barcelona.cat/en/noticia-green-city-hackathon> or in our a github repository <https://github.com/bigsurv18/bigsurv18> (including replication materials if available).

...Big Data Centers och partnerskap



Labour and income Economy Society Regional Corporate Figures

Innovation **Big Data** Open Data Output Contact

Big Data



 Data Science
Campus

Search blog 

Home Partnerships Projects Building capability Team Contact us About us

DATA SCIENCE FOR PUBLIC GOOD

Join us at the MSc in Data Analytics for Government (MDataGov) Symposium 2019!

Registration for the 2019 MSc in Data Analytics for Government (MDataGov) Symposium is now open. The annual Symposium is an opportunity for...



..konferenser

Sponsorer:



BigSurv18

- Monografi (Wiley)
 - The new survey landscape
 - Total error and data quality
 - Big data in official statistics
 - Combining big data with survey statistics: methods and applications
 - Combining big data with survey statistics: tools
 - Regulations, ethics, privacy
- Special nummer i Social Science Computer Review
- BigSurv20 – Statistics Canada är värd för konferensen som kommer att hållas i oktober 2020

**”Surveys are the last
resort”**

Tom Smith (2018)

**“Wider, Deeper, Better,
Quicker and Cheaper”
Tim Holt 2007**



”Wider and Deeper”

Textanalys och webscraping

- En viktig ekonomisk indikator i ett land är antal innovativa företag
- I EU samlar vi in den typen av uppgifter via en undersökning till företag med fler än 10 anställda (Community Innovation Survey-CIS).
- Kan man skrapa av företags webbsidor för att identifiera innovativa företag? (van der Doef et al 2018)
- Genom att skrapa av nätet och även inkludera små företag samt geografisk information så får man mer detaljerad information än vad CIS kan ge

“Wider and Deeper”

- Vakansstatistik genom att skrapa av nätet
- Arbetsförmedlingens platsdatabank och Vakansundersökningen (Wu, Jansson och Elezovic 2018)

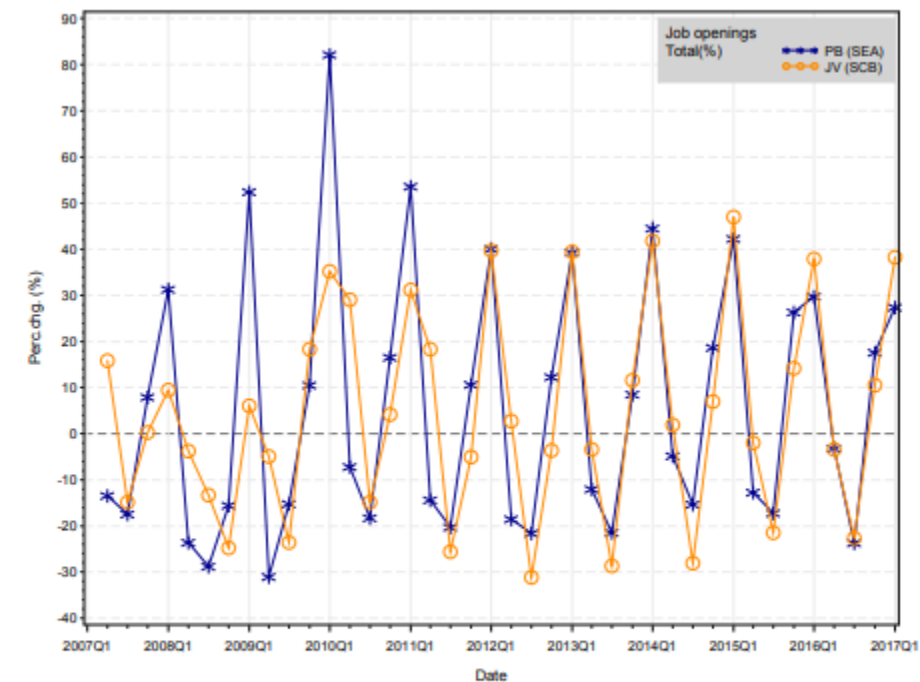
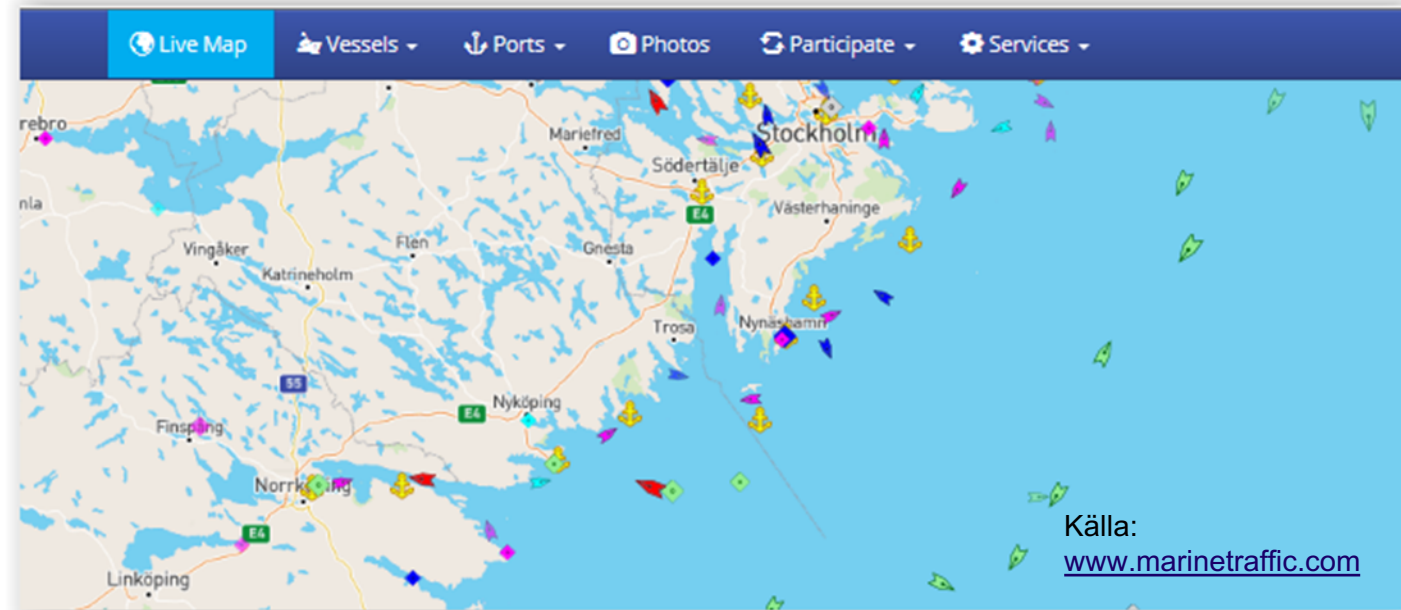


Figure 5.1b PB (SEA) vs JV (SCB) – total number of job openings (Private sector and Public sector, jointly); Growth rates as percentage changes from the previous quarter; quarterly values from 2007q1 to 2017q1.

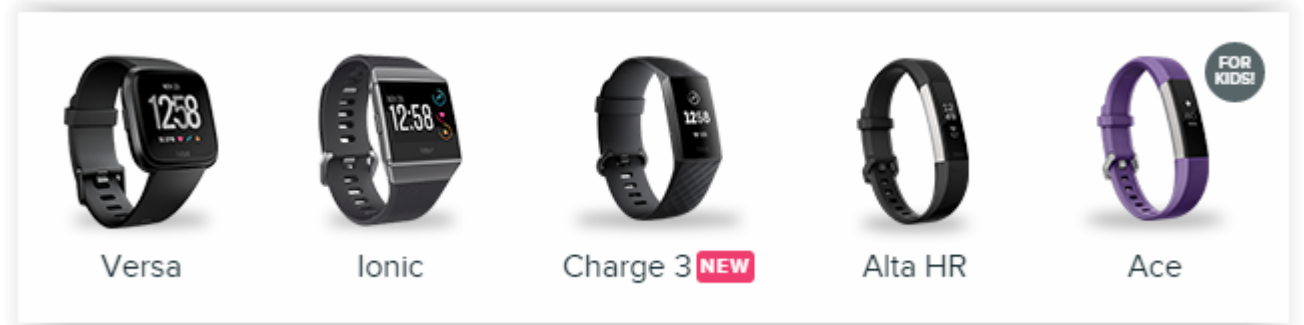
”Better”

AIS-data

- EU-projekt för att utvärdera om AIS-data (*Automatic Identification System*) kan användas för officiell statistikproduktion
- Statistik om sjöfartstrafik för att skatta utsläpp eller för att identifiera områden med tung trafik
- Hamnsstatistik för att t.ex. se hur fartyg rör sig mellan olika hamnar



”Better”



Human wearables

- Att samla in data om fysisk aktivitet är svårt i en undersökning
- Det finns en tendens att överrapportera
- Kan Fitbit-data ge en bättre bild av fysisk aktivitet? (Eckman, Amaya and Furberg 2018). Man studerade även sömndata.
- Kan aktivitetsmätare användas för att mäta ungdomars fysiska aktivitet (the Millenium Cohort Study)? (Calderwood and Gilbert 2018)

”Quicker”

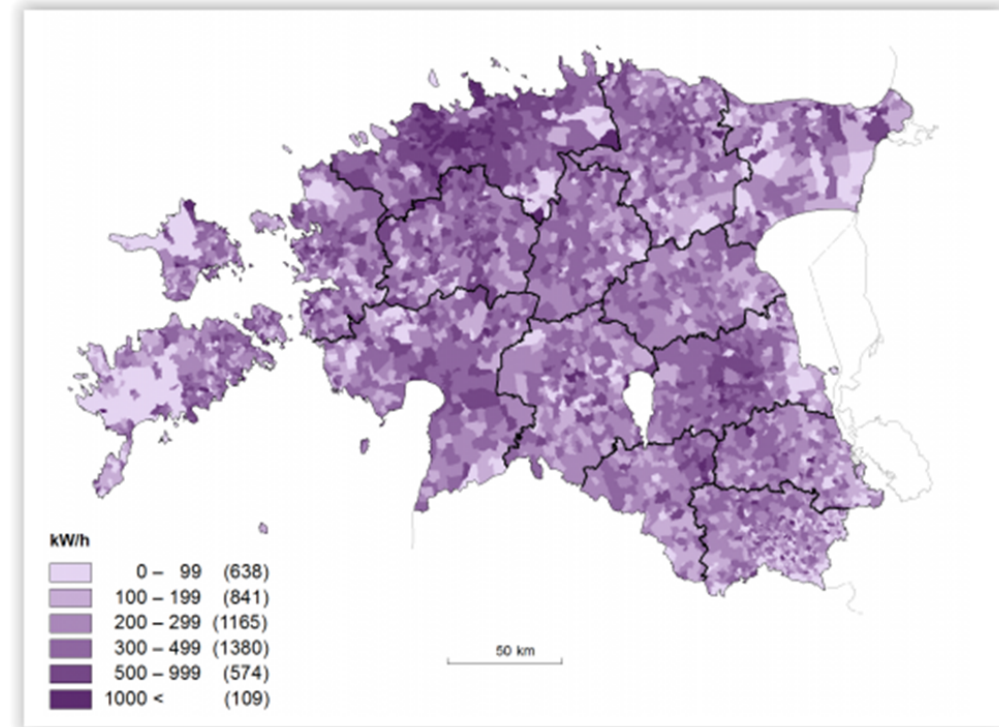
Road density data

- Kan man snabba upp produktionen av BNP? (Luomaranta et al 2018)
- I Finland använde man bl.a. “road density” data för att modellera kvartals-BNP och man kunde snabba upp produktionstiden med 45 dagar
- Konstigt?
- Motsvarande studie har gjorts i Slovenien med liknande resultat

”Cheaper”

- Kan data från smarta elmätare användas för att producera energistatistik? (Ilves and Kirt, 2016)

Average monthly electricity consumption by private persons in Estonia, January 2014



”Cheaper”

Kodning

- I finska Arbetskraftsundersökningen (AKU) frågar man uppgiftslämnarna bl.a. om deras yrke
- 15% av arbetstiden i den finska AKU går till att koda yrke och näringsgren (Kärkimaa and Larja, 2018).
- Kan man använda “machine learning” för att effektivisera kodningsarbetet?

Kvalitet

- Studier har gett mer kunskap om olika felkällor då man använder big data för statistikproduktion
- Kvalitetsramverk
- Kvalitet i olika indikatorer

Ny uppsättning juridiska och etiska frågor

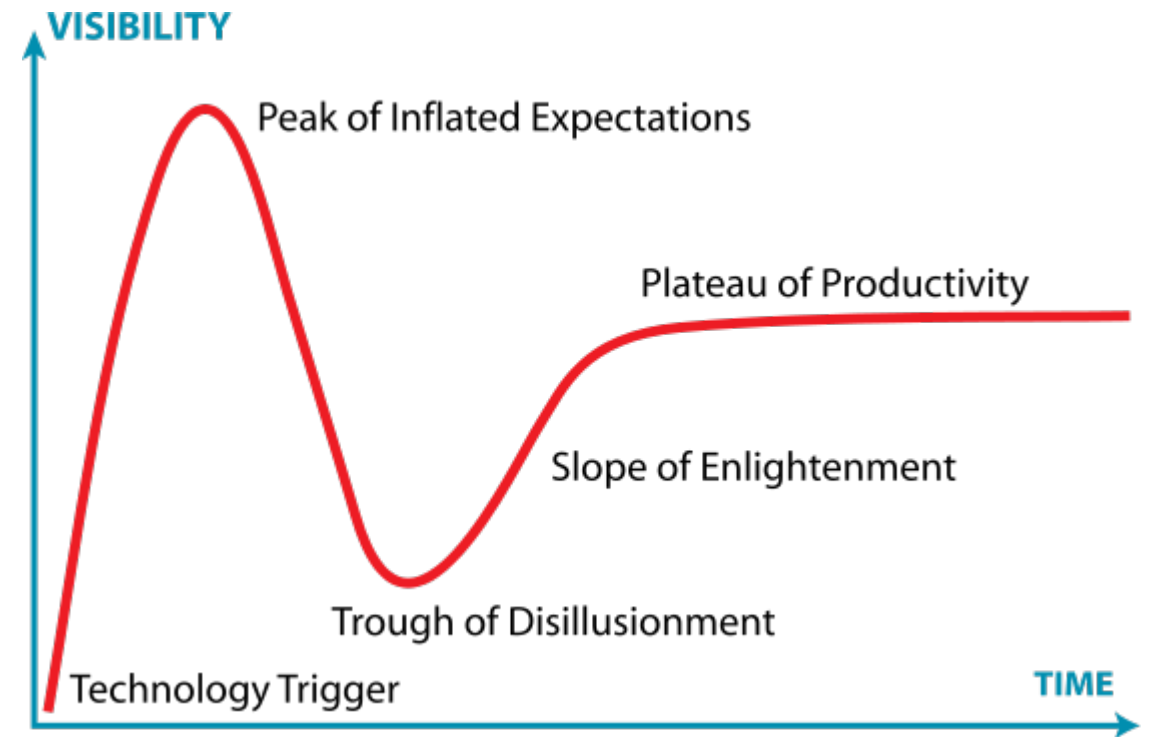
- Data samlas ofta in för ett ändamål men kombineras med andra datakällor och används för ett annat ändamål
- Man kan fundera på vem som äger eller borde äga den här typen av data
- GDPR nämner inte uttryckligen big data

Slutligen

- Hypen kring big data har avtagit
- Var är vi idag på kurvan?

<http://www.moley.com/>

Gartner's hype curve



Source: Wikipedia