



Measuring the internet economy in the Netherlands

A big data analysis

Tommy Span

8 May 2019

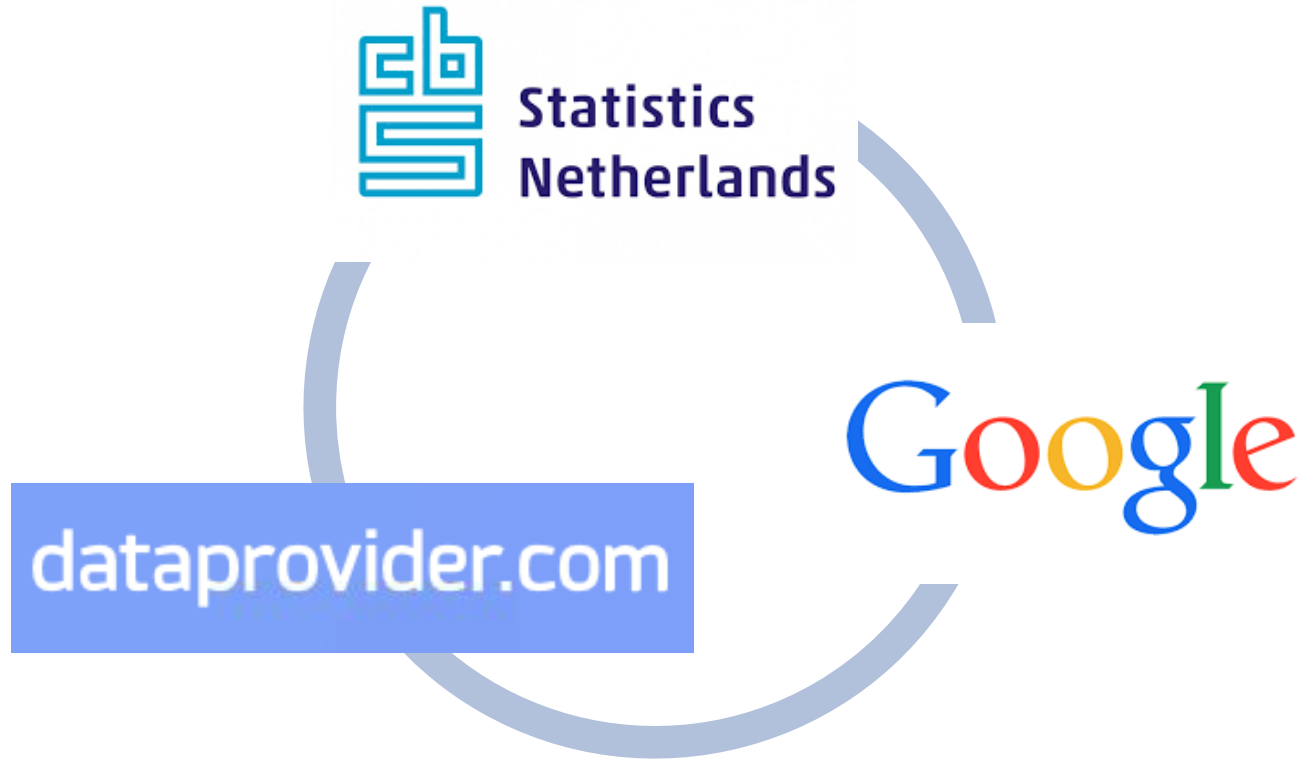
Contents

1. A little bit of background
2. Measuring the internet economy in the Netherlands
 - a) Data sources
 - b) Categories of the internet economy
 - c) Checks using the General Business Register
3. Other big data applications by Statistics NL
4. Questions

Contents

1. A little bit of background
2. Measuring the internet economy in the Netherlands
 - a) Data sources
 - b) Categories of the internet economy
 - c) Checks using the General Business Register
3. Other big data applications by Statistics NL
4. Questions

Three-way partnership



Aim of the study

Main research question:

“What is the importance of the internet economy to the Dutch economy?”

The aim of the research project was fourfold:

1. Determine a pragmatic definition of “the internet economy”;
2. Show the importance and size of the internet economy in the Netherlands;
3. Show the possibilities of new measurement methods with big data;
4. Explain differences from regular statistics/concepts.

Contents

1. A little bit of background
2. Measuring the internet economy in the Netherlands
 - a) Data sources
 - b) Categories of the internet economy
 - c) Checks using the General Business Register
3. Other big data applications by Statistics NL
4. Questions

Dataprovider dataset: 2,6 million Dutch websites

Business information

- Country, address, company name, Chamber of Commerce number, tax number, phone number, e-mail,

eCommerce

- eCommerce probability, shopping cart software, delivery services, payment methods, products, prices,...

Content

- Title, description, keywords, category, language, author....

Other

- Marketing, social media, links, technical and hosting information, ...

Dutch websites

Dataprovider determines nationality of websites

- .nl is always Dutch
- .com with Dutch language: Dutch
- .com but no Dutch language: Dutch if hosted in the Netherlands and Dutch address or phonenumber

Challenge for Latin America!

Call-to-action dataset

- Interaction with website: order, buy, view shopping cart, reservation/booking, subscribe, register.



Business statistics

General Business Register (GBR)

- Backbone of all business statistics based on Chamber of Commerce Trade Register

Production Statistics (PS)

- Financial position, employment, sales

Value Added Tax (VAT)

- Production value, value added, employment, fte

Other

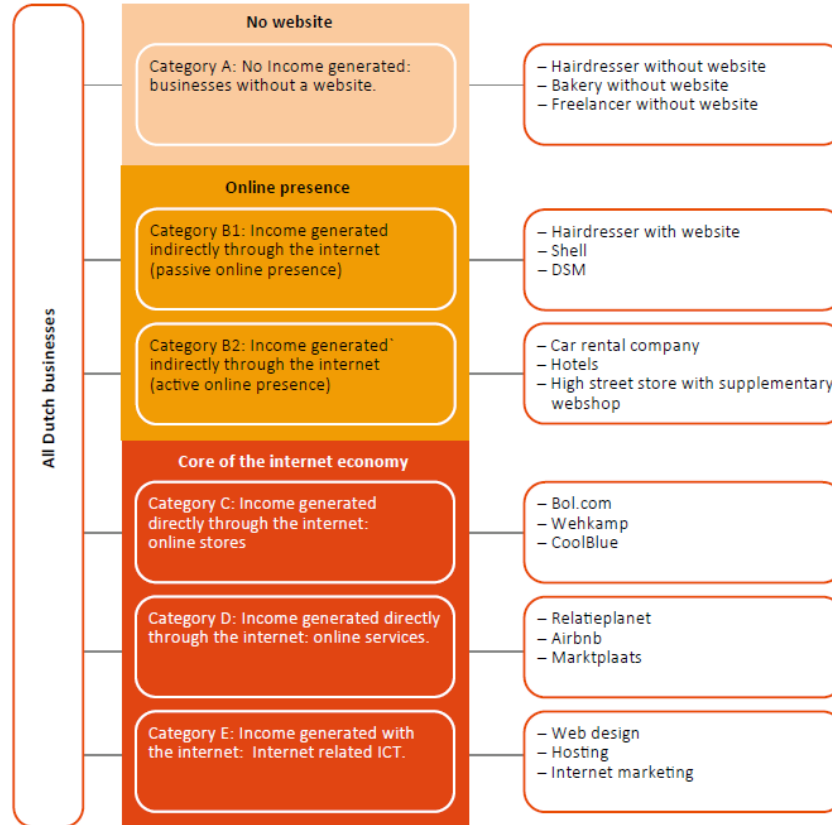
- Turnover statistics, regionalisation of businesses, retail survey for online store turnover, ICT survey

Definition of the internet economy

Is the categorisation on the next sheet applicable to the Colombian situation?

I will come back to this question at the end of the presentation.

Definition of the internet economy



Definition of the internet economy

Category C: e-commerce indicator and Dataprovider variables

Category D & E: keywords followed by manual adjustments

Category B: distinction active or passive online presence using CTA

Keywords always have to be renewed for new countries!

What is excluded?

Categorisation based on data availability.

What is missing?

- Consumer-to-consumer activity: eBay
- Websites that are not linked to
- Facebook or some other social media do not allow indexing

Linking to General Business Register (GBR)

Merging using Chamber of Commerce number, name, address, hostname, e-mail and telephone number

Several challenges

- Businesses with multiple websites in different categories
- Hierarchy: core internet > online presence
- Greatest online footprint or highest employment
- Manual inspection of 100 largest business per category

Linking to General Business Register (GBR)

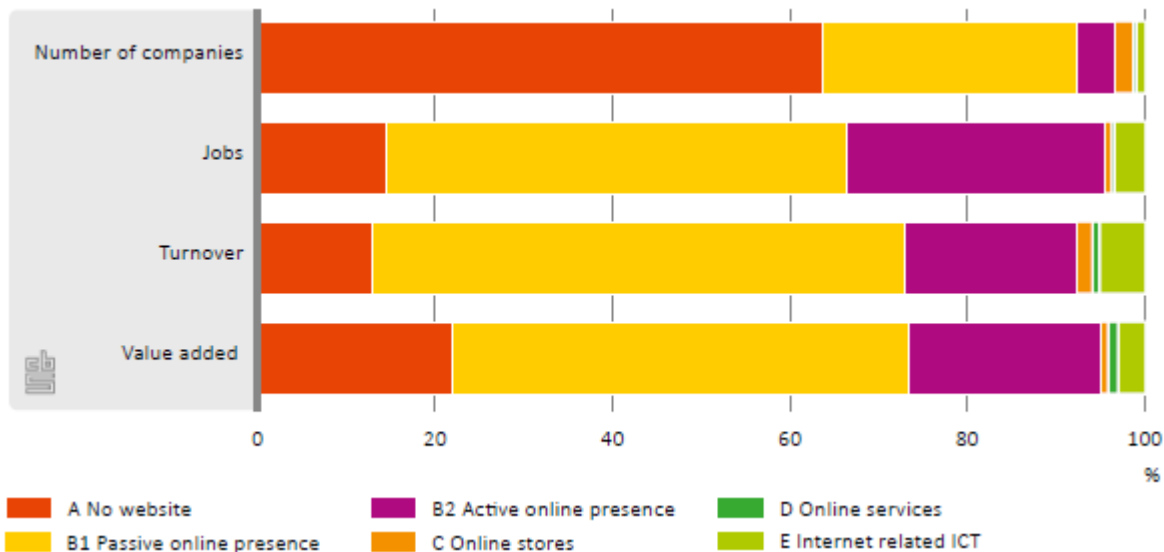
Dual purpose:

1. Helps in internet economy categorisation
2. Backbone for all business statistics

Core of the internet economy:

- 50 000 companies (3,3%)
- € 104 billion turnover (7,7%)
- 345 000 jobs (4,4%)

Linking to General Business Register (GBR)



Discussion

Strengths:

- Combination of big data and microdata
- Economy activity outside of SIC

Limitations

- Data does not cover the entire internet: e.g. Facebook businesses
- Categorisation completely data-driven, not from theory
- Link to GBR: not every website to correct business and multiple websites

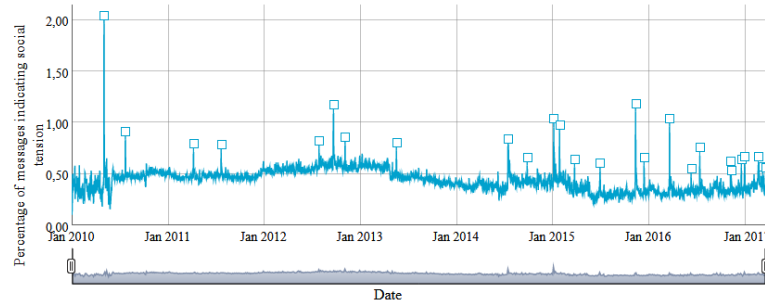
Contents

1. A little bit of background
2. Measuring the internet economy in the Netherlands
 - a) Data sources
 - b) Categories of the internet economy
 - c) Checks using the General Business Register
3. Other big data applications by Statistics NL
4. Questions

Traffic detection loops

- **Traffic loops:** 20 thousand sensors in main roads; 40 thousand in secondary roads
- **Volume:** around 230 million records per day, more than 200 TB in total (*and counting...*)
- **Source:** National Data Warehouse for Traffic Information (NDW)
- **Regular statistics:** Traffic index first released July 2015

Social tension indicator based on Twitter



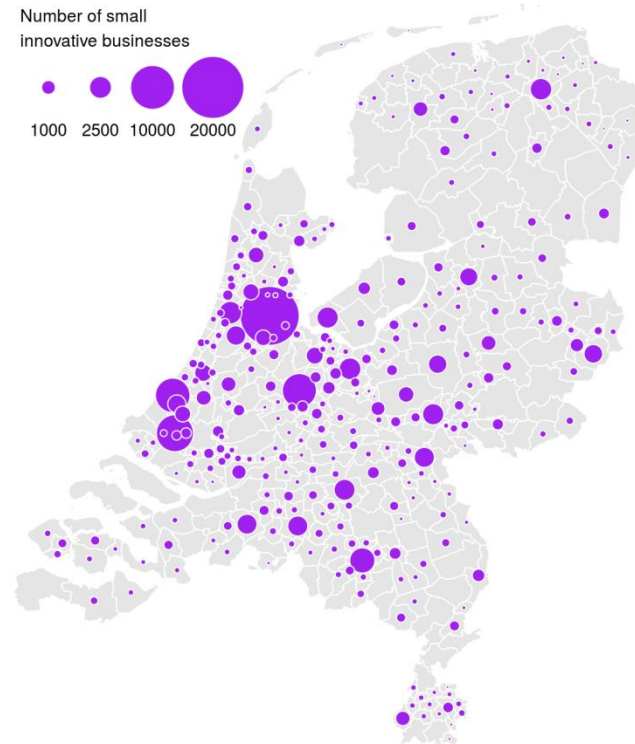
High degree of social tension after Dam Screamer incident and terrorist attacks

This visualisation shows peaks in the social tension indicator. A larger number of messages were posted on or just after the days on which incidents took place which created feelings of unsafety and unrest. The big spike in 2010, for example, is related to the disruption of national Remembrance Day commemorations by the 'Dam Screamer' at Amsterdam Dam Square on 4 May. People's responses to terrorist attacks are also reflected in the social tension indicator: the terrorist attacks in Paris (13 November 2015) and Brussels (22 March 2016) caused a peak in tensions in the Netherlands. In addition, the MH17 disaster (17 July 2014) resulted in strong feelings of unsafety and unrest. Other types of events such as the election of Donald Trump in the United States on 9 November 2016 caused social tensions as well.

Social tensions slightly down in recent years

According to the indicator, overall social tension increased slightly between 2010 and 2013, followed by a slight decrease; this trend can be compared with the results from the annual Safety Monitor, for which a representative group

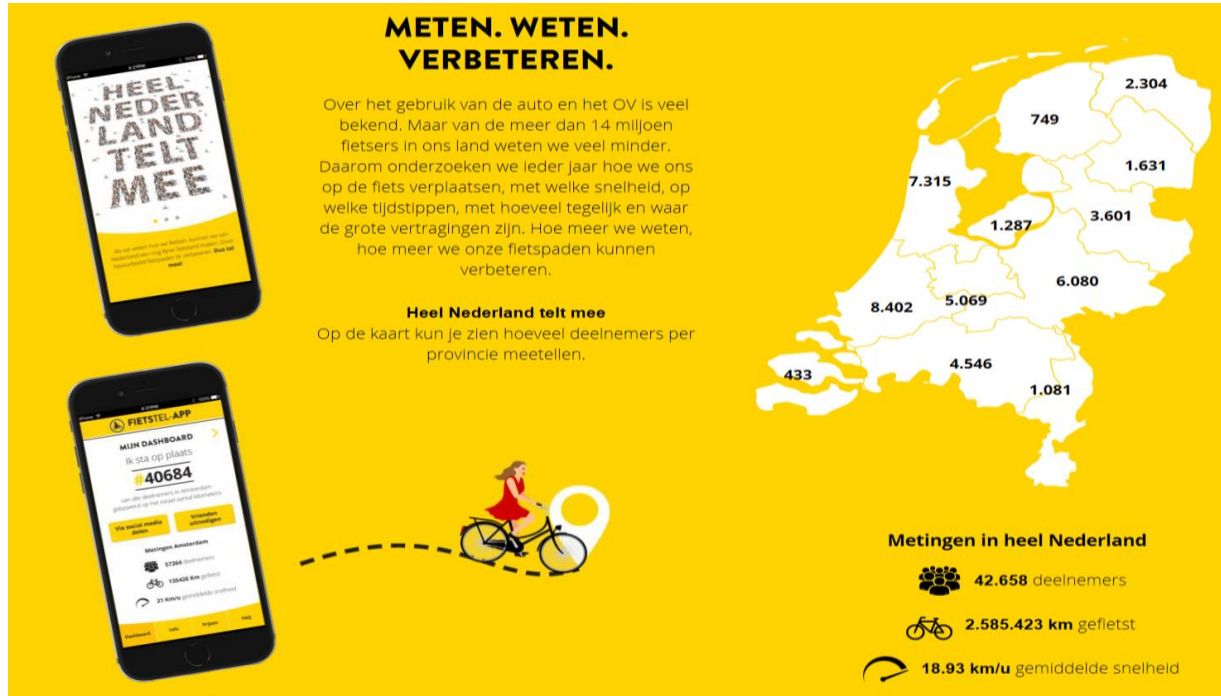
Measuring business innovation



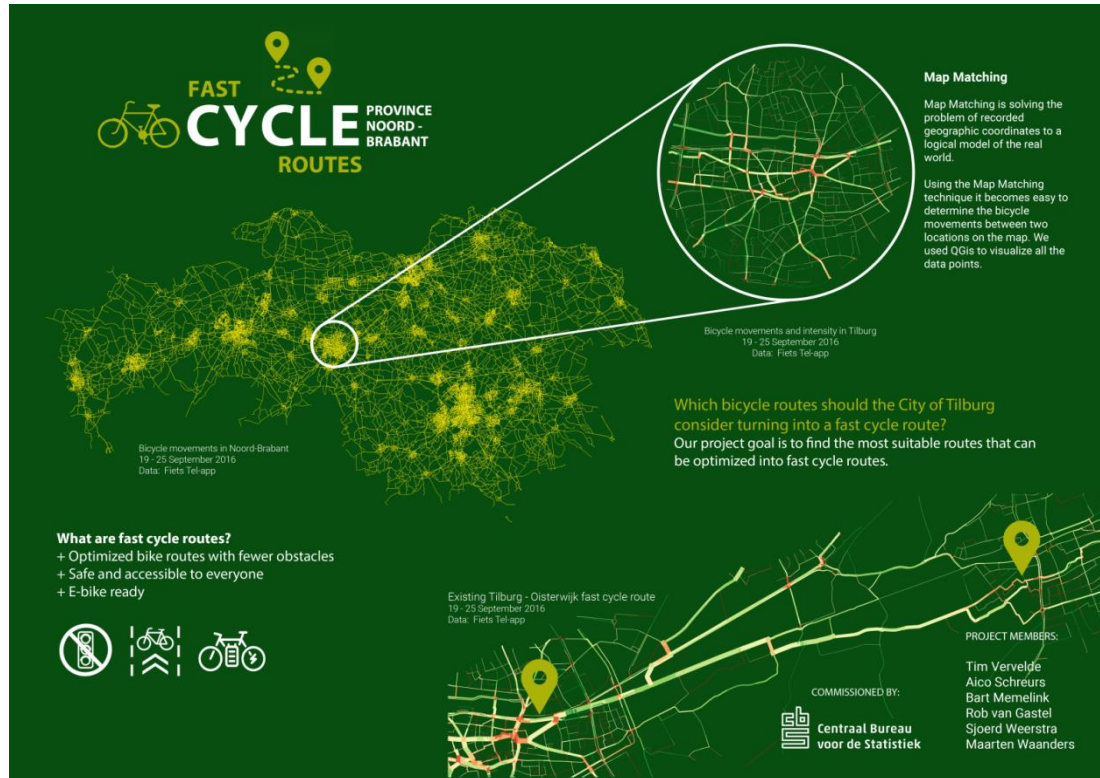
Bicycle count week into statistics



Bicycle count week into statistics



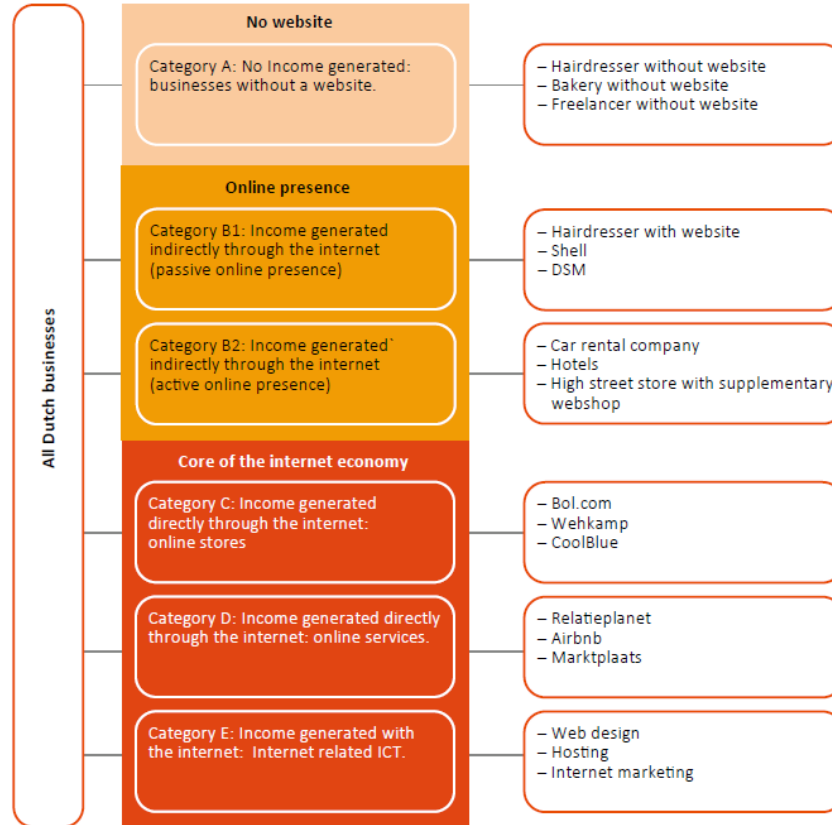
Bicycle count week into statistics



Contents

1. A little bit of background
2. Measuring the internet economy in the Netherlands
 - a) Data sources
 - b) Categories of the internet economy
 - c) Checks using the General Business Register
3. Other big data applications by Statistics NL
4. Questions

Definition of the internet economy



Questions?

