



The use of alternative data sources in business statistics in the Netherlands: experiences and lessons learned

Third meeting of the UN Network of Economic Statisticians

Maarten Alders – director of Business Statistics The Hague – Statistics Netherlands

26 September 2023

Main topics in this presentation

1. What are alternative data sources?
2. Receiving and inspecting the data
3. Processing the data
4. Next steps
5. Summary of pros and cons



1. What are alternative data sources?

In this presentation: any dataset that is not the result of a survey

Catalogue of data sources (234 in use!).

Different types of data sources:

- Administrative data sources
- Registrations
- Scanner data
- 'Big data'
-




1. Catalogue of data sources

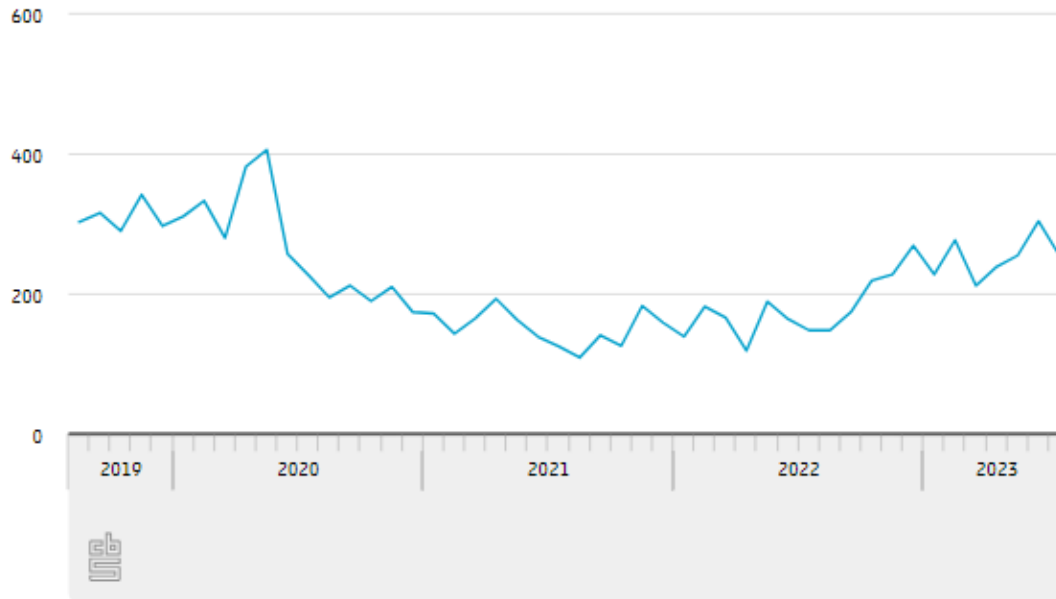
Some examples:

1. Cleared goods at seaports (Import Control System, Customs)
2. Sales of real estate: price, address (Land Registry Office)
3. Turnover tax and Payroll taxes (Tax administration)
4. Subsidies received by companies (Dutch Enterprise Agency)
5. Income and social benefit of all people (UWV)
6. Scanner data: turnover and sold articles (Supermarkets)
7. Bankruptcy (Dutch courts)



1. An easy example: bankruptcies

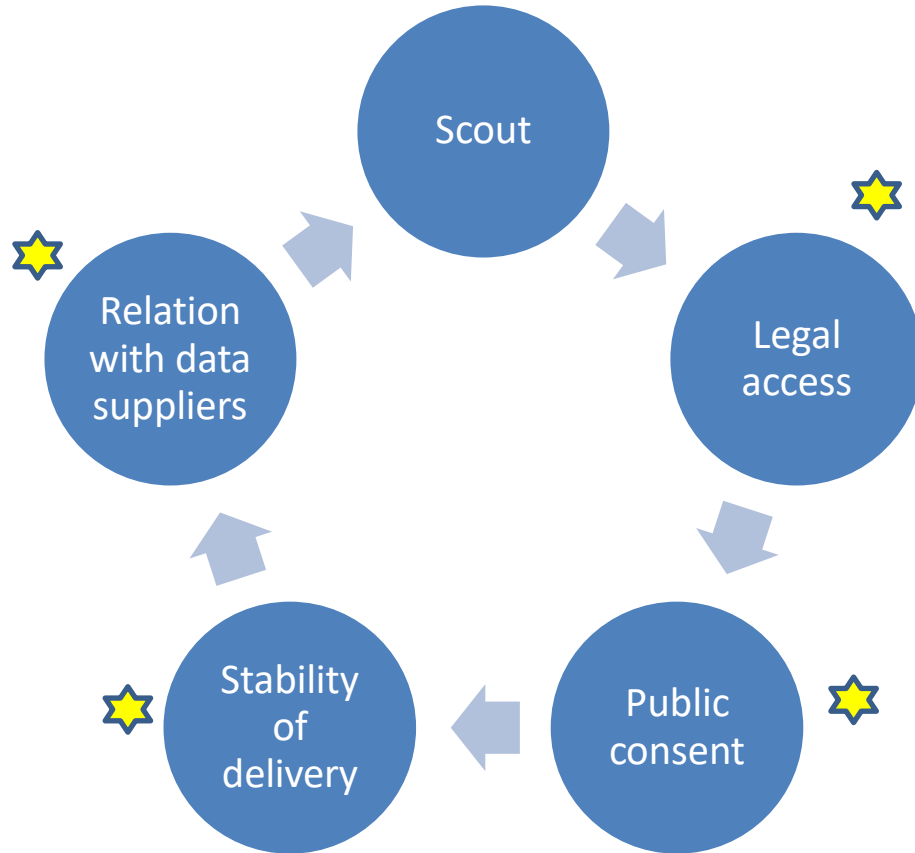
Number of bankruptcies of businesses and institutions (including sole proprietorships) adjusted for court sessions days 



Data supplier:
Dutch Courts



2. Receiving and inspecting the data



Potential issues:

- changes in legislation
- disagreement about data use
- hampering data deliveries
- sudden changes in the data (units, variables)



2. Receiving and inspecting the data

1. Legal access

The Statistics Netherlands Act of 1996 stipulated that:

CBS is required to reduce the administrative burden (of survey taking) to a minimum

and

permitted CBS to access key government registers for statistical purposes

2. Building relationships

3. Inspect the data: metadata and data



3. Processing the data

From actual to target ...:

- Population (issue: duplications, over/under coverage)
- Variables (issue: validity)
- Periods (issue: source too late, different periods)

Linking to business register (issue: different unit types)

- Unique identifier required



3. More about the business register

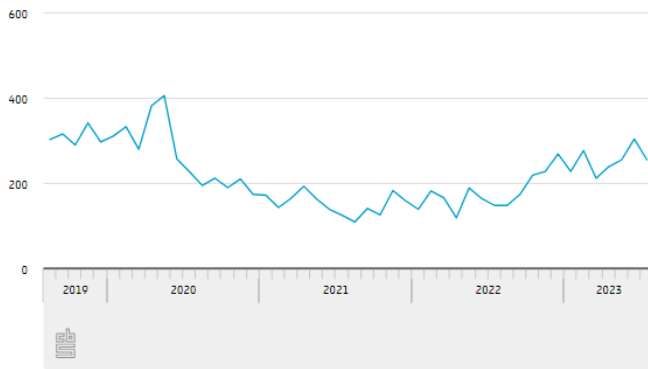
- Mandatory use of statistical units and their characteristics for economic statistics
- Small/medium enterprises: administrative data is leading
- Largest and most complex enterprises: survey data
- Data and structure from the largest EGs is made consistent for several economic variables
- Reporting facilities



3. Let's go back to bankruptcies

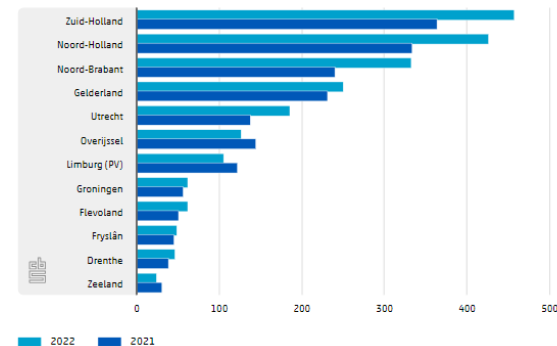
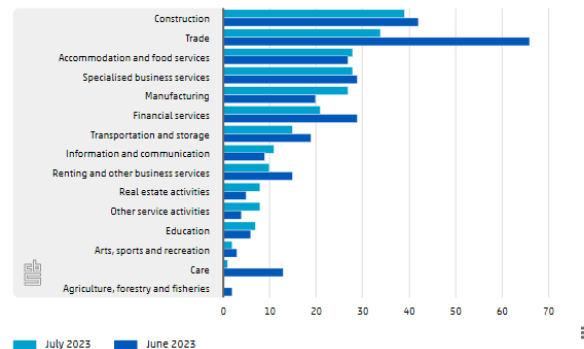
Before linking with business register:

Number of bankruptcies of businesses and institutions (including sole proprietorships) adjusted for court sessions days



Additional information after linking:

Number of bankruptcies of businesses and institutions (including sole proprietorships) by sector



4. Next steps

- More and more data sources, including ‘big data sources’
- More privately held data (i.e. telecom data)
- System to system data collection
- Machine learning techniques

New issues arise:

- Legislation on privately held data
- Data sources even less ‘designed’ for statistics than administrative data



4. Inflation rates and scanner data

Manual retail price observations discontinued

13/01/2020 14:00 / Author: Miriam van der Sangen/Masja de Ree



© Hollandse Hoogte / Frank de Roo

Over the past century, around one hundred interviewers would head out each month on behalf of Statistics Netherlands (CBS) to record the prices of products at 7,000 different shops around the country. These activities have been phased out in recent years. As of 1 January 2020, manual retail price observations are now a thing of the past. CBS now

4. Using website texts to detect innovative companies

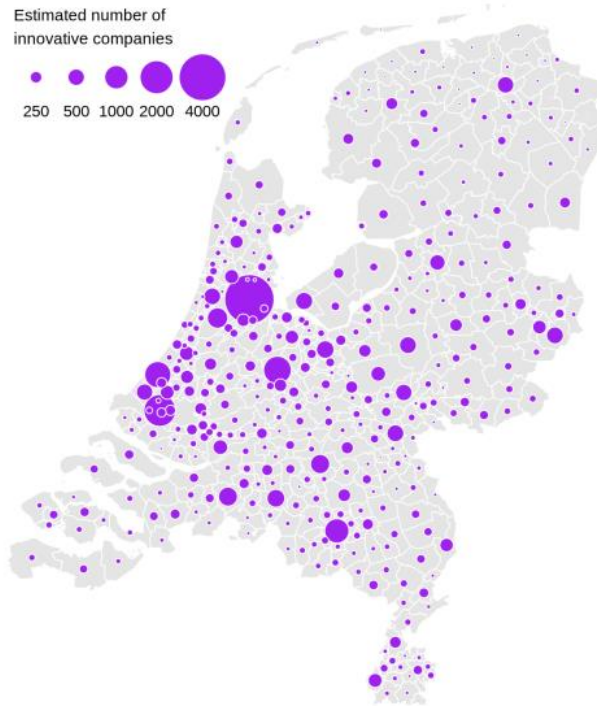


Figure 2. Estimated number of innovative companies with 2 or more working persons in the Netherlands at the municipality level. A total of 52,875 companies are shown.

Source: [Using Website texts to detect Innovative Companies \(cbs.nl\)](#)

2	Combining and processing data	6
2.1	Survey data	6
2.2	URL finding	6
2.3	Webscraping and processing	6
3	Model development	8
3.1	Variable selection	8
3.2	Classification	8
3.3	External validation and application to small companies	10
3.4	Model long-term stability	11



5. A summary of pros ...

- More variables, richer publications
- Reduce data collection costs and response burden
- Increase output granularity
- (More up to date publications)



5. And a summary of cons...

- Suitable legislation required
- Good relation with data owner needed, risk of issues with data delivery
- Alternative data sources have their own purpose: methods are needed to transform these data to statistical output
- Trust in the outcomes is needed: validate the outcomes
- Change in required knowledge



One last thing

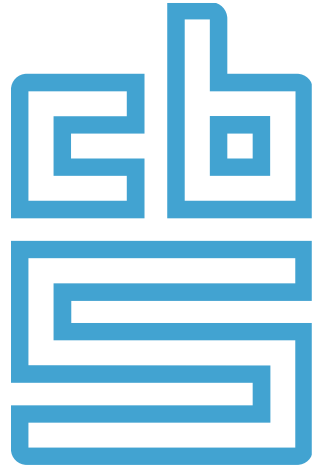
Webinar series on the Global Initiative on Unique Identifiers for Businesses

Organized by: UNSD

UN Committee of Experts on Business & Trade statistics
Global Legal Entity Identifier Foundation (GLEIF)

- **11 October 2023**, 7am – 930am (NY time)
- **7 November 2023**, 7am – 930 am (NY time)
- **19 December 2023**, 7am – 930 am (NY time)





Facts that matter