Preconditions for effective "data science" deployment

Organisational and statistical challenges

Sofie De Broe, Johan van der Valk
17 December 2019
Mission and focus

Using state-of-the-art data science methods and techniques with new and combined datasources to obtain faster, cheaper and better statistical information on burning and relevant policy questions and challenges.
1. Multi-disciplinarity
2. Eco-system
3. Datascouting

<table>
<thead>
<tr>
<th>Bron</th>
<th>Data</th>
<th>Toepassingen</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cargonaut</td>
<td>Aerial cargo in NL</td>
<td>Economy and trade</td>
</tr>
<tr>
<td>Coosto</td>
<td>Social media posts</td>
<td>Desire to move, social tension</td>
</tr>
<tr>
<td>Enologic/Liander</td>
<td>Smart meter data (testset)</td>
<td>Energy consumption</td>
</tr>
<tr>
<td>Tennet</td>
<td>Quantity of current</td>
<td>Energy consumption</td>
</tr>
<tr>
<td>KNMI</td>
<td>Solar radiation data</td>
<td>Solar power production</td>
</tr>
<tr>
<td>Tomtom</td>
<td>Floating Car Data</td>
<td>Traffic intensity</td>
</tr>
<tr>
<td>T-mobile</td>
<td>Signalling data</td>
<td>Movement/displacements</td>
</tr>
</tbody>
</table>
Example: Foreign persons in the Netherlands: Germans (left), Belgians (right) March-April 2018
4. Funding: Deep Solaris Project
5. Relevance: Urban Data Centres
6. Communication

Methods: ML to detect solar panels

Social media and text mining: cyber security

URLs: small innovative companies

Visualisation techniques: dotmaps
7. Research: combining data sources
8. Privacy Preserved Data Sharing and Ethics

![Diagram showing FAIR data station and local data relationships]
Example 1: Analysing combined data within a ‘thrusted secure environment’.

- The set-up of the systems guarantees that both parties will get no insight in the underlying datasets, but are only able to see the results.
9. Implementation

1. Adjust statistical process
2. Software/IT infra
3. Q and stability of the source
4. Train staff
5. ......
10. Questioning output and practises

Are we still making the right statistics?

Can we tackle policy questions nationally?
Are we still making the right statistics?
Half of the EU-region consists of an international border

150 miljoen inhabitants representing 40 procent of the EU territory
Can we tackle policy questions nationally?
Organisational and statistical challenges
Cross-border statistics: an example of new demands that impacts on how we work
Cross-border statistics not available?

➢ Need for comparable data on the border regions: low regional detail and harmonised

➢ **EU data**: Harmonised statistics of Member States with limited regional detail

➢ **National data**: regional detail with specific methods
Production information is challenging

Cross-border statistics are difficult to produce, because they ...

- Require low regional detail,
- using international comparable methods and sources and
- measuring flows of persons & goods

➢ Strong incentive to explore new (BD) sources!
Data portal *Grensdata* for NL Border regions

**Grensdata**

Grensdata is a database with figures over labor market, economy and cooperation for the border regions of the Netherlands with the Dutch provinces Noordrijn-Westfalen, Lower Saxony and the Belgian provinces. You can create graphics and tables, customize and download the data.

**Interreg**

The database results from two projects carried out around the labor market data, which are carried out within the European INTERREG-program for the border regions DE-NL and NL. It is about labor market in border regions DE-NL and Werkveld.

### Date Publication

<table>
<thead>
<tr>
<th>Date</th>
<th>Publication</th>
</tr>
</thead>
<tbody>
<tr>
<td>22-11-2019</td>
<td>Grenspendel werkners: nationaliteit, woonland, werkregio (gemeente)</td>
</tr>
<tr>
<td>15-11-2019</td>
<td>Bevolking; leeftijd, geslacht</td>
</tr>
</tbody>
</table>

**Borderdata portal**
Open data to download

Grensoverschrijdende Open Data

Grenspendel werknemers; nationaliteit, woonland, werkregio (NUTS 3)

Downloads
- Metadata
- Ondervolke dataset
- Dataset voor grafische weergave

Link naar API’s
- Feed (bulk download)
- API (voor Apps)

Meer informatie
- Preview tabel
- Licentie (CC BY 4.0)
- Wat is open data?
- Handleiding odata services
- Handleiding odata voor Excel
- Power pivot

Tabeltoelichting


Deze informatie is samengesteld in het kader van een Interreg-project binnen het programma Duitsland-Nederland en Vlaanderen-Nederland om grensoverschrijdende arbeidsmarkt data beschikbaar te stellen.


Voorstappendoor
Nederland en Duitsland: 2010-2017
België: 2010-2016

Status van de cijfers:
De cijfers in deze tabel zijn definitief

Wijzigingen per 18 oktober 2019
Geen, dit is een nieuwe tabel.
Users vs Producers of cross-border data

Users: Diverse
- In needs
- Different region levels

Producers: Organised
- NSI’s within ESS
- Eurostat, UNECE
SI’s should cover X-border stats

• Methods should be harmonised (inter)nationally
• Is efficient using existing data and expertise
• Equipped for setting up a sustainable system
• Cross-border data is relevant from national perspective
  • Stop depicting a country as an island!
• Fits into mandate of the ESS
  • Cannot rely on Eurostat because of lack of resources and expertise
How should we develop x-border stats?

• Make use of existing data and infrastructures
• Use strategy of constructive collaboration
  • Start with small group of volunteering SI’s
  • Cooperating with stakeholders and relevant organisations
  • Aligned with other initiatives
  • Meet to exchange practices
New demands: change the way we work
Change organisation of producing statistics

➢ Production
  ✓ One source per Statistic -> Multiple sources
  ✓ Micro data available -> not available
  ✓ Few Models -> Many Models

➢ Dissemination
  ➢ Tables & publications -> Open data & Tools & Dashboarding

➢ Different roles within ESS
  ✓ ESS: SI’s have to take more responsibility
  ✓ Eurostat: Coordinating -> Facilitating
Product of statistics: cooperation

Different forms of cooperation is required with

✓ Other SI’s, international organisation
✓ Academics to develop methods
✓ Users (local, regional, national, international)
✓ Data suppliers
✓ Private parties for dissemination
Requiring new skills

➢ New hard skills:
  ✓ Data science
  ✓ Visualisation

➢ New soft skills
  ✓ Communication
  ✓ Relations
  ✓ Collaboration rather than cooperation
### Towards Collaboration

#### From Cooperation to Collaboration

<table>
<thead>
<tr>
<th>Cooperation...</th>
<th>Coordination...</th>
<th>Collaboration...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short term</td>
<td>Longer term</td>
<td>Long term</td>
</tr>
<tr>
<td>Informal Relations</td>
<td>More formal relationships</td>
<td>More pervasive relationship</td>
</tr>
<tr>
<td>No clearly defined mission</td>
<td>Understand mission</td>
<td>Commitment to a common mission</td>
</tr>
<tr>
<td>No defined structure</td>
<td>Focus on a specific effort or program</td>
<td>Results in a new structure</td>
</tr>
<tr>
<td>No planning effort</td>
<td>Some Planning</td>
<td>Comprehensive planning</td>
</tr>
<tr>
<td>Partners share information about the project at hand</td>
<td>Open communication channels</td>
<td>Well defined communication channels at all levels</td>
</tr>
<tr>
<td>Individuals retain authority</td>
<td>Authority still retained by individuals</td>
<td>Collaborative structure determines authority</td>
</tr>
<tr>
<td>Resources are maintained separately</td>
<td>Resources and rewards are shared</td>
<td>Resources are shared</td>
</tr>
<tr>
<td>No Risk</td>
<td>Power can be an issue</td>
<td>Greater risk: power is an issue</td>
</tr>
<tr>
<td>Lower intensity</td>
<td>Some intensity</td>
<td>Higher intensity</td>
</tr>
<tr>
<td>------------------</td>
<td>------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>informal, no goals are defined jointly, no planning together, information is shared as needed.</td>
<td>some planning is required and more communication, thus, a closer working relationship is developed.</td>
<td>working together, having shared commitment and goals, developed in partnership. Leadership, resources, risk, control and results are shared. More accomplished than could have been individually.</td>
</tr>
</tbody>
</table>