New directions for data driven transport safety:
Impact GDPR on road safety policy making

Eric de Kievit | Senior Advisor Transport & Traffic Research | Mobility & Public Space
Manchester, Wednesday November 21th, 2018
Dilemma

- Due to *underregistration*, reliable data on road traffic accidents and casualties has been deteriorating:
  - Fatalities \( \approx 15\% \)
  - Serious Injuries \( \approx 50\% \)
  - Light injuries \( \approx 95\% \)

- Therefore search for coupling with new data sources has become fashionable:
  - Public Record (fatalities)
  - Hospital data
  - Ambulance data

- ‘privacy’ issues lead to loss of information for policy making
What is privacy?

**Who I Am**

Eric de Kievit
10-07-1963

**What I Am**

BSN
ID-papers
Biometrics
Identifier / token

City of Amsterdam
rijsuniversiteit groningen
gemeente Baarn
Personal Data according to GDPR

- any information relating to an identified or identifiable natural person (‘data subject’)

- the identification or authentication of a natural person is unique

- directly or indirectly, in particular by reference to an identifier (i.e. name, address, location, number)
### Database relations

<table>
<thead>
<tr>
<th>customer_id</th>
<th>l_name</th>
<th>f_name</th>
<th>street</th>
<th>p_code</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Bird</td>
<td>Big</td>
<td>Sesame Street 2</td>
<td>1234 AB</td>
</tr>
<tr>
<td>2</td>
<td>Monster</td>
<td>Cookie</td>
<td>Sesame Street 15</td>
<td>1234 AB</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>invoice_id</th>
<th>customer_id</th>
<th>amount</th>
<th>product_id</th>
<th>customer_id</th>
<th>amount</th>
<th>product_id</th>
<th>description</th>
<th>amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>2017_1</td>
<td>1</td>
<td>15</td>
<td>1234</td>
<td>1</td>
<td>15</td>
<td>1234</td>
<td>Bird Flu pill</td>
<td>15</td>
</tr>
<tr>
<td>2017_2</td>
<td>452</td>
<td>200</td>
<td>1235</td>
<td>1</td>
<td>15</td>
<td>1234</td>
<td>Bird Flu pill</td>
<td>15</td>
</tr>
</tbody>
</table>
3 relevant parties according to GDPR

1. ‘Data subject’; identified or identifiable natural person

2. Controller; determines the purposes and means of the processing of personal data

3. Processor; processes personal data on behalf of the controller
GDPR Logic

What are our intentions?

Do we have a common ground?

Thoughtful use of data?

Justifiable goal

1. based on the data subject's consent
2. necessity for the performance of a contract
3. legal basis or a legislative measure
4. vital interest of another natural person
5. performance of a task carried out in the public interest or in the exercise of official authority

Formal requirements GDPR

1. lawfulness, fairness and transparency
2. purpose limitation
3. data minimisation
4. accuracy
5. storage limitation
6. integrity and confidentiality
demand for injury data

For every accident:
- Location and road features;
- Date and time;
- Injury severity (K+SI)
  - #casualties (K+SI)
  - #partners or objects involved;

Per party involved:
- Mode of transport / Object type (tree, wall, guiderail…)
- Age and gender
- Accident type and manoeuvre;
- Some circumstances;

And per road authority (municipality, province, etc) statistical aggregates:
- #casualties by severity, gender and age-class;
- #accidents with alcohol abuse
Into practice:

BRON historically contains 63 features related to car accidents.

A complete BRON record consists of data about:
- Registered traffic accidents with the involved parties and the victims per accident;
- Vehicle detail features;
- The road network of the Netherlands
Red = removed
Blue = new
Black = unchanged

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>DATUM_VKL</td>
<td>Date Accident</td>
</tr>
<tr>
<td>DAG_CODE</td>
<td>Day of the week of Accident</td>
</tr>
<tr>
<td>MND_NUMMER</td>
<td>Month of Accident</td>
</tr>
<tr>
<td>JAAR_VKL</td>
<td>Year of Accident</td>
</tr>
<tr>
<td>TIJDSTIP</td>
<td>Time of Accident</td>
</tr>
<tr>
<td>UUR</td>
<td>Hour</td>
</tr>
<tr>
<td>DDL_ID</td>
<td>Part of the Day</td>
</tr>
<tr>
<td>AP3_CODE</td>
<td>Termination of Accident</td>
</tr>
<tr>
<td>AP4_CODE</td>
<td>Termination of Accident</td>
</tr>
<tr>
<td>AP5_CODE</td>
<td>Termination of Accident</td>
</tr>
<tr>
<td>ANTL_SLA</td>
<td>Amount of Victims</td>
</tr>
<tr>
<td>ANTL_DOD</td>
<td>Amount of Victims/killed</td>
</tr>
<tr>
<td>ANTL_GZH</td>
<td>Amount of Victims/hospitalized</td>
</tr>
<tr>
<td>ANTL_SEH</td>
<td>Amount of Victims/first aid</td>
</tr>
<tr>
<td>ANTL_GOV</td>
<td>Amount of Victims/otherwise injured</td>
</tr>
<tr>
<td>MNE_CODE</td>
<td>Manoeuvre</td>
</tr>
<tr>
<td>DAGTYPE</td>
<td>DAYTYPE &quot;MO-FRI&quot; of &quot;SA-SO&quot;</td>
</tr>
</tbody>
</table>
**Road authority supply after GRPD for injury data (continued)**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>GEBDAT</td>
<td>Date of Birth</td>
</tr>
<tr>
<td>LEEFTIJD</td>
<td>Age</td>
</tr>
<tr>
<td>LKE_ID</td>
<td>Age Category</td>
</tr>
<tr>
<td>GESLACHT</td>
<td>Male or Female</td>
</tr>
<tr>
<td>BLAASTEST</td>
<td>Was there a breath test</td>
</tr>
<tr>
<td>ART8</td>
<td>Use of Alcohol</td>
</tr>
<tr>
<td>MEDICGEBR</td>
<td>Use of Medicine</td>
</tr>
<tr>
<td>GEBJAAR</td>
<td>Year of Birth</td>
</tr>
<tr>
<td>NTT_CODE_B</td>
<td>Nationality of driver</td>
</tr>
<tr>
<td>NTT_CODE_v</td>
<td>Nationality of vehicle</td>
</tr>
<tr>
<td>TDT_ID_1</td>
<td>Cause of Accident1</td>
</tr>
<tr>
<td>TDT_ID_2</td>
<td>Cause of Accident2</td>
</tr>
<tr>
<td>TDT_ID_3</td>
<td>Cause of Accident3</td>
</tr>
<tr>
<td>TDT_AN</td>
<td>Cause of Accident - others</td>
</tr>
</tbody>
</table>

Red = removed
Black = unchanged

Table on vehicle characteristics (details and status) fully removed
Use of Ambulance data

- BRON has bias towards involvement of motorized traffic (Insurance; Legal issues)
- Systematic underreporting of bicycle accidents, therefore …

<table>
<thead>
<tr>
<th>Kenmerken van ongeval</th>
<th>Ambulance-geregistreerde slachtoffers van fietsongevallen in 50km/uur-straten in Amsterdam (N=2824)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Aantal</td>
</tr>
<tr>
<td><strong>Tijdstip</strong></td>
<td></td>
</tr>
<tr>
<td>1-6 uur</td>
<td>363</td>
</tr>
<tr>
<td>7-12 uur</td>
<td>759</td>
</tr>
<tr>
<td>13-18 uur</td>
<td>1.172</td>
</tr>
<tr>
<td>19-24 uur</td>
<td>530</td>
</tr>
<tr>
<td><strong>Dag van de week</strong></td>
<td></td>
</tr>
<tr>
<td>Zondag</td>
<td>302</td>
</tr>
<tr>
<td>Maandag</td>
<td>416</td>
</tr>
<tr>
<td>Dinsdag</td>
<td>385</td>
</tr>
<tr>
<td>Woensdag</td>
<td>419</td>
</tr>
<tr>
<td>Donderdag</td>
<td>425</td>
</tr>
<tr>
<td>Vrijdag</td>
<td>435</td>
</tr>
<tr>
<td>Zaterdag</td>
<td>442</td>
</tr>
<tr>
<td><strong>Naar ziekenhuis vervoerd</strong></td>
<td></td>
</tr>
<tr>
<td>Nee, gering letsel</td>
<td>1.268</td>
</tr>
<tr>
<td>Ja, ernstiger letsel</td>
<td>1.556</td>
</tr>
</tbody>
</table>

- Improving of VRU accident data & -information, especially for accidents with low level of injury seriousness

45% was not transported to hospital
Example of open ambulance data (health service province Flevoland)

https://www.eengezonderflevoland.nl/cijfers/verkeersongevallen-flevoland/
Conclusions

- The introduction of GRDP leads to “information poverty”

- It’s harder to develop tailor made road safety policies because of lack of accurate data

- The use of ambulance data can add value, but only on aggregate level
Questions

- Is a person killed in traffic “a natural person” in the sense of GRPD?

- Is reporting of traffic accidents by (social) media not a more serious threat to our privacy?
The registration process; ‘processing’

**Receiving**
- Remove information in advance that is not necessary for producing BRON
- Security measurements

**Process**
Procedures for storage and management
- Secure access
- Storage terms

**Publication (3 levels)**
Layered Publication
- Open publication: BRON stripped from privacy sensitive information
- Road Administrator: Aggregated BRON
- Scientific:
  - complete BRON

**Diagram**
- Additional Sources
  - input
- Process
  - output
- Storage Database
  - output
- BRON
- Collect
  - input
Level 1: Open data

*Intended for the general public, disclosing no details at all, but aggregated information only:*  

No exact dates/times of accident but only the reported year on annual basis;

Of those involved only “mode of transport” but no “human-features” such as age or even age-groups;

Type of Accident only but no details on manoeuvres.
Levels of publication

Level 2: Road Authorities

Some level of detail in order to monitor policies and design measures.

This version discloses per accident:

- Location and road features;
- Date and time;
- Injury severity (K+SI);
- #casualties (K+SI);
- #partners or objects involved;

Per party involved:

- Mode of transport / Object type (tree, wall, guiderail...);
- Age and gender of driver;
- Accident type and manoeuvre;
- Some circumstances;

And per unit (municipality, province, etc) statistical aggregates:

- #casualties by severity, gender and age-class;
- #accidents with alcohol abuse
Levels of publication

Level 3: Researchers

*Intended for researchers such as SWOV Institute for Road Safety Research:*

*All data provided*

*Use only by signed agreement stating a.o.*:
  - No disclosure of records;
  - Presentations only at levels of publication;
  - Reports in general terms and with aggregated data only;
  - “Raw” material to be destroyed after use