Open Transport Data as an enabler for Multimodal Route Planning

Pieter Colpaert¹, Mathias Van Compernolle², Nils Walravens³, Ruben Verborgh¹, Peter Mechant², Erik Mannens¹, Lieven De Marez², Pieter Ballon¹, Rik Van de Walle¹

1. iMinds – Ghent University – Data Science Lab: pieter.colpaert@ugent.be
2. iMinds – Ghent University – MICT: mathias.vancompernolle@ugent.be
3. iMinds – VUB – SMIT: nils.walravens@vub.ac.be

Abstract
The Public Sector Information (PSI) directive has made Open Data the default within European Public Sector Bodies. End-user multimodal planners need access to government data to make intelligent route planning decisions. In this paper, we study both the needs of the market and the vision of the department of Mobility and Public Works in Flanders using a qualitative research approach (in-depth interviews with 6 market players and 27 governmental data owners, 2 workshops). We found a moderate willingness to publish open data, amongst others, thanks to the European PSI and ITS directives. Furthermore, we found evidence of existing open data reuse among commercial multimodal route planners. We identified 3 caveats: (i) not every dataset will be reused as there is a cost for adoption, (ii) data quality needs to be high enough and (iii) metadata is crucial. We formulate opportunities that lie within the Web's principles to reduce cost of publishing and reusing, and to raise the data quality.

Keywords:
Open Data, e-government, multimodal route planning

Introduction
In the last decade, “open by default” was implemented in many European member states thanks to the Public Sector Information directive. Flanders, a region in Belgium, is not an exception formalising the PSI directive in the concept note of 23d of September 2011¹. Open by default means that, when creating a new governmental dataset, publishing it under an open license becomes the default process. Only when a good reason can be given not to open the dataset up, the dataset can remain unshared or closed. The departments of the Flemish government, among which also the Department of Mobility and Public Works, has started to slowly open up datasets.

¹ https://www.bestuurszaken.be/open-data-bij-de-vlaamse-overheid
Open Transport Data as an enabler for Intelligent Transport Systems

Multimodal route planning is the act of taking into account multiple modes of transport when providing route planning advice. In order to compare between different transport modes, and in order to fulfil the many needs of end-users, a huge amount of data is needed, coming from various sources [1]. Often, these different sources do not take into account that they can be useful for multimodal route planning and are thus published in a format that is not suitable for using them as a basis to provide multimodal route planning advice. For commercial route planning applications, this entails costs in terms of harmonizing certain datasets world-wide.

In this paper, we study which datasets are already used for commercial route planning applications today, and research the needs of the involved market players. Furthermore, we interviewed data owners and directors of different agencies of the department of Mobility and Public Works in Flanders.

Related work
It is early days for (Linked) Open Data within transport data. In 2009 in the UK, where standardization with NAPTAN came a long way, a Linked Data mapping was done of one export. This data dump is still available. We can only assume this was a proof of concept which has now been abandoned.

Within road transport, the DATEXII data exchange format for road events exists. In order to publish transport data, the DATEXII exchange format can also be used as an open data format, which is being done in Flanders. This DATEXII vocabulary has been reused in the Transport Disruption Ontology [2], which is a Linked Data specification for road events, and their impact on the plans travellers would have.

Within Public Transit, the General Transit Feed Specification (GTFS) has become the de facto standard to publish transit schedules. This specification has already been translated into Linked GTFS, which requires more effort for data publisher for persistent identifiers for their schedules' data. GTFS can also be transformed into the actual connections (a departure time, departure location, arrival time and arrival location). A Linked Data specifications for these connections is called Linked Connections [1] and allows multi-modal route planning directly on top of the published data.

In the Share PSI project, a best practise was discussed to rank government datasets using a traffic light system (red-yellow-green), to indicate whether they are ready to become Open Data. We have applied this as well by defining a green dataset as a dataset that can be published without further ado, yellow as a dataset where there is one thing lacking before the dataset can be published, and red as all other datasets.

Method
Open Transport Data as an enabler for Intelligent Transport Systems

A team of interdisciplinary researchers with backgrounds in public administration and communication, business modelling and software engineering, used the ESTEEM participation method [3] to interview all the organisations of the department of Mobility and Public Works about their current view on Open Data and about the needs of data reusers.

Interviewing governmental organisations
Belgium is divided into three regions: Flanders, the Walloon region and Brussels. Each region has its own regional government, which has its own Department of Mobility and Public Works. In Flanders, this department is responsible for facilitating, among others, the water ways, the airports, the highways and the bus company in Flanders. This government department maintains many datasets ranging from geospatial basic registries, to real-time traffic sensor data, to indicators for evidence based decision making. While at the time of writing only the traffic on the highways and traffic indicators are published as open data, we interviewed 12 directors and 15 data maintainers about their vision on Open Data. During these interviews, we captured unresolved questions regarding data publishing and gathered a list of to be published datasets.

Interviewing the market
We have identified six market players within our reach for an interview: Be-Mobile, Prophets, Ally, FlowPilots, Infofarm and Google. Each of these parties reuse, or is planning to reuse, governmental datasets for a use case of multimodal route planning. The questions followed a strict template (see Table 1).

| Table 1 – The template for the interviews with the 6 market players that might reuse Open Government Data for the use case of multimodal route planning |
|-------------------------------------------------|----------------|----------------|----------------|----------------|
|                                                   | Attitude | Bottlenecks | Expectations | Requirements |
| Internal                                           |          |             |              |              |
| External                                           |          |             |              |              |
| Technological                                     |          |             |              |              |

Workshops
We have organised two workshops: one to validate the outcomes of the interviews with the different governmental organisations, the other to align the market needs with the governmental Open Data roadmap. On the first workshop, we welcomed a representative of each organisation within the Department of Mobility and Public Works that we had already met during a one on one interview. We
Open Transport Data as an enabler for Intelligent Transport Systems

had an introductory programme where we summarised the basics of an Open Data policy: the open
definition and the implementation of the PSI directive in Flanders. Furthermore we also gave a short
summary of the results of the interviews with the market stakeholders. In the second half of this
workshop we had three parallel break-out sessions in which we discussed 10 unresolved questions that
came out of the interviews (see further).

In the second workshop, we invited several market player reusing Flemish Open Data, ranging from
companies like We Open Data\(^2\), Be-Mobile and Flow Pilots\(^3\) to stakeholders such as the traveler's
organisation TreinTramBus\(^4\) and the organisation iRail\(^5\), which promotes digital creativity concerning
mobility data in Belgium. As a speaker, we invited CityMapper\(^6\), which outlined what data they need
to create a world-wide multimodal route planner.

Results

Unresolved “publishing Open Data” questions

After the interviews, ten unresolved questions were identified. The questions are ranked by popularity
as discussed during the interviews, as inspired by the ESTEEM method:

1. Should data publishing be centralised or decentralised within the department and what process
should be followed? If decentralised: what about the metadata?
2. How to ensure reusers interpret the data correctly?
3. How to get the right means and knowledge on how to publish open data within our
organisation?
4. How do we know what reusers want?
5. How can we influence what reusers do with the data?
6. How can Open Data help us with evidence based decisions?
7. Should the government be responsible for a certain data quality?
8. How can Open Data raise the government's efficiency?
9. How can we ensure sustainability when a dataset is published once?
10. How do we ensure the technical availability of datasets?

The outcomes of this debate were used to draft of first vision document on Open Data and data
management in general for the department.

Market needs

From the interviews with the market stakeholders, we learned there is willingness to reuse open
governmental datasets for multimodal route planning.

\(^2\) [http://weopendata.com](http://weopendata.com)
\(^3\) [http://www.flowpilots.com/](http://www.flowpilots.com/)
\(^4\) [http://treintrambus.be](http://treintrambus.be)
\(^5\) [http://hello.iRail.be](http://hello.iRail.be)
\(^6\) [http://citymapper.com](http://citymapper.com)
Open Transport Data as an enabler for Intelligent Transport Systems

On the one hand, the market expressed the clear need for Open Data, as defined by the Open Definition. One on one contracts, or having to get in contact with a data owner before being able to access the datasets, are hurdles some are unwilling to take, certainly when trying to reuse as many datasets as possible. On the other hand, it is not because a dataset is opened up, that it will automatically be reused. The uptake of a specific dataset depends on many criteria: sometimes a feature that is already present can be made better with this new dataset, sometimes using the dataset would mean introducing a new end-user feature, in other cases the data is not of a sufficient quality for the specific use case. In order to get guarantees of a certain data quality (e.g., getting a guaranteed availability or being able to retrieve the dataset in a certain specific format), the market is willing to pay for a Service Level Agreement (SLA) on top of the Open Data.

Last but not least, for public transportation, the market expressed interest into reselling tickets through an Application Programming Interface (API). This follows the recent trends towards Mobility as a Service (MaaS) and Open Ticketing.

Catalogue of governmental datasets

We categorised the demand of the market for multimodal route planning in 5 market questions:

1. Where can you drive/walk?
2. How fast are/were you allowed to drive?
3. How fast are/were you able to drive?
4. Where can you park your bike/car?
5. What are the timetables of public transit options?

We have identified datasets that are maintained by the government today, that, in the long run, can be published as Open Data, and summarised them in Table 2. However, these datasets only provide part of an answer to one or more market needs.

<table>
<thead>
<tr>
<th>Description</th>
<th>Roadmap</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traffic events on the Flemish highways maintained by the Flemish Traffic Center</td>
<td>Openly licensed and publicly available: published in both OTAP and Datex2. The core task of the Flemish Traffic Center is limited to the highways. Therefore, it is not planned to include the local roads within this</td>
</tr>
</tbody>
</table>
### Open Transport Data as an enabler for Intelligent Transport Systems

<table>
<thead>
<tr>
<th>Dataset Description</th>
<th>Status</th>
</tr>
</thead>
</table>
| Road database (including parkings) for regional roads maintained by the Agency for Roads and Traffic | Open in Q1 2016  
Already complies to the INSPIRE directive for geospatial data and will be published using WMS, WFS and SHP. |
| Dynamic speed limit signs on the highways maintained by the Traffic Center          | Open Data in ~2016                                                     |
| Location of local dynamic speed limit signs                                         | Open Data in Q1 2016 via the INSPIRE directive                         |
| Loop vehicle detectors on the highways by the Flemish Traffic Center                | On the roadmap to be open as Open Data                                |
| Validated statistics of traffic congestion on the Flemish highways maintained by the Flemish Traffic Center | Opened up through own development. XLS files can be exported.          |
| Kwaliteitsparkings.be                                                                | Website with quality labels given to highway parking facilities. Publicly available, but not openly licensed. |
| Public transit time tables maintained by De Lijn                                    | Planned timetables can be requested through a one-on-one contract.     |

The datasets that can be accessed today, are advertised on an open data page maintained by the department itself at [http://mow.vlaanderen.be/opendata](http://mow.vlaanderen.be/opendata).
Open Transport Data as an enabler for Intelligent Transport Systems

Conclusions and perspectives
Based on our qualitative, interdisciplinary research approach using the ESTEEM participation method we can conclude that a dataset is only going to be reused by the market when the benefits outweigh the costs. These costs can be lowered by Public Administrations by raising the interoperability between datasets worldwide. Sometimes however, extra costs may raise the benefits, such as an optional Service Level Agreement.

In our research we also noted a ‘fear’ in public administrations that opened up data will not be interpreted correctly by third parties. Therefore, providing context and metadata with each dataset is of utmost importance. In Europe a specification exists called the DCAT-AP\(^7\) standard, to promote cross-border metadata sharing. Again, the interoperability of the meaning behind each field within this metadata is of high importance to both data publisher as the reuse market. Furthermore, publishing data for maximized reuse [4] should happen in a cost-efficient manner, as extra funds for opening up datasets are hard to get.

Finally, our research also shows that it is currently unclear how the department will handle demand for datasets that are not yet centralized within the government. For instance, there is no complete database for parking facilities across Flanders, nor a dataset for all traffic signs, nor a database for traffic congestion on both local and regional roads.

The traffic light system (see related work) was hard to implement for our use case: the green-yellow-red system was oversimplifying the complexities of a potential roadmap towards Open Data. For an internal deliverable, we used the colour scheme, yet provided arguments with the roadmap to Open Data after certain investments would be done. Overall, it is too early days for this Open Data policy to conclude whether it is a success or not. In future work, we will describe the outcomes of the project in more detail and provide more conclusive results.

From an academic perspective, raising the interoperability of both metadata and data can be studied on many levels: technological, syntactic, semantic, and querying level of interoperability. The Internet and the Web ensured technological interoperability among the datasets: we can exchange data worldwide. Furthermore, thanks to the standardisation of open Web formats such as HTML, JSON, XML or CSV, we can read these serialisations and syntaxes from different programming languages without problem. The Web, or the HTTP-protocol, and the Representational State Transfer (REST) principles [5], also define ways to raise the semantic interoperability of datasets by using Uniform Resource Identifiers (URI) over HTTP as identifiers. This way, whenever an identifier is used, it can be looked up in a browser. Moreover, the hypermedia principle within REST, specifies a meta-solution that allows self-documenting query services. An experimental framework was introduced, taking into

---

\(^7\) [https://joinup.ec.europa.eu/asset/dcat_application_profile/description](https://joinup.ec.europa.eu/asset/dcat_application_profile/description)
account these principles, for multimodal route planning as Linked Connections [1].

References


