Open Transport Data for maximising reuse in multimodal route planners: a study in Flanders

Pieter Colpaert[1,2], Mathias Van Compernolle[2], Nils Walravens[3], Peter Mechant[2], Jan Adriaenssens[4], Femke Ongenae[2], Ruben Verborgh[2], Erik Mannens[2]

[1]Corresponding author
[2]{firstname}.{lastname}@ugent.be
[3]nils.walravens@vub.ac.be
[4]jan.adriaenssens@imec.be

Abstract:

The European Data Portal shows a growing number of governmental organisations opening up transport data. As end-users need traffic or transit updates on their day to day travels, route planners need access to this government data to make intelligent decisions. Developers however, will not integrate a dataset when the cost for adoption is too high. In this article, we study the internal and technological challenges to publish data from the Department of Transport and Public Works in Flanders for maximum reuse. Using the qualitative ESTEEM research approach, we interviewed 27 governmental data owners and organised both an internal workshop as a matchmaking workshop. In these workshops, data interoperability was discussed on 4 levels: legal, syntactic, semantic and querying. The interviews were summarised in 10 challenges to which possible solutions were formulated. The effort needed to reuse existing public datasets today is high, yet we see the first evidence of datasets being reused in a legally and syntactically interoperable way. Publishing data so that it is reusable in an affordable way is still challenging.

1. Introduction

When asking an audience in Europe how far they live from work, half of them will reply with an answer expressed in time distance (e.g., "20 minutes"). What is an easy unit for humans, appears to be a more difficult data problem for computers. In order to compare between different modes of transport, and in order to fulfil the different requirements and constraints of end-users, the amount of datasets that can contribute to one of many possible answers, are limitless. *Federated route planning* (also referred to as distributed (route) planning) is the act of taking into account multiple modes of transport (multimodal route planning) when providing route planning advice, as well as datasets from sources that do not, such as criminality statistics, wheelchair accessibility or address databases. These route planners only become better when more transport datasets from governmental organisations are available and the ability to federate/distribute queries over multiple data sources increases.

Thanks to the European Public Sector Information (PSI) directive, "open by default" was implemented in European member states [9]. This leads administrations to having to explain why a certain dataset should not be published, rather than having to motivate why it should be open. Moreover, also the Intelligent Transport Systems (ITS) directive helps popularising publicly shar-

ing data, such as with a delegated regulation elaborating on a European Access Point for Truck Parking Data [8]. Finally, also a European directive exists for sharing data within the geospatial domain, called INSPIRE [10]. On the European data portal, at the time of writing, 212,396 transport datasets themselves range from public transport time schedules to road traffic events and locations of parking sites. A study of market players in multimodal route planners [20] showed that the uptake of these datasets is not high. The threshold in the costs/benefits trade-off for when datasets reuse flourishes, remains to be discovered.

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 Transport and Public Works (DTPW). This government department maintains datasets ranging from geospatial base registries, to real-time traffic sensor data, to indicators for evidence based decision making. Its complex organisational structure in a small geographic area makes it an interesting object of research. In this paper, we researched the challenges for lowering reuse cost and for raising benefits when publishing governmental transport data within this region.

The goal of publishing mobility datasets to the Web, is to maximise the adoption in third party systems such as route planning systems, urban planning tools or mobility studies. Today, third parties in other domains reuse public datasets from the Web, yet the adoption of transport datasets published on the European data portals remains limited [2, 20]. Indeed, when studying app stores, there are only few apps today that try to reach a global audience [20]. Other apps are volunteer based apps that work for a limited geographic region, or are apps by the transport agencies themself. Furthermore, it should be easy enough for data consumers to also include other parameters within route planning algorithms, such as accessibility or multimodality. When studying the cost and benefits for these third party developers, we can see that the benefits, yet the cost for adoption can still be lower.

The experiment where an audience was asked how far they live from work, was conducted at several talks of the first author of this paper in Europe, with each time a similar result. We invite the reader to reproduce this.

In the next section, we introduce the term "Data Source Interoperability" in order to introduce a data reuse framework. Next, we discuss the method to study the acceptance of Open Data within governmental organisations, which we based on an existing qualitative method. We study the datasets themself using the interoperability framework, and report on organisational challenges that came from two workshops we organised. Finally, a few recommendations for actions are proposed to the Public Sector Bodies involved.

2. Data source interoperability

Quantifying intoperability has overal been regarded as complex, as it is difficult to agree on what a fully interoperable system exactly is [17, 12, 16]. In order to be able to qualitatively study how easy it would be to add another datasets to an existing system, we introduce the term *data source interoperability*. We define this interoperability as how easy it is to evaluate questions over two different data sources. Problems that occur while querying over these two data sources are called *interoperability problems*, and are illustrated in Figure 1.

The first level is the *legal* level: data consumers must be allowed for these two datasets to be queried together. When for a certain dataset a specific one on one contract needs to be signed before it can be used, the burden for data consumers becomes too high [20]. When two datasets are made available as Open Data, which means it complies with the Open Definition (`http://`
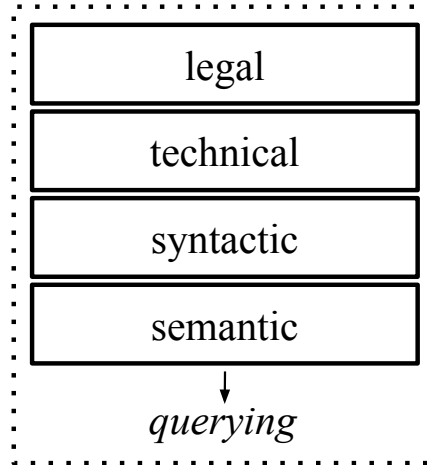
Fig. 1. *The layers of data source interoperability in the context of lowering the cost for the adoption of public datasets, based on existing interoperability models [12]*

`opendefinition.org`) and has an open license attached to it, the interoperability problems will be minimised.

The second level is the *technical* level which entails how easy it is to bring two datasets physically together. Thanks to the Internet, we can assume this is possible today.

The third kind of interoperability describes whether the serialisation/syntax allows for easy merging. On the Web today, commonly open standards are used to serialise data, such as HTML, CSV, XML or JSON. When a program can read these documents into memory using a common library, the *syntactic interoperability* problems have been resolved.

The *semantic interoperability* describes whether the terms/identifiers used do not conflict when bringing two datasets together. In order to overcome this type of problems, Linked Data and the Resource Description Framework (RDF) was introduced. Instead of using local identifiers or words for terms, web addresses are used for every identifier and or word used. This way, when the web address is resolved, the definition can be returned, including links to other interesting resources. Furthermore, in order to be independent of a certain syntax, RDF introduces an abstraction for statements, called *triples*. Each triple can be compared to a small sentence, containing a *subject*, a *predicate* and an *object*.

```
1 <https://pietercolpaert.be/#me> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://
    xmlns.com/foaf/0.1/Person>.
2 <https://pietercolpaert.be/#me> <http://xmlns.com/foaf/0.1/name> "Pieter Colpaert".
3 <https://pietercolpaert.be/papers/iet-otd> <http://www.w3.org/1999/02/22-rdf-syntax-ns#
    type> <http://xmlns.com/foaf/0.1/Document> .
4 <https://pietercolpaert.be/papers/iet-otd> <http://purl.org/dc/terms/creator> <https://
    pietercolpaert.be/#me>.
5 <https://pietercolpaert.be/papers/iet-otd> <http://purl.org/dc/terms/title> "Open
    Transport Data for maximising reuse in multimodal route planners: a study in Flanders"
    .
```

Listing 1. *An example of five triples describing this paper: by using identifiers, the semantic interoperability problems can be resolved faster.*

In Listing 1 an example can be found of five triples describing this paper. As a thought experiment, each of these five statements could be published by five different machines. When hen they would be encountered by user-agents crawling the web, these statement would still be semantically

3

interoperable, as no identifiers will conflict. Just like on the Web of Documents today, everyone can create their own Uniform Resource Identifiers (URIs). It is the responsibility of the maintainer to make sure the URIs are persistent and do not change over time.

```
1  {
2    # Mapping the to be used terms to URIs
3    "@context" : {
4      "Document": "http://xmlns.com/foaf/0.1/Document",
5      "Person": "http://xmlns.com/foaf/0.1/Person",
6      "creator": {
7        "@id": "http://purl.org/dc/terms/creator",
8        "@type": "@id"
9      },
10     "name": "http://xmlns.com/foaf/0.1/name",
11     "title": "http://purl.org/dc/terms/title"
12   },
13   # The actual data
14   "@id": "https://pietercolpaert.be/papers/iet-otd",
15   "@type" : "Document",
16   "title": "Open Transport Data for maximising reuse in multimodal route planners: a study
          in Flanders",
17   "creator": {
18     "@id": "https://pietercolpaert.be/#me",
19     "@type": "Person",
20     "name": "Pieter Colpaert"
21   }
22 }
```

Listing 2. *The same example as Listing 1, now in JSON-LD, illustrating how URIs can be used in more common syntaxes as well.*

It is not the goal to change all data exchange mechanisms to use the triple format as introduced by Listing 1. Instead, existing serialisations can be auto-documented with what terms and elements can be mapped to what URIs. An example for JSON, extended with the JSON-LD format [4], can be found in Listing 2. For XML, RDF/XML exists [3], as well as for HTML, RDFa exists [6] and even for tabular data such as CSV, a metadata specification exists [5]. When no official standard for your serialisation exists, mapping languages can provide the necessary tools to document your dataset with URIs [7].

When the four layers of data source interoperability are achieved, user-agents will not yet be able to also ask questions over the borders of one data source. Two extremes of data interfaces can be identified: on the one hand, *data dumps* can be provided with no server functionality at all, while on the other hand, *query services* allow an infinite amount of questions to be answered on the server of the data publisher. When the publisher exposes a query interface, the questions that can be answered become more specific. The drawback for asking questions over multiple data sources in this case, is the answers to these specific questions are harder to combine with other services, as gradually more specific questions would need to be asked. Furthermore, when the server answers all possible questions for end-users, the servers' availability becomes less cost-efficient or reliable [22]. A trade-off is to be made between these two options, in order to maximise the ability to federate queries while cost-efficiently ensuring high availability of the data source [18, 19].

As a comparison, the Web of Documents today is also a system where data sources publish their knowledge in fragments for the consumption by end-users. End-users do not have to download the entire Web at once, as the Web can be browser by clicking through pages that are downloaded just in time. Trade-offs have already been made to make the information on these pages easily consumable. Representational State Transfer (REST) [24], an architectural style for the Web, defines

constraints for documents and datasets published on the Web for scalability and interoperability. Triple Pattern Fragments interfaces [18] were already suggested as a way to support question answering for RDF data.

In order to come to these five layers, we studied the related work within interoperability frameworks [2, 14, 15, 16, 17]. The Information Modeling and Interoperability (IMI) model [14] has only three levels: the syntax layer, the object layer and the semantic layer. The *syntax layer* is responsible for "dumbing down" object-oriented information into document instances and byte streams. The *object layer*'s goal is to offer applications an object-oriented view on the information that they operate upon. This layer thus defines the hierarchy in the data structures of e.g., an XML or JSON document. The *semantic layer* is the same as the one defined in this paper. The authors argue that each of these three layers should have their own technology to have a fully interoperable service. Today we indeed see that XML syntax has an RDF/XML specification which allows object RDF to be stored within XML. A specific XML stylesheet then defines how the objects should be structured within one file. In our framework, we talk about how data can be fragmented within the querying layer, and how data should be structured within one file is considered to be contained within the syntactic and the semantic layer.

Interoperability problems were also described as integration problems [15, 16]. The goal of an integration process is to provide users with a unified view on one machine. Four types of heterogeneity are commonly discussed. *Implementation heterogeneity* occurs when different data sources run on different hardware and *structural heterogeneity* occurs when data sources have different data models, in the same way as the object level in the IMI model. *Syntax heterogeneity* occurs when data sources have different languages and data representations. Finally, *semantic heterogeneity* occurs when "the conceptualisation of the different data sources is influenced by the designers' view of the concepts and the context to be modeled".

Depending on the specific domain when studying the data source interoperability, different other layers may be specified, such as an *object interoperability*, as within the IMI model, or the *process interoperability*. In order to keep the model generalise, we assume the process interoperability is captured within the semantic interoperability framework, as a different identifier should be used for something resulting from a non-interoperable process. Also the *organisational interoperability* [17, 2] has been the object of research, which focuses on high level problems such as cultural differences and alignment with organisational processes The layers introduced here however are limited to studying the data source interoperability. The organisational aspects are studied separately and are tackled as part of studying the challenges before the data can be published, or before the interoperability itself can be raised.

3. Method

The PSI directive, as well as the INSPIRE and ITS directive, have the goal to maximise the reuse of certain governmental datasets. This policy decision has been taken on a high level, and while the technological aspects of raising the adoption of governmental datasets are listed above, still the change in governmental processes, policy and organisational challenges need to be resolved. Keeping the introduced data source interoperability framework in mind, we reiterate the research question as follows: "what are the organisational challenges for raising the interoperability of (Flemish) government (transport) data?".

An answer to this question is pursued by following the ESTEEM protocol [23]. Its main goal is to assess, in a systemic way, the connection of visions and interests of different stakeholders. The

method was developed as a qualitative participatory method for gaining societal acceptance and provides facilitators (such as project managers, consultants, advisors) guidelines, approaches and milestones in arenas characterised by multi-stakeholders, where competing interests are at play. It consists of 6 steps:

Step 1. Project past and present information is collected about the project for later analysis

Step 2. Vision building different views are created by means of multiple interviews

Step 3. Identifying conflicting issues conflicting issues are identified by studying the interviews and ranked by the project manager

Step 4. Portfolio of options different ways to improve the project acceptance by reviewing a variety of solutions are identified

Step 5. Getting to shake hands open up the process to a larger number of stakeholders and discuss in a wider setting the different challenges and solutions identified

Step 6. Recommendations for action study the results and recommendations are formulated

This method was first developed for striving towards societal acceptance in big (infrastructural) energy projects. In a similar way as the paper introducing ESTEEM, our project starts from an assessment of technology within the context of already made policy decisions, while no real vision and strategy on the execution of this policy is available. We refer in this context to the European PSI directive [9] that puts forward the obligation of opening up data by default and which has to be integrated by administrations in regional policies. Open (Transport) Data is then envisioned to encourage innovation by market players and make public administrations more transparent to the broader public, yet implementation details into concrete actions are yet to be filled in. The deliberative forecasting exercise of the ESTEEM method is based on the confrontation with different stakeholders that may introduce "alternative routes", in order to reduce the trial and error process associated with innovation management, and thus reducing uncertainty and doubt.

An interdisciplinary team of three researchers with backgrounds in public administration and communication, business modeling and software engineering, studied how an Open Data policy can be implemented at the DTPW, between May 2015 and January 2016. A specific focus within multimodal route planning apps was requested.

4. Discussing datasets

Out of 27 interviews with data owners and directors working in the policy domain of mobility, we collected a list of datasets potentially useful for multimodal route planning. Definitions of *a dataset* however diverged depending on who was asked. Data maintainers often mentioned a dataset in the context of an *internal database*, used to fulfil an internal task, or used to store and share data with another team. When talking to directors, a dataset would be a *publicly communicated dataset*, e.g., a dataset for which metadata can be found publicly, a dataset that would be discussed in politics or a dataset the press would write stories about. In other cases, a dataset would exist *informally* as a web page, or as a small file on a civil servant's hard drive.

A data register of the mobility datasets that are part of the Open Data strategy can now be found at `http://opendata.mow.vlaanderen.be/`. The list consists of *publicly communicated datasets* as well as *informal data sources* published on websites. During the interviews, we were

able to gather specific challenges related to specific datasets useful for multimodal route planning, summarised in Table 1: for a dataset to be truly interoperable, all boxes need to be ticked.

Dataset	legal	technologic	syntax	semantic	querying
Traffic Events	open license	yes	XML	no URIs	file
Roads database	open license	yes	XML	no URIs	no
Validated statistics	open license	yes	CSV	no URIs	no
Information websites	no	yes	HTML	no URIs	linked documents
PT timetables	closed license	over FTP	ZIP	no URIs	dump
Road Signs	no license	yes	none	no URIs	no
Address database	open license	yes	XML	PoC	dump & service
Truck Parkings	open license	yes	XML	no URIs	file
Metadata catalogue	open license	yes	XML	yes	dump

Table 1 Selection of studied datasets with their interoperability levels as of October 2016

Traffic events on the Flemish highways
This dataset is maintained by the Flemish Traffic Center, has an open license and is publicly available. It decribes the traffic events, only on the highways, to which the core tasks of the traffic center is limited. The datasets can be downloaded in XML. For the semantics in this XML, two versions in two different specifications (OTAP and Datex2) are available, for which the semantics can be looked up manually. The elements described in the files are not given global identifiers however, making it impossible to refer to a similar object in a different dataset. The dataset is small and is published as a dynamic data dump. As the dataset is small enough to be contained in one file, it can be fetched over HTTP regularly, as well as the updates. The HTTP protocol works well for dynamic files, as caching headers can be configured in order not to overload the server when many requests happen in a short time. The file, except for the *semantic interoperability*, thus provides also as a good dataset for federated route planning queries.

Road database for regional roads
The road database for the regional roads is maintained by the ART. It is a geospatial dataset and already has to comply to the INSPIRE directive. Its geospatial layers are thus already available as web services on the geospatial access point of Flanders: `http://geopunt.be`. The roadmap in 2016 is to also add an open license and to also publish the data as linked files using the TN-ITS project's specification (`http://tn-its.eu/`).

Validated statistics of traffic congestion on the Flemish highways
Today, validated statistics of traffic congestions on the Flemish highways are published under the Flemish Open Data License by the Flemish Traffic Center. A website was developed, which allows someone that is interested to create charts of the data, as well as export the selected statistics as XLS or CSV. The *legal*, *technical* and *syntactic interoperability* are thus fully resolved. Yet when looking at the *semantic interoperability*, no global identifiers are used within the dataset. Furthermore, when looking at the *querying interoperability*, machines are even discouraged from using the files, as a test for whether you are a human (a captcha) is used to prevent machines from discovering and downloading the data automatically. When requesting a CSV file, the server generates the CSV file with historic data on the fly from the database.

Information Websites
Examples of such datasets are the real-time dataset of whether a bicycle elevator and tunnel is operating (`http://fietsersliften.wegenenverkeer.be/`), a real-time dataset of whether a bridge is open or not (`http://www.zelzatebrug.be/`) shows when a bridge north of the city of Ghent will open again when closed), or a dataset of quality labels of car parks next to highways (`http://kwaliteitsparkings.be`). The three examples mentioned can be accessed in HTML. Nevertheless, this as well is a valuable resource for end-user applications, as when the page would be openly licensed and when the data would be annotated with web addresses, the data can be extracted and replicated with standard tools and questions can be answered over these different data sources. These three examples are always only *technologically* and *syntactically* interoperable, as they use HTML to publish the data, yet there are no references to the meaning of the words and terms used. Furthermore, there is no open license on these websites, not explicitly allowing reuse of this data. Finally, as the data can easily be crawled by user-agents and thus replicated, we reason that in a limited way, the data would be able to be used in a federated query.

Public transit time tables maintained by De Lijn
Planned timetables, as well as access to a real-time webservice, can be requested through a one-on-one contract. This contract results in an overly complex legal interopability. First, a human interaction need to request access to the data, which can be denied. Furthermore, in the standard contract, it is not allowed for a third party to sublicense the data, which makes republishing the data, or a derived product, impossible. The planned timetables can be retrieved in the General Transit Feed Specification (GTFS) specification, which is an open specification, making the dataset *syntactically interoperable*. The identifiers used within this dataset for e.g., stops, trips or routes do not have a persistency strategy. Therefore, the *semantic interoperability* cannot be guaranteed. As a dump is provided, potential reusers have access to the entire data source reliably. The querying interoperability could be higher when the dataset would be split in smaller fragments.

Road Sign Database (RSD)
The database, in October 2016, is still only available through a restricted application. It is a publicly discussed dataset, as its creation was commissioned by a decree. On a regional level, the RSD is in reality two data stores: one database for regional road signs, managed by ART, and a database which collected the local road signs, managed by the department itself. Some municipalities would however also keep a copy of their own road signs on a local level, leading to many interoperability problems when trying to sync. Sharing this data with third parties however only happens over the publicly communicated RSD, which is only accessible through the application of the RSD itself.

Address database
A list of addresses is maintained as well by another agency, called Information Flanders. The database has, just like the RSD, to be updated by the local administrators. Thanks to the simplicity of the user-interface and the fact that it is mandatory to update the database while changing, removing or adding addresses, the database is well adopted by the local governments. It is licensed under an open license, and it is published on the Web in two ways: a data dump is updated regularly, and a couple of web services, which work on top of the latest dataset. Currently, Information Flanders is creating a Proof of Concept (PoC) to expose the database as Linked Data: every address will get a URI.

Truck parkings on the highways
This dataset needs to be shared with Europe, which in its turn makes this dataset publicly available at the European Union's data portal (`http://data.europa.eu/euodp/en/data/dataset/`

etpa) [8]. The dataset is available publicly, under an open license, as XML, using the Datex2 stylesheet. The file however does not contain persistent identifiers, thus it is impossible to guarantee the *semantic interoperability*. As with the traffic events, the file allows for querying by downloading the entire file.

Open Data portal's metadata
In order for datasets to be found by e.g., route planning user agents, they need to be discoverable. The metadata from all datasets in Flanders are available at `http://opendata.vlaanderen.be` in RDF/XML. The metadata is licensed under a Creative Commons Zero license, and for each dataset and its way to be downloaded (distribution), a URI is available. In order to describe the dataset, the URI vocabulary DCAT is used, which is a recommendation by the European Commission in order to describe data catalogues in an interoperable way. However, within INSPIRE, another metadata standard was specified for geospatial data sources. GeoDCAT-AP is at the time of writing being created to align INSPIRE and DCAT (`https://joinup.ec.europa.eu/node/139283`). It is thus far, the only dataset that complies in an early form to all the interoperability levels.

5. Challenges and workshops

We 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, both as prescribed by the ESTEEM method. In the first workshop, we welcomed a representative of each organisation within the DTPW that we had already met during a one on one interview. In the first half, we had an introductory program where we summarised the basics of an Open Data policy: the open definition, the implementation of the PSI directive in Flanders and the interoperability layer model. Furthermore we also gave a short summary of the results of the interviews with the market stakeholders. The key challenges were listed and discussed, initially identified by the heads of division of the DTPW. In order to identify these challenges, all interviews were first analysed in search of arguments both pro and con an open data policy. In the second half of this workshop we had three parallel break-out sessions in which we discussed unresolved questions that came out of the interviews. The arguments that returned most often were bundled and summarised into ten key challenges:

1. Should data publishing be centralised or decentralised within the department and what process should be followed?

2. Ensuring reusers interpret the data correctly

3. Acquiring the right means and knowledge on how to publish open data within our organisation

4. Knowing what reusers want

5. Influencing what reusers do with the data

6. Supporting evidence based policy-making

7. Creating responsibility

8. Raising government's efficiency

9. Ensuring sustainability once a dataset is published

10. Ensuring the technical availability of datasets

These challenges were discussed in smaller groups during the workshop in order to formulate solutions. By giving answers or providing "ways out" of these questions, the participants were challenged to think together and develop a solution that is carried by everyone in the organisation.

In the second workshop, we invited several market player reusing Flemish Open Data. As a keynote speaker, we invited CityMapper (`http://citymapper.com`), which outlined what data they need to create a world-wide multimodal route planner.

6. Recommendations for action

The three directives (PSI, INSPIRE and ITS) were often regarded as the reference documents to be implemented. The best-practices for PSI, as put forward by the "Interoperability solutions for public administrations, business and citizens" (ISA^2) programme, focus on Linked Data standards for semantic interoperability. However, the INSPIRE directive for geospatial data, brings forward a national access portal for geospatial data services, in which datasets are made available through services. There is a metadata effort, called GeoDCAT-AP, which brings the metadata from these two worlds together in one Linked Data specification. The ITS directive also puts forward their own specifications, such as NETEX at `http://netex-cen.eu/`, Datex2 at `http://www.datex2.eu/` and SIRI at `http://www.siri.org.uk/`. These specifications do not require persistent identifiers, and do not make use of URIs for the data model. We advised the department to first comply to the ISA^2 best practices, as getting persistent, autodocumented identifiers is the only option today to raise the *semantic interoperability* on web-scale. For datasets that already complied to the INSPIRE or ITS directive, the department would also make these available as data dumps (e.g., as with the roads database).

The Flemish government has *style guidelines* for their websites. We advised to implement extra guidelines for the addition of structured data, e.g. with RDFa [6]. Next, a conclusion from the first workshop was to invest in *guidelines* for the creation of databases. This should ensure each internal and externally communicated dataset is annotated with the right context.

In order to overcome the many organisational challenges, recommendations for action were formulated and accepted by the board of directors:

- Keeping a private data catalogue for all datasets that are created (open and non-open)

- All ICT policy documents need to have references to the Open Data principles outlined in the vision document

- The department of DTPW is responsible for following up these next steps, and will report to the board of directors.

- Opening up datasets will be part of the roadmap of each sub-organisation within DTPW

- On fixed moments, there will be meetings with the Agency Information Flanders to discuss the Open Data policy

Finally, also specific recommendations to data owners, as exemplified in Table 1, were given.

7. Conclusion

As the goal of an Open Data policy is to share data with anyone for any purpose, studying an Open Data policy means studying *how to maximise the potential reuse of datasets*. Reuse of datasets is going to happen only when the cost for adoption is lower than the benefits of adopting it, and this cost for adoption can be lowered when the overall interoperability of data sources increases. In this paper, we studied how to raise the interoperability of public mobility datasets. First, a framework was introduced which studies data source interoperability, as illustrated in Figure 1. Next, we interviewed 27 data maintainers and directors in the DTPW in Flanders. We listed datasets that are available and can be used for multimodal route planning advice, for which a small selection was summarised in Table 1.

Flanders took the first big leap in Open Data by implementing the PSI directive and having standard licenses for their open datasets. Next, also the technical interoperability is converging towards the Internet and using the HTTP protocol. Furthermore, the syntaxes chosen by some key datasets are open formats, while other datasets on the open data portal are still published as e.g., non-interoperable PDF files. Flanders is at this moment experimenting with the first steps towards Linked Data. They have the first iteration of a URI strategy and publish Linked Data of the metadata catalogue, a proof of concept on local decisions, and a proof of concept for the address database. Some datasets are even already usable for federated querying, as some dataset is small enough to be able to publish it in a small file. Some datasets also have query interfaces or data dumps, which make them usable for respectively querying locally or remotely, yet still the cost for adoption is high for federated querying in both cases. It is remarkable to see that the system of e.g. the validated traffic statistics at the Traffic Center invested in an export application that is only usable by humans, but still have an open license attached to it. Also in other countries, datasets are often hidden behind registration forms, making the dataset not discoverable by machines. We can conclude that, when a developer wants to create a new global intermodal route planner, the cost for adoption is still too high. For each dataset in Flanders, at least one interoperability problem needs to be resolved. We formulated recommendations for the DTPW to take into account in Section 6.

Using the ESTEEM method, ten organisational challenges were identified in Section 5. By listing these challenges, we can estimate the readiness of an organisation for Open Data. We invite surveys within other organisations following a similar method, to compare these challenges within different organisations. Furthermore, when the readiness for Open Data evolves, new challenges may arise and old ones may disappear, thus the evolution can be studied within a big organisation.

Today, we see the implementation of three European directives cause confusion on how to comply with different specifications. PSI and the INSPIRE directives are slowly resolving conflicts by enabling Linked Data within the geographic domain and aligning their vocabularies of terms. The ITS directive however reintroduces conflicts: proposed specifications such as TN-ITS, Datex2, NETEX, SIRI, do not require persistent identifiers. The risk for existing open transport datasets is that transport data publishers will have to invest not in enhancing the semantic interoperability and ability to federate queries, but in complying with a new syntactic set of rules.

Finally, federated route planning is currently still far from reality. We see a limited amount of market players picking up data with a low interoperability level. However, we do not yet see the wide adoption of open datasets in multimodal route planners that the Open Data movement promised. If we want route planning to become as a commodity addition to existing services, we need to lower the cost for adoption of such data sources, and thus raise the interoperability of our open transport datasets.

8. References

8.1. Websites

[1] Schmachtenberg M., Bizer C., Jentzsch A., Cyganiak R.: 'Linking Open Data cloud diagram 2014', `http://lod-cloud.net/`, accessed 10th of October 2016

[2] 'Transport Group - The European Data Portal', `http://www.europeandataportal.eu/data/en/dataset?groups=transport`, accessed 16th of August 2016

[3] 'RDF 1.1 XML Syntax', `https://www.w3.org/TR/rdf-syntax-grammar/`, accessed 19th of September 2016

[4] 'JSON-LD 1.0: A JSON-based Serialization for Linked Data', `https://www.w3.org/TR/json-ld/`, accessed 19th of September 2016

[5] 'Model for Tabular Data and Metadata on the Web', `https://www.w3.org/TR/2015/REC-tabular-data-model-20151217/`, accessed 19th of September 2016

[6] 'RDFa Core 1.1 - Third Edition: Syntax and processing rules for embedding RDF through attributes', `https://www.w3.org/TR/rdfa-syntax/`, accessed 19th of September 2016

[7] 'RDF Mapping Language (RML)', `http://rml.io/spec.html`, accessed 19th of September 2016

[8] 'Article 5 of Commission delegated Regulation (EU) No 885/2013 in the framework of the ITS Directive 2010/40/EU', `http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32013R0885`, accessed 2d of October 2016

[9] 'Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information', `http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A02003L0098-20130717`, accessed 2d of October 2016

[10] 'Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE)', `http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32007L0002`, accessed 2d of October 2016

8.2. Journal articles

[11] Risto Ö., Evgeni M., Timo Olavi K.: 'Interoperability of eCall and ERA-GLONASS in-vehicle emergency call systems', IET Intelligent Transport Systems, 2015, 9, (6), p. 582–590

[12] Colpaert P., Van Compernolle, M., De Vocht L, Dimou A., Vander Sande M, Verborgh R, Mechant P, Mannens E.: 'Quantifying the Interoperability of Open Government Datasets', IEEE Computer, 2014, p.50–56

[13] Van Cauter, L.: Government-to-government information system failure in Flanders: an in-depth study, 2016

[14] Melnik, S., Decker, S.: 'A layered approach to information modeling and interoperability on the Web', In Proceedings of the ECDL'00 Workshop on the Semantic Web, 2000

[15] Verstichel, S., Ongenae, F., Loeve L., Vermeulen, F., Dings, P., Dhoedt, B., Dhaene, T., De Turck, F.: 'Efficient data integration in the railway domain through an ontology-based methodology', Transportation Research Part C: Emerging Technologies, 2011, 19, p.617–643

[16] Sheth, A. P.: 'Changing focus on interoperability in information systems: from system, syntax, structure to semantics', Interoperating geographic information systems, 1999, p.5–29

[17] Rezaei, R.: 'Interoperability evaluation models: A systematic review', Computers in Industry, 2014, p.1–23

8.3.   Conference Paper

[18] Verborgh, R., Hartig, O., De Meester, B., Haesendonck, G., De Vocht, L., Vander Sande, M., Cyganiak, R., Colpaert P., Mannens, E.: 'Querying datasets on the web with high availability', International Semantic Web Conference, 2014

[19] Colpaert, P., Llaves, A., Verborgh, R., Corcho, O., Mannens, E., Van de Walle, R.: 'Intermodal public transit routing using Linked Connections', Proceedings of the 14th International Semantic Web Conference: Posters and Demos, 2015

[20] Walravens N., Van Compernolle M., Colpaert P., Ballon P., Mechant P., Mannens E.: '"Open Government Data"-based Business Models: A Market Consultation on the Relationship with Government in the Case of Mobility and Route-Planning Applications'. Proceedings of the 13th International Conference on e-Business, Lisbon, Portugal, 26-28 July 2016.

[21] Buyle R., Colpaert P., Mechant P., Volders V., Verborgh, R., Mannens, E.: "Local Council Decisions as Linked Data: a proof of concept", Proceedings of the 15th International Semantic Web Conference: Posters and Demos (2016).

[22] Buil-Aranda, C., Hogan, A., Umbrich, J., Vandenbussche, P.: 'SPARQL Web-Querying Infrastructure: Ready for Action?', Proceedings of the 12th International Semantic Web Conference 2013, 2013, p.277–293 year=2013

8.4.   Book, book chapter and manual

[23] Raven, R. P., Jolivet, E., Mourik, R. M., & Feenstra, Y. C. (2009). ESTEEM: Managing societal acceptance in new energy projects: A toolbox method for project managers. Technological Forecasting and Social Change, 76(7), 963–977.

[24] Fielding, R. T. (2000). Architectural styles and the design of network-based software architectures (Doctoral dissertation, University of California, Irvine).