Creating unified data APIs for Rijkswaterstaat

Team

- Tom Ootes: front end developer specialising in data visualisation
- Jeremie Coullon: PhD student in mathematical modelling of traffic flow on motorways

Motivation

We both joined this challenge for a chance to solve interesting transportation problems by working with Rijkswaterstaats' datasets, analysing it, and making insightful visualisations. When we made the team our first step was to understand what were the problems faced by the experts working within Rijkswaterstaat. Although we learned that there are many interesting challenges within each discipline, but we especially noticed that the data was scattered in different locations and organised in an inconsistent way. There is no consistent way to access data, so one has to become aquainted with the particular quirks of each API before even being able to start working with it!

The difficulty of accessing and using data makes it difficult to find solutions to the transportation problems Rijkswaterstaat.

Obstacles

- Different provinces and disciplines publish their data in different API's
- Data has to be downloaded in an ad-hoc way (ex: links to servers to download zipped xml files)
- Broken links to download data
- Inconsistent API means that there's a learning curve before even loading up a new dataset.
- Some of the sites managing the data have a confusing UX (ex: circular or broken links)

Though acknowledging that this would be a huge undertaking, we think that building a consistent API for Rijkswaterstaats' datasets would be an important step in solving transportation problems. This would allow easy access for all kinds of datasets managed by Rijkswaterstaat for internal as well as for external use.

We believe Transport for London (TFL)'s API[1] is a nice model of the kind of solution we have in mind. TFL's API allows standardised access to different kinds of data about London's transport: status of tube lines/stops, location/status of bike stops, accidents on roads etc...

Idea

Make data (both realtime and static) easily accesible for everyone both internal and external to Rijkswaterstaat. This would accelerate he development of creative solutions to Rijkswaterstaat's transportation problems using data.

Users and their requirements:

- Internal users: need easy access to relevant databases for analysis. The interdisciplinary nature of these user's work means that they need to be able to access differenent types of data easily without learning a new API
- External users: need easy access to 3rd party developers and researchers. Unlike the interal users (who can contact their collegues), they may not necessarily have specialised help in accessing and using the data.
- Documentation/UX: Clear documentation and UX allows new users to find and query data with the least amount of friction as possible

Solution: a consistent and well-documented API

The implementation of this idea can be realised by creating a API with static and realtime information on assets of Rijkswaterstaat.

The development of this simple interface for data would unlock the development of data-driven transportation solutions.

Concrete plan: API with realtime and static/historical data

Create realtime and static data for relevant Rijkswaterstaat insfrastructure such as:

- Locks: various measurements from sensors (ex: energy use, local wind speed and direction etc. . .)
- Roads (density and speed over road sections)
- Accidents (lcoation of accidents on the road)
- state of the road: road, bridge, and tunnel closures, or faults in the road/bridge/tunnel

Case Studies

To illustrate the empowering effect of a simple API for developers and data analysts, we first use TFL's API to create a simple interactive map. We then give examples of how a similar API for Rijkswaterstaat could be used to help solve transportation problems.

TFL - bike stations

London has a public bicycle hire scheme which consistents of bikes in stations distributed evenly throughout London. One can buy a subscription to use bikes at any of the stations. One can use TFL's API to retrieve location of these bike stations.

We use this API to develop a very simple web application at http://jeremiecoullon.github.io/TFL-maps/: one can click on a point on the map of London and get all bike stations within a selected radius.

To acheive this we used open source tools:

- Angular: front end javascript framework
- Openlayers3: javascript library for creating interactive maps

TFL's API:

- documentation: each dataset has a description of it's content and how to query it
- common format and structure: returns JSON or XML

As TFL adds or removes bike stations and this API gets updated, the interactive map is also updated. Furthermore as all the datasets are presented in the same format, there is less of a learning curve in querying another database.

Rijkswaterstaat

We present some example to illustrate how a clear API could be used for Rijkswaterstaat.

1. Visulalise on a map the position of all bridges on a map along with their status (closed or open) similarly to the TFL example. This allows a very visual representation in the road layout for decision-makers
2. Exploring the causes of faults on bridges: one could overlay on a map data such as the cumulative vehicles counts on all bridges along with the status of the faults on the bridges. This kind of visualisation could help as an exploration tool to understand what the causes of the deterioration of roads/bridges and how one might deal with it.
3. Asset management: One can overlay existing data (such as vehicles counts on roads, congestion, VMS, deterioration of roads etc..) to make decisions about resource allocation. As the data gets updated, the visualisation could be used continuously. Furthermore, as this kind of analysis needs input from many different datasets, a consistent API would be very helpful in creating this.

Conclusion

There is a need for a consistent way to query real-time and static datasets managed by Rijkswaterstaat, both for internal as well as external use. The example of TFL's API shows that this is both possible and effective in boosting the use of data for research and applications. Investigating in more detail how TFL's API was built and considering carefully the unique properties of Rijskwaterstaats' datasets will be essential in understanding how this can be done in practice.

---

[1] https://api.tfl.gov.uk/