The need for unified data APIs for Rijkswaterstaat

Campus Party 2016 - Make'm Smarter

web application here:

Team

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

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 formed 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, 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. We highlight this fact and illustrate the benefits that can be gained from a unified data API.

Current obstacles to accessing data

- 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 APIs means that there is a learning curve to overcome before even loading up a new dataset.
- Some of the sites managing the data have a confusing user experience (UX) (ex: circular, redundant, 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 and stops, location and status of bicycle stations, the location of accidents on the roads etc...

Solution: a consistent and well-documented API

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

Users and their requirements:

- Internal users: These 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: Similarly, 3rd party developers and researchers need easy access to open databases. However, unlike the interal users (who can contact their collegues), they may not necessarily have specialised help in accessing and using the data.
- Documentation/UX: Both types of users need clear documentation. A good UX allows new users to find and query data with the least amount of friction as possible.

The data and some challenges

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 help boost the development of data-driven transportation solutions.

Some datasets for Rijkswaterstaat's insfrastructure that would need to be unfiied include:

- 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

Some challenges in implementing this include:

- A legacy database management system to migrate: there are many various databases dispersed in different locations.
- Differenet workflows: current users of these databases would need to rebuild their data workflows to work with a new API
- Highly heterogeneous data: each dataset has it's own specific structure so the API must be flexible to accomadate all of these (for example, the larger databases must be optimised to enable fast queries)

The following case studies illustrate that doing this is possible as well as beneficial.

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 - bicycle stations

London has a public bicycle hire scheme which consists of bicycles in stations distributed evenly throughout London. One can buy a subscription or pay on-the-go to use a bicycle at any of the stations. TFL's API makes it easy to query the locations of these bicycle stations.

We use this API to develop a very simple web application here: one can click on a point on the map of London and get all bicycle stations within a selected radius.

To acheive this we used open source tools:

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

Some features of TFL's API include::

- Documentation: each dataset has a description of it's content and how to query it
- Common data format and structure: returns JSON or XML
- Designed for live (real-time) and high-volume use

On the TFL website, one can input the parameters for a data query, and the URL that fetches the data is then generated. The parameters in the URL are intuitive so recreating this user input in a web application (as in the current example) or any other use case is straightforward.

If TFL adds or removes bicycles 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 - possible uses

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

1. Visulalise the position of all bridges on a map along with their status (closed or open) in a similar way to the TFL example. This allows a very clear visual representation in the road layout for decision-makers. An illustration of this example using artificial data is shown in the web application linked on page 1.
2. Exploring the causes of faults on bridges: one could overlay on a map data such as the cumulative vehicles counts on all bridges as well as 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/