

Beleidsimpacterend dashboard op basis van mobiliteitsdata

Cédric Reyniers, Simon Vermeersch

Promotoren: prof. dr. ir. Sofie Verbrugge, prof. dr. ir. Filip De Turck

Begeleiders: dr. ir. Marlies Van der Wee, dr. Stijn Verstichel, dr. Femke Ongenae

Masterproef ingediend tot het behalen van de academische graad van
Master of Science in de industriële wetenschappen: informatica

Vakgroep Informatietechnologie
Voorzitter: prof. dr. ir. Bart Dhoedt
Faculteit Ingenieurswetenschappen en Architectuur
Academiejaar 2016-2017



Beleidsimpacterend dashboard op basis van mobiliteitsdata

Cédric Reyniers, Simon Vermeersch

Promotoren: prof. dr. ir. Sofie Verbrugge, prof. dr. ir. Filip De Turck

Begeleiders: dr. ir. Marlies Van der Wee, dr. Stijn Verstichel, dr. Femke Ongenae

Masterproef ingediend tot het behalen van de academische graad van
Master of Science in de industriële wetenschappen: informatica

Vakgroep Informatietechnologie

Voorzitter: prof. dr. ir. Bart Dhoedt

Faculteit Ingenieurswetenschappen en Architectuur

Academiejaar 2016-2017



UNIVERSITEIT
GENT

Woord vooraf

Dankzij dit eindwerk hebben we hopelijk een stap in de goede richting gezet voor het openbaar vervoer. Hier bijkomend hebben we tevens beter leren samen werken.

We zouden in de eerste plaats onze promotoren prof. dr. ir. Sofie Verbrugge en prof. dr. ir. Filip De Turck willen bedanken. Zonder hen zou de verwezenlijking van deze masterproef niet mogelijk zijn geweest. Een speciaal dankwoord ook aan onze begeleiders dr. ir. Marlies Van der Wee en dr. Stijn Verstichel en dr. Femke Ongenae. Hun wekelijkse opvolging zorgde ervoor dat we nooit te lang bleven stilstaan bij problemen. Verder zouden we ook IDLab willen bedanken voor het gebruik van hun faciliteiten gedurende deze maanden.

Graag zouden we ook onze ouders en vrienden willen bedanken voor hun steun, op de moeilijke momenten konden we steeds op ze rekenen. Hierbij nog een extra dankjewel aan onze medestudenten op IDLab met wie we regelmatig problemen besproken hebben: Remi Helaers, Jerico Moeyersons en Maxim Ronsse.

Cedric Reyniers en Simon Vermeersch, 2017

Abstract

Mobiliteit is één van de belangrijkste pijlers van de hedendaagse maatschappij, maar komt hand in hand met een aantal uitdagingen. Het aantal personen dat gebruik maakt van het wegennet neemt jaarlijks toe en de infrastructuur is op vele plaatsen in de wereld niet opgewassen tegen deze grote drukte. Eén van de manieren om tegemoet te komen aan deze uitdagingen is door in te zetten op verbeteringen van het openbaar vervoer. Deze verbeteringen worden best gedreven door de problemen die voorkomen te identificeren en daarop te handelen. Daarom wordt in deze thesis een dashboard gecreëerd voor het management van openbare vervoersmaatschappijen, meer specifiek voor spoorbedrijven. Elk bedrijf heeft een aantal belangrijke indicatoren (Key Performance Indicators - KPIs) waarop gefocust wordt. Dit platform zal gemeenschappelijk bruikbaar zijn door verschillende bedrijven met behulp van een gemeenschappelijk vocabularium; hierdoor kunnen ondernemingen kwalitatief met elkaar vergeleken worden.

Elke KPI heeft een bepaalde waarde die berekend kan worden, bijvoorbeeld de stiptheid in een bepaalde periode. Hiernaast heeft elke KPI ook een bepaalde invloed op het geheel; de gemiddelde vertraging kan bijvoorbeeld voor tien procent meetellen in de totaalscore van het bedrijf.

In het dashboard, de applicatie bovenop het platform (Proof-of-Concept), worden dergelijke indicatoren met elkaar in verband gebracht via een configureerbare boomstructuur om zo een globale performantiescore van de onderneming te bekomen. Met behulp van de boomstructuur kunnen ook de belangrijkste knelpunten ontdekt worden. Er werd dus een dashboard gecreëerd waarbij de gebruiker nieuwe indicatoren kan aanmaken gebaseerd op de verzamelde data uit verschillende databronnen van België en het Verenigd Koninkrijk. Na het creëren van deze indicatoren kan de gebruiker de resultaten ervan analyseren en ook de vergelijking maken tussen verschillende bedrijven in België en het Verenigd Koninkrijk.

De resultaten zijn positief in die zin dat er duidelijke knelpunten zichtbaar werden en het was mogelijk – rekening houdend met de beperktheid van de data – om een totaalscore voor een onderneming op te stellen. Deze score evenals de indicatoren konden individueel vergeleken worden tussen verschillende ondernemingen om ze de verschilpunten duidelijk te maken. Op deze manier is het mogelijk om bedrijven kwalitatief met elkaar te vergelijken.

Policy-impacting KPI dashboard based on mobility data

Cédric Reyniers and Simon Vermeersch

Supervisors: prof. dr. ir. Filip De Turck, prof. dr. ir. Sofie Verbrugge
Counsellors: dr. ir. Marlies Van der Wee, dr. Stijn Verstichel, dr. Femke Ongenae

Abstract – Mobility is one of the most important aspects in modern society, but it presents quite a few challenges to this society. There are a lot of people that use their car to go to work. This is not healthy because the roads are getting congested. One way to solve this is by improving public transport. It would be better if commuters start using public transport instead of driving their car. The aim of this master thesis was to find a way to help public transport companies in improving their services. To help these companies a dashboard has been created. On this dashboard, the management can select the most important Key Performance Indicators (KPI) they want to see. The way these KPIs are calculated is shared between companies but the data used in the calculations is dependent on the company. This is done by using a common vocabulary. The purpose of it all is to find obstacles and resolve them. In the end, it became clear that it's possible to compare different companies by making use of a common vocabulary. However, it would be better if the common vocabulary was drawn up by experts.

Keywords – KPI, public transport, semantics, reasoning, LimeDS, OSGi

I. INTRODUCTION

Today, a lot of commuters use their car instead of public transport. This should change for various reasons, i.a. traffic jams and the environment. Most people like to be stuck in traffic jams rather than using public transport, but why? The reason why people do not use public transport might be resolved by improving the services of public transport companies.

In this study, the operation of different railway companies is analyzed. In the first place to detect problems for a specific railway company and secondly to compare different railway companies.

This introduction is followed by four chapters. First, the research that has been done about KPIs, railway companies and ontologies is listed. The implementation comes after research and describes the architecture, the proposed ontology and the possibilities of the application. This chapter is followed by the results of railway companies in Belgium and the UK. At the end, there is a conclusion about the implementation and the results.

II. RESEARCH

A. Key Performance Indicators

Key Performance Indicators (KPI) are values that represent a measurement of the performance of a company on a business objective. These indicators can help companies to improve their service by pointing out obstacles. KPIs can also be used to compare companies in a qualitative manner.

KPIs can be either objective or subjective [1]. The difference between the two can be shown in an example: the amount of delayed services is an objective KPI whereas the customer satisfaction is a subjective KPI. In this, mainly objective KPIs are dealt with.

There are a lot of KPIs for public transport companies and it goes without saying that not all those KPIs will be dealt with in the implementation of this dashboard.

B. Stakeholders

An important question to ask before creating this platform is: who needs this platform? First and foremost, the people using the services of railway companies will benefit from this platform. The platform will help to improve the services of the companies which will in turn benefit the passengers. This leads to the second important stakeholder: the management of a railway company. They will monitor the KPIs and use the dashboard. In that way, they will detect problems and they will know which things should be improved.

C. Belgium vs UK

Railway companies in Belgium and in the UK work in a different way. In Belgium, the railway is nationalised whereas in the UK it is privatised. This may lead to different results when comparing the KPIs of both countries.

In both countries, there are for example different ways to determine whether a service is delayed or not and who's responsible for it. Another interesting aspect is the ticket price; where does the money from the train passengers go to and is it justified?

D. Ontology

“An ontology is an explicit specification of a shared conceptualization”, Gruber [2]. This definition is one of the most widely used ones for ontologies, though it does not say much. An ontology conceptual model of domain knowledge, which defines a common vocabulary for researchers who need to share this knowledge, which should be defined by domain experts. It is most commonly used as metadata schemes which due to their formal explicit specification on concepts and relations in the domain knowledge are machine-readable.

III. IMPLEMENTATION

A. Technical description of the platform

The architecture can be divided into four big parts, as can be seen in Figure 1.

Firstly, the main component is the *Datahandler* which is based on the LimeDS¹ framework. This component, as the name suggests, handles the end-to-end dataflow and processing of this data. The most commonly used data sources are Darwin [3] and Twitter. Darwin is used to find the delay time and delay reason of the services in the UK. Twitter is used to get the delay reasons in Belgium which are mentioned in the tweets of NMBS.

Secondly, the platform uses a RDF-triple store database where all the instances of the used ontology scheme are being saved. This component is implemented as a micro-service and is connected to the *Datahandler* through a REST-service. A micro-service is a light-weight, fine-grained and loosely coupled service used in service-oriented architecture.

Thirdly, there is a second micro-service which is implemented the same way as the other micro-service. This micro-service is used for rule-based semantic reasoning on a subset of the database.

And finally, the client-side dashboard, the application on top of the platform. This dashboard uses AJAX calls to communicate with the *KPIEditor*. The *KPIEditor* is a segment in the *Datahandler* which is used for calculating and creating of KPIs.

By using a service-oriented architecture style the components can easily be replaced. The heart of this platform is found in the *Datahandler*, the rest is built around it and can be changed.

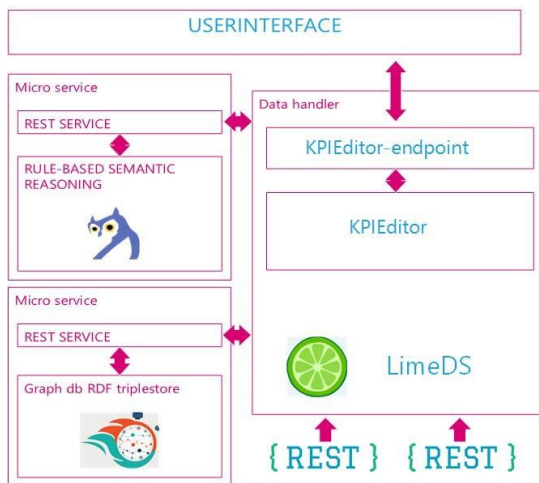


Figure 1: Dataflow between the components of this platform.

B. Mobility Data – Ontology

The ontology we created is based on KPIOnto² and Transit schema³. Unfortunately, this was deemed insufficient, so we added a few new concepts. This was necessary because delay events were not specified in the other ontologies. Also, a new concept was created for Twitter data, because Twitter is a frequently used data source in this Proof-of-Concept.

The way railway companies gather information is different for each company which leads to heterogeneous data between companies [4]. This creates a problem because it makes these datasets difficult to compare. Ontologies can help in resolving these issues. If, for example, one would want to compare the reason for delay in services of two different companies, the data should first be mapped to an ontology to be able to compare them. Delay events, defined in the ontology, were divided into categories like Defective, Interference, Third-party etc. By doing this and by mapping the delay reasons retrieved from data sources of Belgian and British train companies to these categories, comparing these companies became possible. These delay events categories can be used in KPIs.

C. Application

1) Create KPIs

To be able to create KPIs, one should have the knowledge to create queries. This means managers cannot use this application on their own. Indicators are divided into two main categories as can be seen on Figure 2: Measurements and High-Level-KPIs.

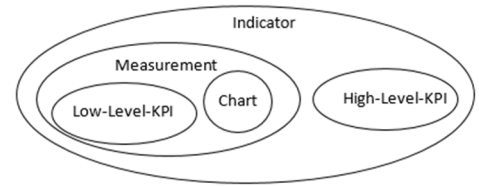


Figure 2 : Indicator set.

Measurements are created by defining a SPARQL and a SPARQL-DL [5] query. These are ontology schema query languages which respectively query from the database and the semantic reasoning component. When there is only one variable selected in the SPARQL-DL query then the Measurement is categorized as a Low-Level-KPI and an aggregation function like count or sum must be defined as well. If there is more than one variable selected, the KPI is categorized as a Chart which means they will be shown as a chart on the dashboard. To summarize, a Low-Level-KPI represents a one-dimensional result and a chart represents a multi-dimensional result.

To make sure KPIs can be used by different agencies and countries, there are two additional variables which can be used in the SPARQL queries: agency and country. When these are used the result of the query will depend on the user of the dashboard. When two different users sign in into the dashboard, the results of those KPIs will be different. In the example shown in Figure 3, the user is NMBS from Belgium and so this means the variable ‘&{COUNTRY}’ will be replaced by ‘:Belgium’ to match the instance of the country.

¹ <http://limesds.be/>

² <http://kdmg.dii.univpm.it/kpionto/specification/>

³ <http://vocab.org/transit/>

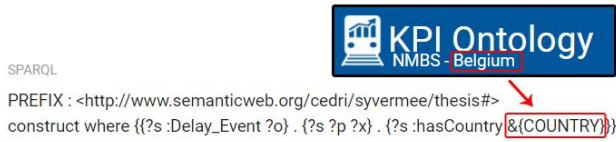


Figure 3 : The use of variables in SPARQL-queries.

High-Level-KPIs are built on Low-Level-KPIs and other High-Level-KPIs. The user of the application can create a formula that can contain these KPIs and numerical variables.

By using a combination of High-Level-KPIs and Low-Level KPIs a KPI tree can be built. An example of that can be seen on Figure 4.

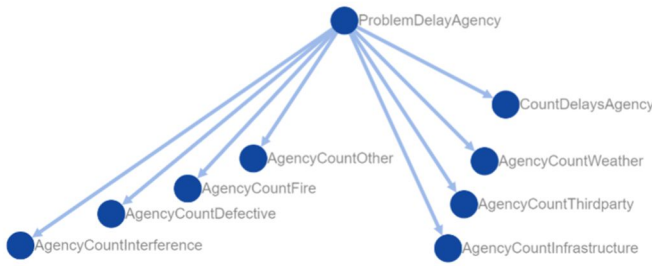


Figure 4 : Overview of a High-Level-KPI (*ProblemDelayAgency*) and the Low-Level-KPIs it is built on, shown in a KPI tree.

2) Dashboard

The dashboard can be used by the management; the user can pick the KPIs he wants to see. The results of these KPIs will be personalized for the user as explained earlier.

In this dashboard, the user can decide how the KPIs and charts will be shown. For a Low-Level-KPI, a maximum score for this KPI will have to be provided and the user will have to answer whether a low score is positive or negative. For example, when the user wants to monitor the amount of delayed services, a low score will be positive as can be seen in Figure 5.

For charts, there are some more questions that will have to be answered by the user. As mentioned before, when creating a chart, the user has to define which variables are selected. In the dashboard, the user has to decide which of these variables will be used to group by. For example, when the user has created a query with variables 'day' and 'services', he will probably want to group by 'day'. The second question for the user is which group by function he wants to use, to name a few: count, max, sum, etc. The last question for the user is whether the chart has to be visualised either as a bar chart or as a pie chart as can be seen in Figure 6.

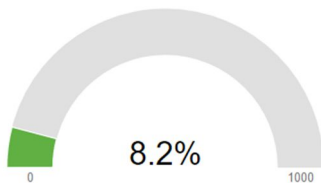


Figure 5 : Example of a High or Low-Level-KPI, shown in a gauge chart.

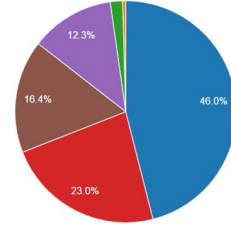


Figure 6 : Example of a Chart, shown in a pie chart. By hovering over the chart on the dashboard, the categories will become clear.

IV. RESULTS

After retrieving data from both Belgian and British train companies the big question is: is it possible to learn something out of it?

In Belgium, the results from this dashboard were compared with the monthly reports of previous months. As mentioned in those reports, a third of the delays are caused by third-party. But in this dashboard, it is possible to move a level deeper and see what kind of issues are predominant in third-party delay events. By doing that, trespassing has been found as the most common delay event in third-party. Infrabel, the rail infrastructure manager in Belgium, are already campaigning against trespassing and the risks involving running on the rail tracks [6]. This shows the dashboard can point out issues that need attention.

When looking at the United Kingdom, a clear difference between the working of the different rail companies was found. In this study, four companies were considered: they all had differences in the reasons for delay. The punctuality of the services turned out to be better than expected.

In comparison, Belgium and the United Kingdom do not differ all too much in the reason for delays. The biggest difference is in third-party incidents. In Belgium, around 30% of all delays are caused by these types of incidents while in the UK it is only around 15%. This shows that Belgium should work harder in preventing things like trespassing and make sure these kinds of incidents do not disrupt the service too much.

V. CONCLUSION

It has been proven that using ontologies can help in detecting issues as has been mentioned in the results.

While our goal has been reached, it must be emphasized that mapping of data and the terms used in the ontology should be done by domain experts. This should improve the results of the KPIs.

Still, the results show that the used strategy works.

REFERENCES

- [1] Z. Li and D. A. Hensher, "Crowding in public transport: A review of objective and subjective measures," 2013.
- [2] T. Gruber, "Ontology definition," 2007. [Online]. Available: <http://web.dfc.unibo.it/buzzetti/IUcorso2007-08/mdidattici/ontology-definition-2007.htm>. [Accessed: 20-Feb-2017].
- [3] National Rail Enquiries, "Darwin Data Feeds," 2017. [Online]. Available: <http://www.nationalrail.co.uk/100296.aspx>
- [4] Verstichel, S., Ongenaes, F., Loeve, L., Vermeulen, F., Dings, P., Dhoedt, B., ... & De Turck, F. (2011). Efficient data integration in the railway domain through an ontology-based methodology. *Transportation Research Part C: Emerging Technologies*, 19(4), 617-643.
- [5] A. Caf , "SPARQL-DL - Theory & Practice," 2013. [Online]. Available: <https://www.slideshare.net/adrielcafe/sparqldl-theory-practice>. [Accessed: 20-Mar-2017].
- [6] Infrabel, "Infrabel voert strijd tegen spoorlopers," 2017. [Online]. Available: <https://www.infrabel.be/nl/pers/infrabel-voert-strijd-tegen-spoorlopers>. [Accessed: 17-May-2017].

Inhoudsopgave

Woord vooraf	i
Abstract	ii
Extended Abstract.....	iii
Inhoudsopgave.....	vi
Lijst van afkortingen.....	ix
Hoofdstuk 1: Inleiding	1
1.1. Probleemstelling.....	1
1.2. Doelstelling	2
1.3. Tijdlijn.....	2
Hoofdstuk 2: Analyse van de functionaliteit	4
2.1. Use Case	4
2.1.1. Use case-diagram.....	4
2.1.2. Mogelijkheden voor gebruikers	5
2.2. Technische omschrijving	7
Hoofdstuk 3: Literatuurstudie	8
3.1. Key Performance Indicators in het openbaar vervoer	8
3.1.1. Wat is een KPI?.....	8
3.1.2. Case: Openbaar vervoer	9
3.1.3. Overzicht van mogelijke KPI's	11
3.1.4. Conclusie	31
3.2. KPI Dashboard	32
3.2.1. Waarom gebruik maken van een dashboard?	32
3.2.2. Voorbeelden van dashboards.....	32
3.3. Case Studie: toepassing op het treinverkeer.....	34

3.3.1.	Stakeholders bij treinondernemingen.....	34
3.3.2.	België	38
3.3.3.	Verenigd Koninkrijk	42
3.4.	Ontologieën.....	44
3.4.1.	Definitie	44
3.4.2.	Waarom worden ontologieën gebruikt?.....	45
3.4.3.	The Semantic Web	45
3.4.4.	Reasoning	48
3.4.5.	Semantic Web Rule Language	49
3.5.	Conclusie.....	50
Hoofdstuk 4: Implementatie.....		51
4.1.	Frameworks en technologieën.....	51
4.2.	Databronnen.....	54
4.3.	Package diagram.....	57
4.3.1.	Microservices	58
4.3.2.	LimeDS.....	59
4.4.	Ontologieën.....	60
4.4.1.	Key Performance Indicators	60
4.4.2.	Railway	62
4.4.3.	Twitter Delay Events	63
4.5.	Database.....	65
4.6.	Semantic Reasoning	66
4.7.	KPIEditor	68
4.7.1.	Inloggen.....	68
4.7.2.	Dashboard	70
4.7.3.	Creator	73

4.7.4. Analyzer	75
4.8. Conclusie	75
Hoofdstuk 5: Resultaten	76
5.1. België	76
5.1.1. Inleiding in verband met de opgehaalde data	76
5.1.2. Analyse van de resultaten	77
5.1.3. Conclusie	85
5.2. Verenigd Koninkrijk	86
5.2.1. Inleiding in verband met de opgehaalde data	86
5.2.2. Analyse van de resultaten	86
5.2.3. Conclusie	93
5.3. Vergelijking tussen België en het Verenigd Koninkrijk	94
5.3.1. Analyse van de resultaten	94
5.3.2. Conclusie	95
Hoofdstuk 6: Testing	96
6.1. Performance testing	96
6.1.1. Load/Stress testing	96
6.2. Conclusie	101
Hoofdstuk 7: The past and the future	103
Besluit	105
Referenties	106

Lijst van afkortingen

API	Application Programming Interface
CRUD	Create – Read – Update - Delete
CSV	Comma-Seperated Values
DB	Databank
DL	Description Logics
EHBO	Eerste Hulp Bij Ongelukken
HSP	Historic Service Performance (API van Darwin)
HTTP	HyperText Transfer Protocol
JSON(-LD)	JavaScript Object Notation (for Linked Data)
KPI	Key Performance Indicator
LimeDS	Lightweight modular environment for Data-oriented Services
NMBS	Nationale Maatschappij der Belgische Spoorwegen
OSGi	Open Services Gateway initiative
OWL	Web Ontology Language
OWL 2 EL	Web Ontology Language 2 Expressive Language
OWL 2 QL	Web Ontology Language 2 Query Language
OWL 2 RL	Web Ontology Language 2 Reasoning Language
OWL DL	Web Ontology Language Description Logics
RDB	Relationele DataBank
RDF	Resource Description Framework
REST	Representational State Transfer
RML	RDF Mapping Language
SOA	Service-Oriented Architecture
SPARQL(-DL)	Simple Protocol And RDF Query Language (Description Language)
SQL	Structured Query Language
SWRL	Semantic Web Rule Language
UI	User Interface
URI	Uniform Resource Identifier
XML	eXtensible Markup Language

Hoofdstuk 1: Inleiding

1.1. Probleemstelling

In moderne tijden speelt mobiliteit een grote rol in de maatschappij, het bevolkingsaantal op aarde blijft stijgen en het merendeel is dagelijks op stap. De meeste mensen maken gebruik van de auto voor hun woon-werkverkeer [1]. Een groot probleem is echter dat de infrastructuur niet overal voorzien is op dit grote aantal reizigers [2]. Een mogelijke oplossing voor dit probleem is dat men gebruik maakt van het openbaar vervoer. Dit heeft als voordeel dat er minder auto's op de baan zullen zijn, bijgevolg zal dit dus leiden tot minder files. Een ander voordeel is dat het ten goede komt van het milieu, een prangend probleem in de 21^e eeuw. Een belangrijk punt zijn de stakeholders, wie ondervindt er allemaal invloed van de werking van dit soort bedrijven? De meest voor de hand liggende stakeholder is de potentiële gebruiker van het openbaar vervoer zelf.

Eén van de grote vragen is bijgevolg: waarom maken sommige mensen geen gebruik van het openbaar vervoer? Het antwoord op hierop zou door gebruik te maken van een KPI-dashboard ontdekt kunnen worden. Indien een openbaarvervoerbedrijf zijn pijnpunten ontdekt, kan het hier in de toekomst rekening mee houden en verbeteringen aanbrengen. Wanneer de pijnpunten opgelost zijn zullen – hopelijk – meer mensen gebruik maken van de diensten van het bedrijf.

Key Performance Indicators (KPI's) zijn op dit vlak cruciaal voor een onderneming, ze helpen namelijk om de zwakke punten in een onderneming bloot te leggen. Een KPI is een meetbare waarde die weergeeft hoe effectief een onderneming een belangrijk *business objective* bereikt.

Ook spoorwegbedrijven moeten hun performantie evalueren en hebben de benodigde systemen hiervoor. Gezien ieder bedrijf zijn eigen systeem heeft om KPI's te evalueren en beheren is het een lastige opgave om deze met elkaar te vergelijken. Samenwerking is een belangrijke factor om een competitief voordeel voor beide partijen te behalen. Collaboratie steunt op gegevens uitwisselen, waar de betekenis van de informatie goed op voorhand gedefinieerd is. Hier komen semantische technieken in het spel, deze beschrijvende technologieën helpen om het probleem van heterogene data op te lossen.

1.2. Doelstelling

Het doel van ons platform is om spoorwegondernemingen de mogelijkheid te geven hun werking te vergelijken met andere spoorwegondernemingen en om te kijken of de pijnpunten van ondernemingen op een eenvoudige manier kunnen worden weergegeven. Dit bestaat dus uit twee delen: enerzijds één onderneming evalueren en anderzijds het vergelijken van meerdere ondernemingen. De low-level KPI's worden rechtstreeks afgeleid uit de data, de high-level KPI's worden afgeleid uit meerdere low-level of high-level KPI's. Op deze manier kan een KPI-boom gecreëerd worden die een overzicht geeft van de werking van de onderneming. Uiteindelijk moet het op deze manier mogelijk zijn om verschillende bedrijven met verschillende datasyntax op een kwalitatieve manier met elkaar te vergelijken.

1.3. Tijdlijn

Om een beter beeld te krijgen van het concept van KPI's werd eerst een onderzoek uitgevoerd om te ontdekken wat een KPI exact is en wat voorbeelden van KPI's in het openbaar vervoer zijn, dit gebeurde in sectie 3.1. Er werd in sectie 3.2 gekeken naar bestaande KPI-dashboards en waar ze op focussen. Verder werd ook nagegaan welke stakeholders er allemaal meespelen bij treinondernemingen (sectie 3.3), in de eerste plaats om te kijken met welke stakeholders een bedrijf rekening moet houden en anderzijds om te bepalen welke stakeholders gebruik zullen maken van het platform. Bij het onderzoeken van de stakeholders werd ook al een eerste link gelegd met de werking van de Nationale Maatschappij der Belgische Spoorwegen (NMBS) en Infrabel, in het uitwerken van de casestudies kwamen deze bedrijven nog meer aan bod. De Britse spoorwegen werden ook behandeld, de werking in Groot-Brittannië verschilt namelijk sterk met die in België. Als laatste deel van de literatuurstudie werd ook gekeken naar ontologieën: wat is het precies en op welke manier kan dit principe van nut zijn in het platform?

Hierna volgde de analyse van het platform in Hoofdstuk 2, er werd gekeken wat er precies mogelijk moet zijn voor de gebruikers van het platform. Aan de hand hiervan werd een eerste ontwerp gemaakt van de opbouw van het platform en naar de te gebruiken technologieën.

De volgende stap in het proces was de implementatie in Hoofdstuk 4. De gekozen technologieën werden geïmplementeerd en ook de verschillende databronnen werden uitgewerkt. Hierop volgend ontwikkelden we en gebruikten we ontologieën. Vanaf dit moment konden query's

uitgevoerd worden, hiervoor werd een website gecreëerd. Het doel van de website is om KPI's te creëren en om ze te tonen op een dashboard.

Uiteraard is het van belang dat het gaat om een 'beleidsimpacterend' dashboard. Hierom werden KPI's gecreëerd en geanalyseerd door middel van opgehaalde mobiliteitsdata. De resultaten van deze KPI's worden verder onderzocht en gekaderd binnen de context in Hoofdstuk 5. Op deze manier wordt nagegaan of het platform beleidsmakers kan helpen.

Als laatste stap wordt er ook nog testing uitgevoerd in Hoofdstuk 6, op deze manier kunnen mogelijke problemen ontdekt worden.

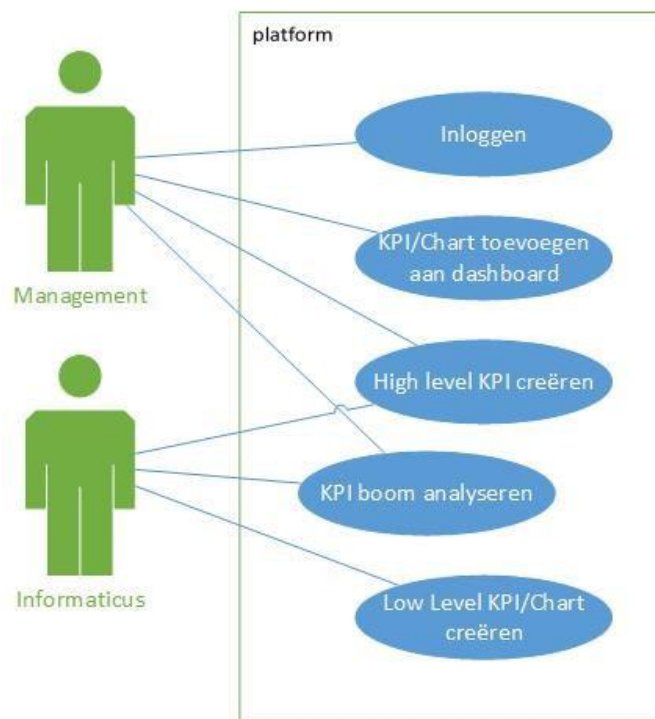
Hoofdstuk 2: Analyse van de functionaliteit

In de volgende paragrafen wordt enerzijds beschreven wat de gebruiker zal kunnen in de applicatie en anderzijds de technische omschrijving van de structuur van het volledige platform.

2.1. Use Case

Tijdens de analyse worden de functionaliteiten van de applicatie onderzocht en weergegeven in een use case-diagram. Achteraf worden deze afzonderlijk uitgewerkt.

2.1.1. Use case-diagram



Figuur 1: Use case-diagram

Op Figuur 1 is een algemeen use case-diagram te zien van de verschillende mogelijkheden voor de gebruikers. Deze worden verder besproken in de volgende sectie.

2.1.2. Mogelijkheden voor gebruikers

In deze sectie wordt kort besproken wat de functionaliteiten zijn waar de gebruikers van het platform gebruik van kunnen maken.

Inloggen

Doelstelling: De gebruiker kan een organisatie kiezen om in te loggen als user.

Actoren: Management

Postconditie: De gebruiker is ingelogd als een bepaalde organisatie

KPI/Chart toevoegen aan dashboard

Doelstelling: De gebruiker kan uit een lijst van beschikbare KPI's/Charts kiezen welke hij wilt toevoegen aan zijn gepersonaliseerde dashboard

Actoren: Management

Preconditie: Gebruiker moet ingelogd zijn

Postconditie: Na het kiezen van de KPI wordt deze waarde achterliggend berekend en weergegeven op een informatieve wijze op het dashboard.

Successscenario:

1. Gebruiker selecteert een KPI om te voegen
2. Keuze wordt doorgegeven en berekend
3. Extra pop-ups voor personalisatie moeten worden ingevuld
4. KPI is toegevoegd aan het dashboard

Alternatieve scenario:

- 2a. Fout bij het berekenen resulteert in geen toevoeging

High level KPI creëren

Doelstelling: De gebruiker kan een nieuwe high level KPI toevoegen die te gebruiken is in het dashboard

Actoren: Management/Informaticus

Postconditie: Succes bericht

Successscenario:

1. De gebruiker vult de benodigde velden in.
2. Gebruikt de rekenmachine om formules op te stellen voor het berekenen van de KPI
3. Krijgt een succes bericht en kan de KPI gebruiken in het dashboard

Alternatieve scenario:

- 3a. Krijgt onsuccesvol bericht

Low level KPI/Chart creëren

Doelstelling: De gebruiker kan een nieuwe low level KPI/Chart toevoegen die te gebruiken is in het dashboard

Actoren: Informaticus

Postconditie: Succes bericht

Successscenario:

1. De gebruiker vult de benodigde velden in.
2. Krijgt een succes bericht en kan de KPI gebruiken in het dashboard

Alternatieve scenario:

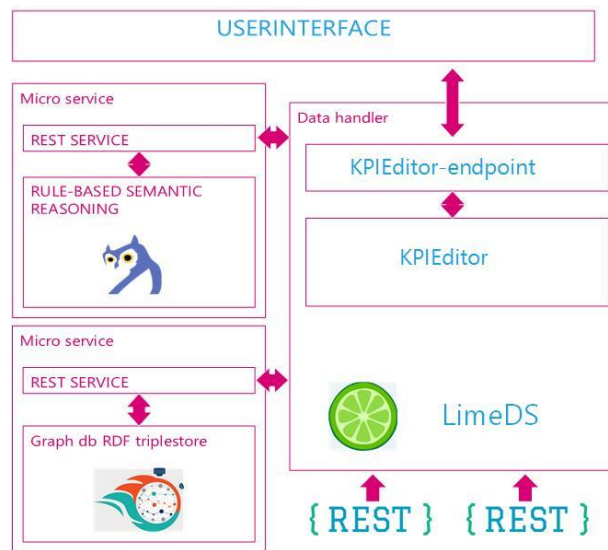
- 2a. Krijgt onsuccesvol bericht
- 2b. Kan KPI toch niet gebruiken door foutieve query's

KPI-boom analyseren

Doelstelling: De gebruiker kan de volledige KPI-boom analyseren zodanig hij een volledig zicht krijgt op alle afhankelijkheden.

Actoren: Management/Informaticus

2.2. Technische omschrijving



Figuur 2 : Dataflow van de platform.

Het platform bestaat uit vier grote blokken, zoals te zien is op Figuur 2.

Allereest is er de *Datahandler* die in eerste plaats instaat om ruwe data op te halen van de, deze nadien semantisch te verwerken en tot slot naar de databank te pushen. De *Datahandler* zorgt ook voor communicatie tussen de andere blokken. Een belangrijk onderdeel van de *Datahandler* is de KPIEditor waar het dashboard rond gebouwd is.

Ten tweede is er de databank, deze maakt gebruik van een Representational State Transfer (REST) service als communicatiemiddel met de overige blokken. Doordat deze component geïmplementeerd is als microservice is het eenvoudig om het te vervangen door een andere microservice dat gelijkaardige functies uitvoert.

De derde component is eveneens een microservice dat dient om *rule-based* semantisch te redeneren op een subset van de databank. deze maakt ook gebruik van REST om te communiceren met de rest van de blokken.

Tot slot is er de User Interface (UI) die in communicatie staat met de KPIEditor voor het gebruik van het dashboard.

Hoofdstuk 3: Literatuurstudie

Om een platform te bouwen kan best eerst een stap terug genomen worden. Er zijn namelijk een aantal vragen die nog beantwoord moeten worden. Wat is nu precies een Key Performance Indicator (KPI)? Wat zijn KPI's in het openbaar vervoer? Welke KPI-dashboard platforms bestaan er al? Welke stakeholders in de spoorwegsector hebben baat bij een dergelijk platform? Welke casestudies zullen behandeld worden? Wat is nu precies een ontologie? Waarom is er een vocabularium voor het beschrijven van vervoerssystemen, routes, tussenstops en dergelijke meer nodig en hoe werkt dit precies? Op al deze vragen wordt een antwoord gezocht in onderstaande paragrafen.

3.1. Key Performance Indicators in het openbaar vervoer

In dit onderzoek zal dieper ingegaan worden op de mogelijke KPI's in het openbaar vervoer. In de eerste plaats zal gedefinieerd worden wat een KPI precies is. Hierna zal gekeken worden welke KPI's van toepassing kunnen zijn in het openbaar vervoer in het algemeen en met welke zaken best rekening gehouden kan worden. Als laatste wordt een conclusie gevormd.

3.1.1. Wat is een KPI?

Key Performance Indicators zijn meetwaarden die gebruikt worden om de performantie van een onderneming te analyseren. Deze indicatoren zijn van belang om verschillende ondernemingen op een kwalitatieve manier met elkaar te vergelijken. Verder kan aan de hand van deze indicatoren bepaald worden welke aanpassingen een positieve invloed hebben op de toekomstige werking van het bedrijf [3]. Een mogelijk voorbeeld hierbij is dat het verhogen van de stiptheid een positief effect heeft op de klanttevredenheid. Die klanttevredenheid zal mogelijk ook voor meer reizigers en dus meer inkomsten zorgen.

Een eerste onderscheid tussen verschillende KPI's kan gemaakt worden op basis van objectiviteit. Objectieve meetwaardes zijn bijvoorbeeld het aantal reizigers of het aantal afgelegde kilometers. Subjectieve meetwaardes focussen eerder op klanttevredenheid via bijvoorbeeld een enquête [4]. In de volgende paragrafen worden voornamelijk objectieve indicatoren behandeld, enkel in *Activiteiten* bij het puntje 'Klanttevredenheid' vinden we subjectieve KPI's terug.

Key Performance Indicators steunen op prestatiemetingen. Maar wat is nu eigenlijk een correcte manier om prestatiemetingen uit te voeren? In volgende lijst [5] bevinden zich een aantal eigenschappen en vereisten van prestatiemetingen.

- Meetbaarheid: meetbaar met middelen die binnen het budget passen
- Voorspelbaarheid: bruikbaar om toekomstige strategieën te beoordelen
- Duidelijkheid: verstaanbaar voor zowel management als het grote publiek
- Bruikbaarheid: directe meting van een bepaald probleem
- Multimodaliteit: bruikbaar voor alle relevante modes binnen het probleem (bv. zowel voor treinen als bussen)
- Tijdelijkheid: metingen moeten analyseerbaar zijn binnen bepaalde tijdsperiodes
- Veelzijdigheid: metingen moeten geschikt zijn voor elk doel van het systeem
- Controle: metingen moeten de organisaties in staat stellen om de gemeten karakteristiek te controleren en te verbeteren
- Relevantie: metingen moeten besluitvormers relevante informatie geven om beslissingen te maken

Om een goede werking met KPI's te voorzien is het belangrijk om rekening te houden met een aantal factoren. In de eerste plaats moet je voldoende KPI's hebben om een goed beeld te krijgen van de performantie van het bedrijf, anderzijds kan een overaanbod aan KPI's ervoor zorgen dat er te veel data opgehaald en bewaard moet worden. Hierom is het van belang dat de belangrijkste KPI's bepaald worden vooraleer er begonnen wordt met het ophalen van data.

3.1.2. Case: Openbaar vervoer

In deze masterproef wordt een onderzoek gedaan naar de verschillende KPI's in het openbaar vervoer. Onder openbaar vervoer worden treinen, bussen, taxi's, vliegtuigen, autodelen en fietsdelen beschouwd. In de praktijk omvat dit dus alle gedeelde transportmethodes.

Het is van belang dat een consistente naamgeving wordt gebruikt die van toepassing is op zowel treinen, bussen en andere vervoersmiddelen. In de volgende paragrafen zal de term **halte** worden gebruikt om een station, bushalte, luchthaven en dergelijke meer aan te duiden. Met een **voertuig** wordt bijvoorbeeld een treinwagon of een vliegtuig bedoeld. De spoorverbinding of de

weg tussen twee haltes wordt het **traject** genoemd. Het uitvoeren van de verplaatsing op een traject wordt de **service** genoemd.

Om een strategisch prestatiegericht plan op te stellen om de werking van transportsystemen te verbeteren wordt best rekening gehouden met volgende factoren [5]:

- Definieer algemene doelen en doelstellingen voor het transportplan.
- Identificeer specifieke prestatiegerichte doelstellingen in een objectieve, meetbare vorm.
- Identificeer specifieke prestatiegerichte maatregelen om te gebruiken bij het meten en beoordelen van de relevante resultaten van elk component van het transportplan.
- Herken de factoren die aangepast kunnen worden om de prestatie van het transportsysteem positief te beïnvloeden.
- Beschrijf de benodigde bronnen om de doelen te bereiken.

Een plan is niets zonder uitwerking, om de metingen van de prestatie op een correcte en eenduidige manier te laten verlopen wordt best rekening gehouden met volgende overwegingen [5]:

- Prestatiemeting moet de tevredenheid van de klant over de service in rekening brengen zonder de belangen van de systeemoperatoren te verwaarlozen.
- Prestatiemeting voor, tijdens en na de service kan de mogelijkheid om problemen te ontdekken en op te lossen makkelijker maken.
- Prestatiemeting kan er voordeel bij doen om live feedback te krijgen van gebruikers.
- Transportorganisaties mogen geen zachte maatregelen (bv. veiligheidsgevoel van reizigers) verwaarlozen ten voordele van harde maatregelen (bv. het aantal ongevallen).
- Prestatiemeting moet een afweging maken tussen behoeften en voordelen op korte en lange termijn

Naast al deze voorwaarden is het uiteraard ook van belang waar de benodigde data gevonden kan worden. Enerzijds kan data gehaald worden uit enquêtes, deze kunnen uitgevoerd worden op werkplaatsen, aan de haltes en dergelijke meer [5]. Verder wordt de meeste data gehaald uit het gebruik van interne en externe API's die de reisinformatie van het openbaar vervoer beschikbaar stellen.

3.1.3. Overzicht van mogelijke KPI's

Om een goede keuze te kunnen maken tussen de KPI's die behandeld zullen worden wordt eerst op zoek gegaan naar verschillende mogelijke KPI's. Onderstaande categorieën worden afgeleid uit het verslag van InteGRail; deze onderverdeling is een consensus tussen verschillende treinoperators in verschillende Europese landen [3]. In de onderverdeling binnen deze categorieën worden hier en daar aanpassingen doorgevoerd om relevante KPI's, die gevonden worden op andere bronnen, in te passen. Wanneer de KPI ontdekt werd door onszelf staat er geen bronvermelding bij. De vier grote categorieën zijn: voertuigbeheer, activiteiten, verkeersbeheer en infrastructuur.

Voertuigbeheer

Een eerste categorie waar rekening mee gehouden wordt is het voertuigbeheer. Dit omvat alles wat te maken heeft met de voertuigen: wat kan er misgaan, op welke manier wordt dit opgelost of voorkomen en hoe efficiënt gebeurt dit? Tevens wordt ook rekening gehouden met de milieuefficiëntie van de voertuigen.

1. Onbeschikbaarheid

Voertuigen zijn alleen nuttig wanneer ze effectief gebruikt kunnen worden, de tijd die het voertuig doorbrengt in onbeschikbare toestand is per definitie verloren tijd.

1.1 Ongepland

Deze onbeschikbaarheid ontstaat wanneer een voertuig plots faalt of beschadigd is. Hierin wordt het onderscheid gemaakt tussen de tijd die verloren gaat en het aantal voertuigen dat buiten gebruik is. Aangezien deze reparaties dan zonder voorbereiding verlopen kan het zijn dat er eerst gewacht moet worden op de juiste materialen om de reparaties uit te voeren. In tweede instantie zal er dan overgegaan worden op de effectieve reparatie door personeel.

KPI	Beschrijving	Bron
Ongepland	Hoeveel voertuigen zijn onbeschikbaar ten gevolge van ongeplande onderhoudswerken aan de voertuigen?	[3]
Schade	Hoeveel voertuigen zijn onbeschikbaar door beschadigde componenten ten gevolge van externe factoren?	[3]
Falen	Hoeveel voertuigen zijn onbeschikbaar door het niet werken van bepaalde componenten ten gevolge van interne factoren?	[3], [6]
Materiaalgebrek voor reparatie	Hoeveel tijd gaat verloren ten gevolge van een gebrek aan materiaal om reparaties uit te voeren?	[3]
Reparatietijd	Hoelang duurt het om de reparaties uit te voeren?	[3]

1.2 Gepland

Om een goede werking van de voertuigen te verzekeren is het vanzelfsprekend dat er periodiek onderhoud wordt uitgevoerd. Hierin wordt het onderscheid gemaakt tussen de tijd die verloren gaat en het aantal voertuigen dat buiten gebruik is.

KPI	Beschrijving	Bron
Gepland	Hoeveel voertuigen zijn onbeschikbaar ten gevolge van geplande onderhoudswerken aan de voertuigen?	[3]
Onderhoudstijd	Hoeveel tijd gaat verloren ten gevolge van geplande onderhoudswerken en reparaties aan de voertuigen?	[3]

2. Betrouwbaarheid

Het is uiteraard niet alleen van belang hoeveel tijd de voertuigen doorbrengen in onbeschikbare toestand, het is mogelijks nog belangrijker dat de tijd tussen opeenvolgende onbeschikbaarheden zo lang mogelijk is. In dit opzicht is het belangrijk na te gaan hoe veel tijd er zit tussen opeenvolgende reparaties aan voertuigen.

KPI	Beschrijving	Bron
Tijd tussen onbeschikbaarheid	Hoeveel tijd zit er gemiddeld tussen opeenvolgende onbeschikbaarheden van een voertuig?	[3]

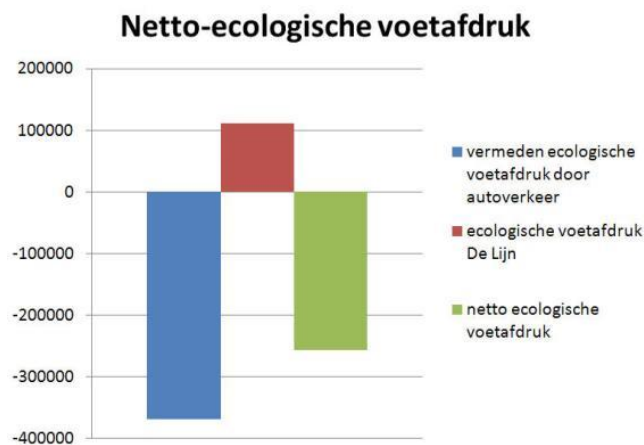
3. Uitgaven

Om een winstgevende onderneming te zijn mogen de uitgaven niet te hoog oplopen; het is dan ook van groot belang om de uitgaven in rekening te brengen. Indien de uitgaven gereduceerd kunnen worden zonder dat de geleverde service kwalitatief achteruit gaat heeft dit uiteraard alleen maar voordelen. Hierboven werd reeds aangehaald dat bij onderhoud en reparatie er tijd verloren gaat; deze tijd bestaat uit de periode dat het personeel werkt en eventueel de tijd die gewacht moet worden op de materialen. Dit is niet enkel tijd die verloren gaat maar ook geld, om personeel te laten werken is een bepaalde loonkost nodig en om over het nodige materiaal te beschikken moet dit eerst worden aangekocht.

KPI	Beschrijving	Bron
Onderhoud	Welk bedrag is nodig om het geplande onderhoud te financieren?	
-Personeel	Hoeveel van deze kosten gaan naar personeel?	[3]
-Materiaal	Hoeveel van deze kosten gaan naar materialen?	[3]
Reparatie	Welk bedrag is nodig om reparaties te financieren?	
-Personeel	Hoeveel van deze kosten gaan naar personeel?	[3]
-Materiaal	Hoeveel van deze kosten gaan naar materialen?	[3]
Beheer	Welk bedrag is nodig om het logistieke beheer van de voertuigen te financieren?	
Productie	Welk bedrag is nodig om een voertuig te produceren?	

4. Milieu

In de huidige tijd speelt milieu een belangrijke rol, er kans best gezorgd worden dat de ecologische voetafdruk van de voertuigen zo klein mogelijk is. Wat is de levensduur van een voertuig, hoeveel schadelijke stoffen worden uitgestoten en wat is de voetafdruk bij de productie of afbraak van een voertuig zijn een aantal belangrijke vragen waarop een antwoord gezocht moet worden. Een ander interessant punt is hoeveel bespaard wordt door gebruik te maken van deze voertuigen in tegenstelling tot de meest gebruikelijke voertuigen (auto's) voor verplaatsingen, een voorbeeld hiervan is te zien op Figuur 3.



Figuur 3: Wanneer het verbruik en de productie in rekening wordt gebracht met de besparing kan je er de netto-ecologische voetafdruk uit afleiden. Dit is kwalitatief interessanter dan de zuivere verbruiks- en productievoetafdruk. Op de verticale as ziet u de ecologische voetafdruk in hectaren uitgedrukt. [7]

KPI	Beschrijving	Bron
Verbruik	Wat is de ecologische voetafdruk die wordt nagelaten door de voertuigen per afgelegde km?	[7]
Besparing	Wat is de ecologische voetafdruk die vermeden wordt door het gebruik van openbaar vervoer t.o.v. autoverkeer per afgelegde km?	[7]
Productie	Wat is de ecologische voetafdruk die wordt nagelaten ten gevolge van de productie van een voertuig?	
Levensduur	Hoe lang gaat een voertuig mee?	
Afbraak	Wat is de ecologische voetafdruk die wordt nagelaten bij de afbraak van een voertuig?	

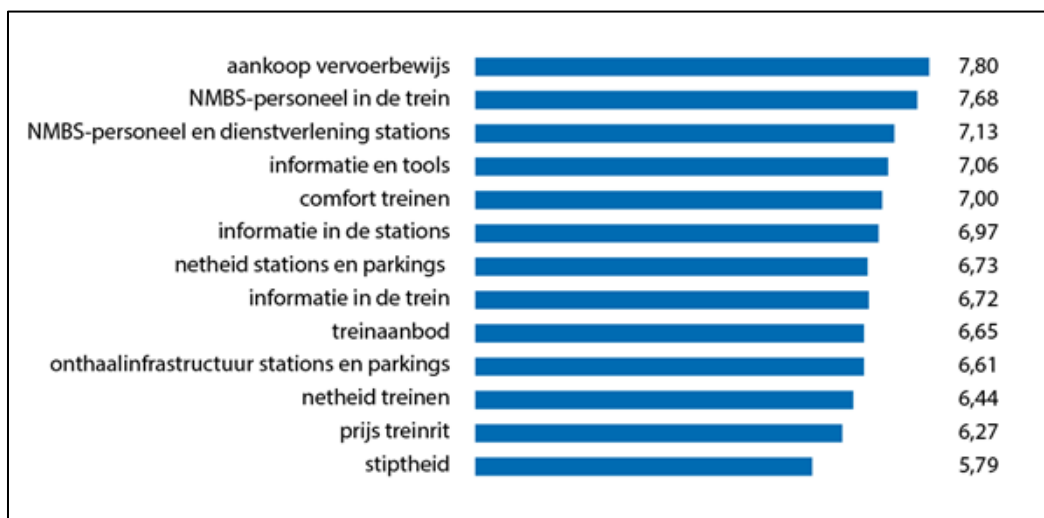
Activiteiten

Activiteiten gaan over de geleverde service. In eerste instantie wordt gepeild naar de tevredenheid van de klanten over de service, hierna wordt in rekening gebracht hoeveel het kost om de service (de reis van punt A naar punt B) te financieren. Verder wordt bekeken hoeveel de klanten betalen voor de service, hoeveel klanten gebruik maken van de service en hoe het gesteld is met de veiligheid van de klanten gedurende de service.

1. Klanttevredenheid

De klanttevredenheid is van groot belang voor elk bedrijf. Indien de klanten tevreden zijn over de geleverde service zullen ze ook in de toekomst gebruik willen maken van deze service. Bij voorkeur zal dan ook op regelmatige basis gepeild worden naar de tevredenheid van de klanten over een aantal van onderstaande aandachtspunten betreffende de service. In de eerste plaats willen de reizigers voldoende en correct geïnformeerd worden over de reisinformatie. Daarnaast is het ook van belang dat de klanten tevreden zijn over de stiptheid van de service. Een andere factor is comfort: voor bepaalde reizigers is dit erg belangrijk en voor anderen is dit slechts een bijzaak.

De NMBS voert voortdurend enquêtes uit in de stations, ze laten reizigers surfen naar een link waarop ze een vijftien minuten durende vragenlijst kunnen beantwoorden. Hierbij wordt een subjectieve score gevraagd op verschillende zaken zoals stiptheid en comfort. De resultaten voor 2015 zijn te zien op Figuur 4 [8].



Figuur 4 : Resultaten tevredenheidsenquête NMBS 2015 [8].

KPI	Beschrijving	Bron
Accuraatheid informatie	In welke mate zijn de klanten tevreden over de geleverde informatie omtrent de service. Hierbij wordt vooral de vraag gesteld of de aangeleverde informatie effectief overeenkomt met de werkelijke situatie.	[3], [6]
Beschikbaarheid informatie	In welke mate zijn de klanten tevreden over de geleverde informatie omtrent de service. Hierbij wordt vooral de vraag gesteld hoe gemakkelijk het is voor de klant om informatie in verband met de service terug te vinden.	[3], [6]
- Voertuig	Hoe tevreden zijn de klanten over de beschikbaarheid van informatie in het voertuig?	[9]
- Halte	Hoe tevreden zijn de klanten over de beschikbaarheid van informatie aan de halte?	[9]
Punctualiteit	Hoe tevreden zijn de klanten over de stiptheid van de service?	[6]
Comfort	Hoe tevreden zijn de klanten over het comfort?	
- Netheid	Hoe tevreden zijn ze over de hygiënische omstandigheden in de voertuigen?	[6], [9]
- Drukke	Hoe druk vinden de reizigers het in de voertuigen?	[4]
- Capaciteit zitplaatsen	Hoe tevreden zijn ze over het aantal beschikbare zitplaatsen in het algemeen?	[4]
Prijsbeoordeling	Hoe tevreden zijn de klanten over de prijs ten opzichte van de geleverde service?	[8]
Service-aanbod	Hoe tevreden zijn de klanten over de mogelijke bestemmingen/verbindingen?	[8]

Bereikbaarheid halte	Hoe tevreden zijn de klanten over de bereikbaarheid/plaatsing van de haltes?	[8], [9]
Snelheid tot eindbestemming	Hoe tevreden zijn de klanten over de snelheid waarmee ze van plaats A naar plaats B reizen?	
Aantal klachten	Hoeveel klachten worden er ontvangen?	[6]
Snelheid afhandeling klachten	Hoe tevreden zijn de klanten over de snelheid waarmee hun klachten worden afgehandeld?	[6]
Resultaat afhandeling klachten	Hoe tevreden zijn de klanten over de reactie de verkregen werd omtrent hun klacht?	[6]
Beschikbaar voor rolstoelgebruikers	Hoe tevreden zijn de rolstoelgebruikers over de mogelijkheden om van de service gebruik te maken? Hoe goed is de service voor rolstoelgebruikers?	[6], [9]
Veiligheidsgevoel	Hoe is het gesteld met het veiligheidsgevoel van de reizigers op verschillende plaatsen horend bij de service?	[9]
- Voertuig	Hoe veilig voelen de reizigers zich in de voertuigen?	[9]
- Halte	Hoe veilig voelen de reizigers zich aan de halte?	[9]

2. Uitgaven

Om een winstgevende onderneming te zijn mogen de uitgaven niet hoog oplopen; het is dan ook van groot belang om de uitgaven in rekening te brengen. Indien de uitgaven gereduceerd kunnen worden zonder dat de geleverde service kwalitatief achteruit gaat heeft dit uiteraard alleen maar voordelen. Dit is echter een simplistische manier, om een gefundeerde conclusie te trekken zal allicht een bijkomende kostenanalyse nodig zijn.

KPI	Beschrijving	Bron
Gebruik voertuigen	De kosten die gepaard gaan met het gebruik van de voertuigen.	[3]
Gebruik personeel	De kosten die gecreëerd worden door het gebruik van personeel met betrekking tot de service.	
- Planning	De kosten die worden gemaakt om de verschillende werkschema's van de werknemers in te plannen.	[3]
- Opleiding	De kosten die worden gemaakt om het personeel op te leiden voor enerzijds hun basistaak uit te voeren en anderzijds bijkomende cursussen voor veiligheid (bijvoorbeeld Eerste Hulp Bij Ongelukken (EHBO)).	[3]
- Loon	De gemiddelde maandelijkse loonkost aan een personeelslid.	[3]
Compensatie	Gaande van terugbetalingen door te grote vertragingen voor de klanten tot compensatie van secundaire vertragingen bij andere operatoren.	[9]

3. Inkomsten

Elk bedrijf heeft een zekere omzet nodig om te kunnen bestaan. De inkomsten zijn dus van cruciaal belang.

KPI	Beschrijving	Bron
Prijs per kilometer	Wat is de gemiddelde prijs voor de reiziger per kilometer?	[10], [11]
Prijs per rit	Wat is de gemiddelde prijs voor de reiziger per rit?	[10]
Aantal betalers	Hoeveel van de gecontroleerde reizigers hebben ook effectief betaald voor hun vervoersbewijs?	[6]

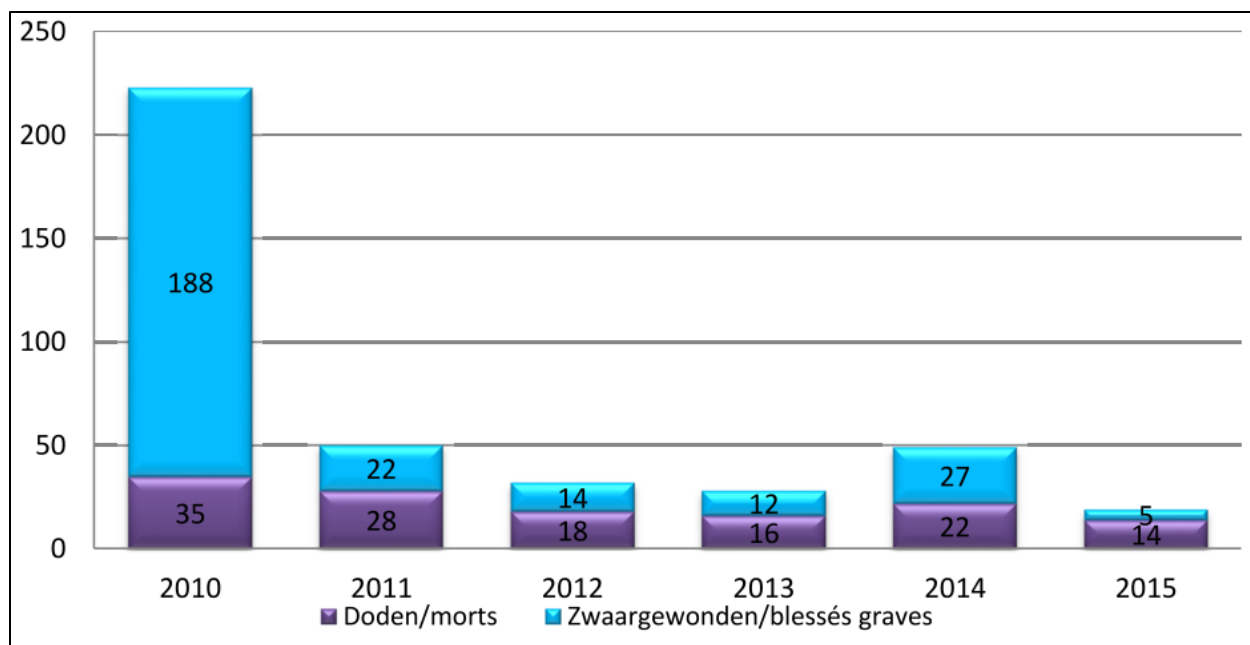
4. Bezetting

Er moet een afweging gemaakt worden tussen het aantal zitplaatsen en het aantal reizigers, het heeft geen nut om te investeren in zitplaatsen die zelden of nooit gebruikt worden. Tevens is het belangrijk te weten hoeveel reizigers gebruik maken van de service.

KPI	Beschrijving	Bron
Bezetting voertuig	Hoeveel reizigers zitten er in een voertuig?	[3], [6], [11]
Reizigerkilometers	Het aantal kilometer dat reizigers afleggen.	[6], [11]
Niet-reizigerkilometers	Het aantal kilometer dat de voertuigen gebruikt worden zonder service, bijvoorbeeld om terug te keren naar de stelplaats.	[12]

5. Veiligheid en beveiliging

Veiligheid is één van de grootste pijlers bij het openbaar vervoer, mensen willen op een veilige manier van punt A naar punt B kunnen reizen. Het is dan ook vanzelfsprekend dat bijgehouden wordt welke feiten het vaakst voor onveilige situaties zorgen en wat hiervan de oorzaken zijn. Ook het aantal doden of zwaargewonden zijn een belangrijke meetwaarde. Uiteraard is het de bedoeling om dit aantal zo laag mogelijk te houden. Niet alleen de hoeveelheid is van belang maar ook de kost die hiermee gepaard gaat. Alle indicatoren hieromtrent zijn gedefinieerd door de Europese Unie, ze worden Common Safety Indicators of kortweg CSI genoemd. Wanneer in onderstaande tabellen referentie [13] vermeld wordt, betekent dit dat het gaat om een CSI. Een voorbeeld hiervan is te zien op Figuur 5.



Figuur 5: Evolutie van het aantal doden/zwaargewonden bij Infrabel, in 2015 was dit cijfer lager dan voorheen.
[13]

Een ander aspect van veiligheid en beveiliging zit hem in de manier waarop het personeel reageert op onverwachte omstandigheden. Terroristische activiteiten zijn spijtig genoeg een realiteit en komen alsmaar meer in het nieuws. Niet enkel vliegtuigen worden gevisieerd zoals in 9/11, maar ook steeds vaker het dagelijkse pendelverkeer. De twintig doden na een bomaanslag in de metro van Brussel klinkt ons nog bekend in de oren. Onderstaande KPI's leggen dan ook de focus op risicomanagement in openbaar vervoer, met name veiligheid en beveiliging bij het gebruiken van het transport als service.

5.1 Feiten

In deze onderverdeling wordt een telling bijgehouden van het aantal ongevallen en criminele feiten.

KPI	Beschrijving	Bron
Botsingen	Hoeveel botsingen tussen een voertuig en een ander voertuig of object vinden er plaats?	[13]
Afwijkingen van het traject	Hoeveel keer wijken voertuigen af van het voorbestemde traject? Hierbij worden bijvoorbeeld ontsporingen bedoeld.	[13]
Persoonsongevallen	Hoeveel persoonsongevallen vinden er plaats? Dit omvat ook personen die van of in de voertuigen vallen of door losse voorwerpen in het voertuig geraakt worden.	[13]
Brand	Hoeveel keer ontstaat er brand in de voertuigen?	[13]
Andere types ongevallen	Hoeveel ongevallen vinden er plaats die niets te maken hebben met bovenstaande redenen?	[13]
Criminaliteit	Hoeveel criminele feiten (exclusief zwartrijden) vinden er plaats?	[14]

5.2 Gevolgen

Hoeveel dodelijke slachtoffers of zwaargewonden er vallen in onderstaande categorieën ten gevolge van de service.

KPI	Beschrijving	Bron
Reizigers	Hoeveel zwaargewonde of dode reizigers zijn er ten gevolge van een ongeval? Zie Figuur 5.	[13]
Personeel	Hoeveel zwaargewonde of dode personeelsleden zijn er ten gevolge van een ongeval?	[13]

Indringer	Hoeveel zwaargewonde of dode indringers zijn er ten gevolge van een ongeval? Indringers zijn alle mensen die zich op de voertuigen of infrastructuur bevinden zonder toestemming.	[13]
Andere persoon	Hoeveel zwaargewonde of dode personen die niet in bovenstaande categorieën passen zijn er ten gevolge van een ongeval?	[13]

5.3 Oorzaak

Hoeveel dodelijke slachtoffers of zwaargewonden er vallen ten gevolge van onderstaande oorzaken.

KPI	Beschrijving	Bron
Infrastructuurproblemen	Hoeveel ongevallen zijn het gevolg van problemen bij de infrastructuur? Bij treinen zijn dit bijvoorbeeld gebroken rails of knikken in het spoor.	[13]
Voertuigproblemen	Hoeveel ongevallen zijn het gevolg van problemen bij het voertuig? Bij treinen is dit bijvoorbeeld een gebroken wiel/as.	[13]
Foutieve seingeving	Hoeveel ongevallen zijn het gevolg van foutieve seingeving?	[13]
Negeren seingeving	Hoeveel ongevallen zijn het gevolg van het negeren van de seingeving door de bestuurder?	[13]
Weersomstandigheden	Hoeveel ongevallen zijn het gevolg van de weersomstandigheden?	

5.4 Gevolgen

Wat is de kost die moet betaald worden vanwege problemen bij de geleverde service?

KPI	Beschrijving	Bron
Personen	De kosten die gemaakt moeten worden ten gevolge van ongevallen met personen. Hierbij wordt rekening gehouden met drie factoren: medische kosten, proceskosten en het productieverlies (waarde van de goederen/diensten die de persoon in kwestie had kunnen produceren voor de samenleving indien er geen ongeval was)?	[13]
Milieuschade	Hoeveel kosten moeten er gemaakt worden door de onderneming om de milieusituatie te herstellen naar de staat waarin het zich bevond voor het ongeval?	[13]
Materiële schade	Hoeveel kosten worden er gemaakt om de infrastructuur of voertuigen terug te herstellen ten gevolge van ongevallen?	[13]
Vertraging	Wat is de economische impact van vertragingen? Dit omvat bijvoorbeeld het niet op tijd aankomen op het werk van werknemers die gebruiken maken van de service.	[13]

5.5 Inbreuk op de beveiliging

Bij het gebeuren van deze inbreuk is het een prioriteit dat deze snel wordt gedetecteerd en wordt opgelost.

KPI	Beschrijving	Bron
Convergentie	De tijd tussen het detecteren en oplossen van het conflict.	[15]
Communicatie	Hoe verloopt de communicatie tussen het personeel en geven ze uniforme richtlijnen aan de aanwezige klanten?	[15]

5.6 Evacuatie

Snelheid is een topprioriteit bij het plaatsvinden van een evacuatie. Deze hangt af van verschillende factoren zoals duidelijke aangeving van noodweg en onbelemmerde doorgang alsook de effectiviteit en efficiëntie van veiligheidsofficieren.

KPI	Beschrijving	Bron
Aangeving nooduitgang	Bij evacuatie is het van belang dat de nooduitgang snel en eenvoudig kan gevonden worden.	[15]
Convergentie	De tijd tussen het detecteren van een probleem en het veilig evacueren.	[15]
Communicatie	Hoe verloopt de communicatie tussen het personeel en geven ze uniforme richtlijnen aan de aanwezige klanten?	[15]

5.7 Controle over panieksituaties

Ook hier is snelheid een topprioriteit om paniek zo snel mogelijk weer onder controle te krijgen. Voorzien van EHBO-posten, genoeg veiligheidsfunctionarissen en een duidelijke communicatielijn zijn een must.

KPI	Beschrijving	Bron
Communicatie	Hoe verloopt de communicatie tussen het personeel en geven ze uniforme richtlijnen aan de aanwezige klanten?	[15]
Veiligheidsapparatuur	Hoe bereikbaar zijn de aanwezige hulpmiddelen?	[15]

Verkeersbeheer

In verkeersbeheer wordt de nadruk gelegd op het managementaspect. Hoe verloopt de communicatie richting reizigers, hoe groot is de vertraging en vanwaar komt ze? Als laatste wordt ook nog in rekening gebracht in welke mate het beschikbare netwerk ten volle benut wordt.

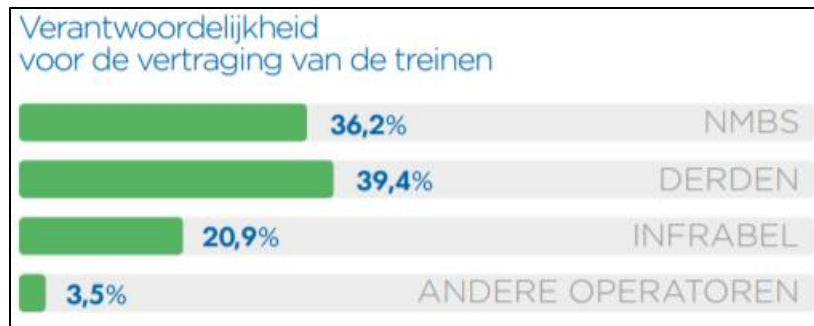
1. Informatie

Voor een vlotte werking van de service is het belangrijk dat alle mensen die geïnformeerd moeten worden over de status van de service, deze informatie snel en correct kunnen ontvangen. In de eerste plaats is het van belang dat de reizigers tijdig op de hoogte worden gebracht over de status van de service waarvan ze gebruik willen maken, anderzijds moet ook het personeel de nodige informatie krijgen om met de veranderde situatie om te kunnen gaan. Als laatste is het ook van belang dat alle service-informatie bij het management terechtkomt zodat analyse mogelijk is.

KPI	Beschrijving	Bron
Informatie voor reizigers	Hoe snel worden reizigers op de hoogte gebracht van vertragingen?	[3], [6]
Informatie voor personeel	Hoe snel wordt personeel op de hoogte gebracht van wijzigingen in de planning?	[3], [6]
Informatie voor management	De informatie die beschikbaar is voor het management, gaande van logs tot verslagen.	[3], [6]

2. Regulatie/Herschikking van de planning

In een ideaal scenario verloopt alles volgens het vooropgestelde plan. In de praktijk is dit vaak anders, de service kan vertraging oplopen en dit kan in een aantal verschillende vormen. Het is uiteraard ook van belang dat een onverwachte storing het verkeer voor de rest van de dag niet in de war stuurt. Er wordt een onderscheid gemaakt tussen verschillende vormen van vertraging, de oorzaak ligt namelijk niet altijd bij het voertuig zelf, zie ook Figuur 6.



Figuur 6: Uit de cijfers van Infrabel blijkt dat bijna 40% van de vertraging veroorzaakt wordt door derden. [16]

KPI	Beschrijving	Bron
Primaire vertragingen	Gemiddelde vertraging van een bepaald voertuig die niet veroorzaakt werd door andere voertuigen.	[3]
Secundaire vertragingen	Gemiddelde vertraging van andere voertuigen veroorzaakt door een bepaald voertuig.	[3]
Geïnduceerde vertraging	Gemiddelde vertraging van een bepaald voertuig ten gevolge van primaire vertraging van andere voertuigen.	
Vertraging door derden	Hoeveel vertraging wordt veroorzaakt door derden? Derden zijn alle mensen die niet voor de bevoegde transportbedrijven werken, dit zijn dus reizigers of onbevoegden.	[16]
Afschaffingen	Het aantal afgeschafte services.	[6]
Onvoltooide services	Het aantal onvoltooide services.	[6]
Aantal services volgens schema	Het aantal voertuigen waarbij de vertraging kleiner is dan de maximale vooropgestelde vertraging (bij Infrabel: 6 minuten [16]) zodat het bestaande schema geen impact ondervindt.	[6]
Aantal services volgens schema tijdens piekuren	Het aantal voertuigen waarbij de vertraging kleiner is dan de maximale vooropgestelde vertraging zodat het bestaande schema geen impact ondervindt. In dit specifieke geval wordt enkel	[12]

	gekeken naar de piekuren, dit is dus in de weekdagen 's ochtends en 's avonds. Op deze momenten zijn de meeste pendelaars op stap en is het aantal reizigers bijgevolg het hoogst.	
Convergentie	De tijd die gemiddeld nodig is om wanneer er afgeweken wordt van het bestaande schema terug te evolueren naar de normale werking volgens het bestaande schema.	
Vermijdbare ongelukken	Het aantal ongevallen die vermeden kunnen worden. Deze worden veroorzaakt door slecht onderhoud of fouten bij de besturing.	[6]

3. Bezetting

Er moet geteld worden hoeveel treinen passeren langs een bepaald traject of een bepaald knooppunt, dit is van belang om te kunnen analyseren of het netwerk ten volle benut wordt.

KPI	Beschrijving	Bron
Bezetting traject	Hoeveel van de beschikbare trajecten worden er effectief gebruikt?	[3]
Doorstroom	Hoeveel voertuigen passeren er?	
- Halte	Langs een halte.	[10]
- Traject	Langs een bepaald traject.	[10]
Doorstroomcapaciteit	Hoeveel voertuigen kunnen er maximaal passeren?	
- Halte	Langs een halte.	
- Traject	Langs een bepaald traject.	

Infrastructuur

In deze categorie wordt alles in verband met de infrastructuur waarvan de voertuigen en reizigers gebruik maken beschreven. Dit bevat dus de trajecten waar de voertuigen gebruik van maken maar ook bijvoorbeeld de gebouwen waar tickets aangekocht kunnen worden. Hierin wordt tijdverlies door onbeschikbaarheid, de kosten en de milieuvriendelijkheid behandeld.

1. Onbeschikbaarheid

De trajecten zijn alleen nuttig wanneer ze effectief gebruikt kunnen worden, de tijd die het traject doorbrengt in onbeschikbare toestand is enerzijds verloren tijd en anderzijds een investering in de toekomst.

1.1 Ongepland

Het ongeplande onderhoud is van toepassing wanneer er plots een probleem opduikt met de infrastructuur en dit opgelost moet worden.

KPI	Beschrijving	Bron
Onverwacht onderhoud	Hoeveel tijd gaat verloren ten gevolge van onverwacht onderhoud?	[3]
Reparatie na falen	Hoeveel tijd gaat verloren door reparatie na falen van componenten?	[3]

1.2 Gepland

Het geplande onderhoud zal periodiek plaatsvinden. Het is van belang dat hierbij niet te veel tijd verloren gaat.

KPI	Beschrijving	Bron
Onderhoud	Hoeveel tijd gaat verloren ten gevolge van geplande onderhoudswerken aan de infrastructuur?	[3]

2. Betrouwbaarheid

Het is uiteraard niet alleen van belang hoeveel tijd de infrastructuur onbeschikbaar is, het is mogelijks nog belangrijker dat de tijd tussen opeenvolgende onbeschikbaarheden zo lang mogelijk is.

KPI	Beschrijving	Bron
Tijd tussen onbeschikbaarheid	Hoeveel tijd zit er gemiddeld tussen opeenvolgende onbeschikbaarheden van de infrastructuur?	[3]

3. Conditie

Onder conditie verstaan we de mate waarin de infrastructuur nog voldoet aan zijn originele capaciteiten.

KPI	Beschrijving	Bron
Conditie	De conditie van de componenten in de infrastructuur.	[3]

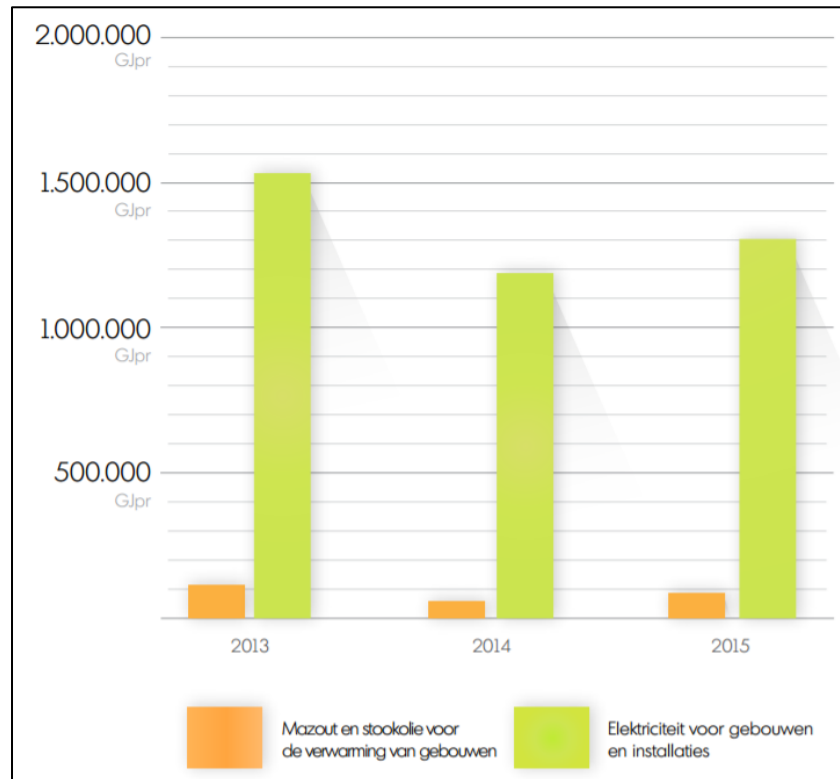
4. Uitgaven

Om een winstgevende onderneming te zijn moeten de uitgaven in evenwicht zijn met de inkomsten. Het is dan ook van groot belang om de uitgaven in rekening te brengen. Indien de uitgaven gereduceerd kunnen worden zonder dat de geleverde service kwalitatief achteruit gaat heeft dit uiteraard alleen maar voordelen.

KPI	Beschrijving	Bron
Aanleg/Productie	Hoeveel kosten worden er gemaakt om de infrastructuur te plaatsen?	
Onderhoud	Hoeveel kosten worden er gemaakt om de infrastructuur waar de service gebruik van maakt te onderhouden?	[3]
Gebruik	Hoeveel kosten worden er gemaakt om van de infrastructuur gebruik te maken?	[3]

5. Milieu

Eén van de belangrijkste zaken in moderne tijden is de ecologische voetafdruk van de onderneming. Welke impact heeft de productie en afbraak van de infrastructuur op het milieu? Bij de infrastructuur wordt er zoals eerder vermeld ook de gebouwen gerekend, zie ook Figuur 7.



Figuur 7: Energieverbruik Infrabel 2013-2015 in GigaJoule primaire energie [17].

KPI	Beschrijving	Bron
Verbruik	Hoeveel energie wordt er per jaar verbruikt in de infrastructuur?	[17]
Productie	Wat is de ecologische voetafdruk die wordt nagelaten ten gevolge van de productie van de infrastructuur?	
Levensduur	Hoe lang gaat de infrastructuur mee?	
Afbraak	Wat is de ecologische voetafdruk die wordt nagelaten bij de afbraak van de infrastructuur?	

3.1.4. Conclusie

Wat kan er besloten worden uit deze studie over KPI's in het openbaar vervoer? Om te beginnen blijkt dat er een groot aantal meetwaarden zijn die als KPI's beschouwd kunnen worden. Dit maakt het natuurlijk niet gemakkelijk om al deze waarden te meten en te bewaren voor analyses moet het nodige budget voor handen zijn. Het is dus van groot belang om te kijken welke KPI's potentieel de grootste invloed kunnen hebben op de prestaties van een onderneming. Hier schuilt dus een probleem, om te weten te komen of een KPI een grote invloed heeft moet deze eerst geanalyseerd kunnen worden. In dit onderzoek wordt verder niet meer gefocust op de potentiële kost om alle KPI's te bewaren, maar het is wel een belangrijke factor in een realistische situatie.

Het is duidelijk dat het overzicht van de KPI's slechts een indicatie geeft van de mogelijke KPI's. Elke KPI kan wanneer gewenst nog verder opgesplitst worden, zo kan bijvoorbeeld het verbruik in de infrastructuur opgedeeld worden per component (halte, traject, ..). Het omgekeerde kan natuurlijk ook, het is bijvoorbeeld perfect mogelijk om primaire en secundaire vertraging samen te nemen tot één KPI. Welke scope de KPI's behandelen hangt af van de beschikbare data, bij Darwin (sectie 3.3.3) wordt bijvoorbeeld een enorme lijst van mogelijke oorzaken voor vertragingen en afschaffingen bijgehouden. Hierbij kunnen dus veel meer KPI's betreffende de oorzaak van ongevallen of vertragingen gedefinieerd worden dan in het bovenstaande overzicht.

Uiteindelijk zullen de KPI's voor elke onderneming waarop gefocust wordt licht verschillen. De ene onderneming zal eerder focussen op klanttevredenheid en een andere onderneming zal bijvoorbeeld vooral winstgericht zijn. Daarom zal elke onderneming ook elke KPI een ander gewicht geven in de bepaling van de eindscore. Deze eindscore wordt dan gebruikt om de globale werking van het bedrijf te kwantificeren.

In dit platform zullen voornamelijk dit soort eenduidige KPI's aan bod komen in de vorm van een percentagescore. Deze KPI's worden berekend op basis van opgeslagen data.

3.2. KPI Dashboard

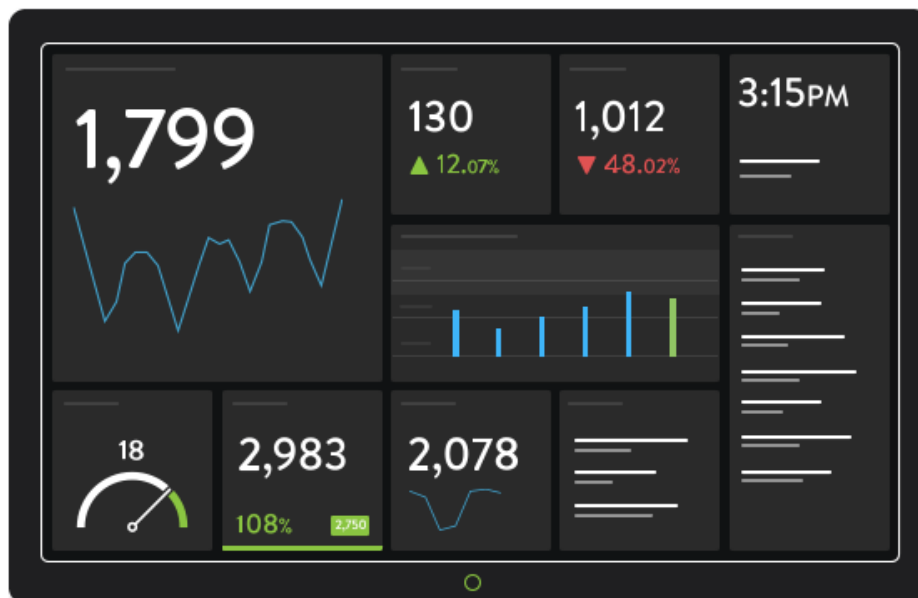
In de eerste plaats wordt gekeken wat precies een dashboard is en daarna worden voorbeelden van reeds bestaande dashboards weergegeven.

3.2.1. Waarom gebruik maken van een dashboard?

Een dashboard is nuttig om de belangrijkste KPI's in één oogopslag te zien. De verschillende KPI's worden op een grafische manier gerepresenteerd met een bepaalde score. Door deze manier van weergegeven is het makkelijk om snel een beeld te krijgen van de werking van de onderneming.

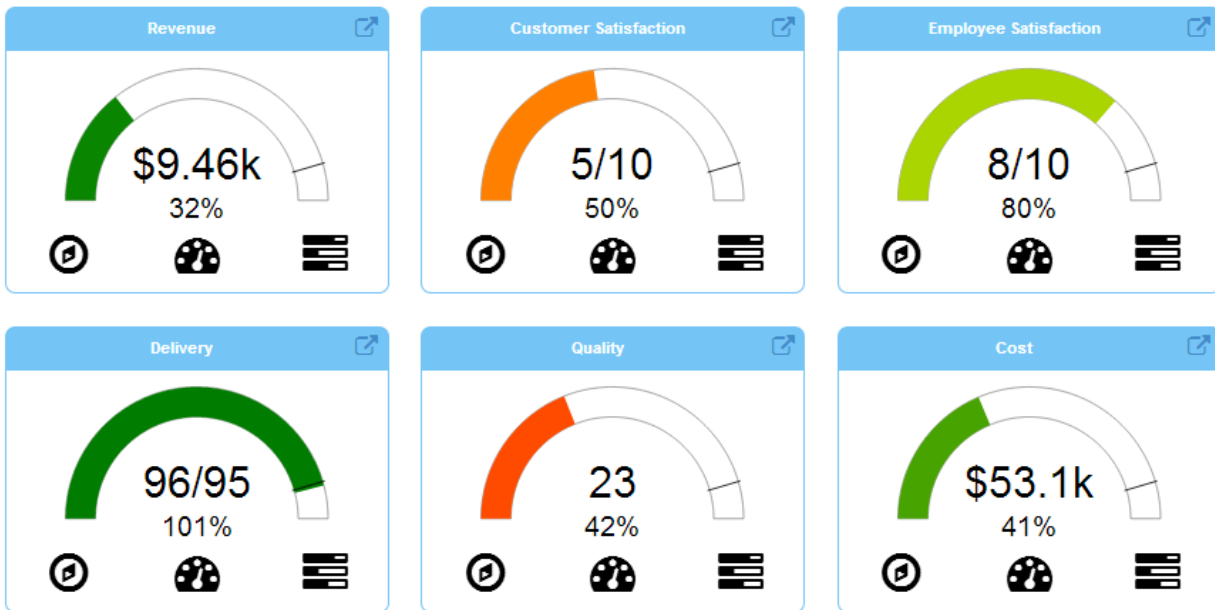
3.2.2. Voorbeelden van dashboards

Eén van de bekendste bedrijven die zich bezighouden met het maken van dergelijke dashboards is Geckoboard [18]. Deze software laat de gebruiker verschillende databronnen (zoals bijvoorbeeld een Twitterfeed) toevoegen om weer te geven op het dashboard. Het is ook mogelijk om bijvoorbeeld Excel spreadsheets en andere datasets toe te voegen. Voor de datasets is er een API beschikbaar die het toevoegen op een eenvoudige manier mogelijk maakt. De verschillende widgets op het dashboard worden aan de hand van een gekozen interval geüpdatet. De evolutie van de waarden en het bereiken van bepaalde doelen wordt ook weergegeven aan de hand van kleurcodes, zoals ook te zien op Figuur 8.



Figuur 8: Schets van een dashboard van Geckoboard.

Een ander voorbeeld is KPI Fire [19], zoals te zien op Figuur 9, hierbij zien we een soortgelijke manier van werken als bij Geckoboard. De verschillende KPI's worden grafisch weergegeven met een bepaalde score en doelscore. Er wordt ook gezorgd voor een extra weergave waar de evolutie van de verschillende KPI's in rekening wordt gebracht. Dit kan bijvoorbeeld gebeuren per kwartaal of per maand.



Figuur 9: Een voorbeeld van een dashboard van KPI Fire.

Andere applicaties zoals Target Dashboard [20] bieden soortgelijke mogelijkheden aan.

3.3. Case Studie: toepassing op het treinverkeer

Bij het bepalen van de KPI's lag de focus op openbaar vervoer in het algemeen. In de casestudies wordt enkel en alleen het treinverkeer behandeld. Om te bepalen welke KPI's geanalyseerd kunnen worden en bijgevolg welke specifieke casestudie uitgewerkt kan worden moet nagegaan worden wat de mogelijkheden zijn van de bijhorende API's. Eerst zal echter een kleine schets gemaakt worden van de werking van de treinondernemingen in België en het Verenigd Koninkrijk, hierbij wordt gestart met een onderzoek naar de verschillende stakeholders bij deze ondernemingen.

3.3.1. Stakeholders bij treinondernemingen

De manier van werken en de aangeboden service is niet alleen van belang voor de operatoren. Elk bedrijf heeft stakeholders. Een stakeholder wordt gedefinieerd als een individu of groep, binnen of buiten het bedrijf, die een aandeel heeft in of invloed heeft op de werking van het bedrijf [21]. In de Cambridge Dictionary wordt een stakeholder gedefinieerd als *“a person or group of people who own a share in a business”* of *“a person such as an employee, customer, or citizen who is involved with an organization, society, etc. and therefore has responsibilities towards it and an interest in its success”* [22]. Deze definities sluiten aan bij de betekenis van een stakeholder in deze literatuurstudie.

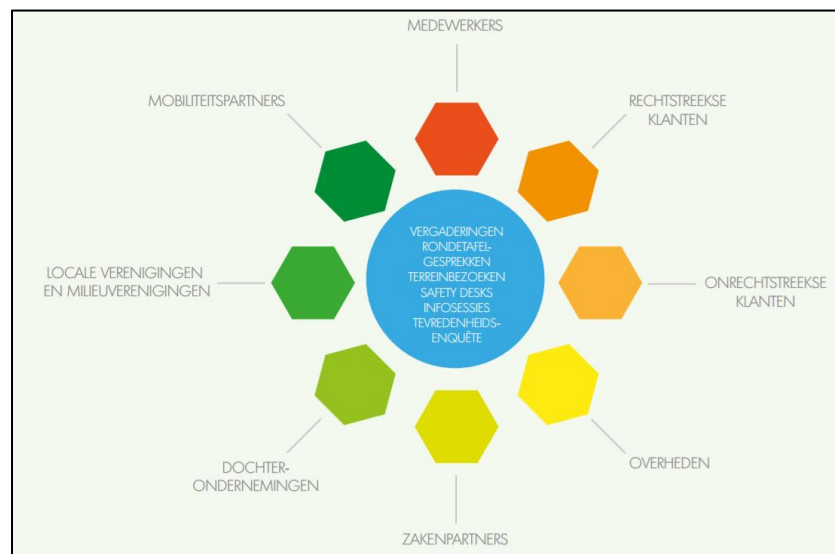
Hoewel alle bedrijven verschillende types stakeholders kunnen hebben, kan men ze voornamelijk opdelen in vijf categorieën [21]: klanten, werknemers, leveranciers (van goederen of services), eigenaars en de gemeenschap. Deze vijf categorieën kunnen ook verdeeld worden over twee grote categorieën. Enerzijds spreekt men over omgevingsstakeholders, deze bevat de klanten, eigenaars en de gemeenschap. Anderzijds spreekt men over processtakeholders, deze bevat de werknemers en de leveranciers. De processtakeholders zorgen ervoor dat er een service wordt geleverd, de omgevingsstakeholders ondervinden invloed van deze service.

Een groot (overkoepelend) spoorbedrijf zoals Infrabel definieert acht categorieën stakeholders waarmee ze op gepaste momenten in dialoog gaan, te zien op Figuur 10 [17]. De medewerkers van Infrabel zijn het personeel. De rechtstreekse klanten zijn de operatoren die gebruik maken van het spoorwegennet (bijvoorbeeld NMBS), de onrechtstreekse klanten zijn de reizigers die gebruik maken van de dienst van NMBS. De overheden kunnen opgesplitst worden in de federale,

regionale of lokale overheid. Leveranciers en ondernemers zijn de zakenpartners van Infrabel. Infrabel heeft ook twee dochterondernemingen. TUC Rail staat in voor grote infrastructuurprojecten. HR Rail is gedeeld met NMBS: die onderneming zorgt voor het personeelsbeheer en is bijgevolg de werkgever van het personeel van NMBS en Infrabel. Onder lokale verenigingen en milieuverenigingen worden onder andere buurtbewoners gerekend en de mobiliteitspartners zijn bijvoorbeeld havens of maatschappijen voor openbaar vervoer [23]–[25].

Op de site van NMBS [26] vermeldt men in verband met stakeholders: “Reizigers- en milieuverenigingen gelinkt aan mobiliteit, volksvertegenwoordigers, overheden en onderzoekscentra... De lijst van NMBS-stakeholders is lang en divers.”

Wanneer we naar het buitenland kijken zien we soortgelijke groepen vermeld worden. Bij Southern Railway [27] heeft men het over reizigers, lokale gemeenschappen, raadsleden, parlementsleden en transportautoriteiten. Ook bij Arriva [28], een Britse vervoersorganisatie, komen volgende stakeholders terug: reizigers, *contract-awarding clients*, werknemers, zakenpartners en de hele gemeenschap. De *contract-awarding clients* maken gebruik van de kennis van Arriva in verband met het vervoer van reizigers.



Figuur 10: Infrabel onderscheidt acht grote categorieën stakeholders. [17]

Bij Infrabel is gebleken dat de voornaamste thema’s waarin stakeholders geïnteresseerd zijn de stiptheid, veiligheid en financiën zijn. Aangezien Infrabel deze thema’s kent kunnen ze er in de toekomst rekening mee houden en bijgevolg de service verbeteren [17].

Stakeholders zijn dus alle mensen of groepen van mensen die op de een of andere manier in contact komen met een bedrijf. Elk bedrijf moet er naar streven om zijn stakeholders tevreden te houden. Om dit te bewerkstelligen kan best op gepaste momenten in discussie gegaan worden met de stakeholders. Het doel van deze bijeenkomsten is om problemen uit te klaren en uiteindelijk de service voor elke stakeholder beter te maken.

Elke stakeholder heeft verschillende belangen in het openbaar vervoer. Het is bijgevolg van belang om te definiëren voor welke stakeholder(s) het dashboard bedoeld is. In dit onderzoek gaan we uit van een operator (zijnde bijvoorbeeld NMBS, niet Infrabel). Er wordt een consensus gezocht tussen de onderverdelingen die teruggevonden werden in de hierboven beschreven bronnen.

Stakeholder	Definitie
Klanten	De reizigers zijn de rechtstreekse en tevens enige klanten van de onderneming.
Medewerkers	De medewerkers zijn alle personeelsleden die taken uitvoeren die vanuit het management gedelegeerd worden. In het treinverkeer zijn dit bijvoorbeeld conducteurs of machinisten.
Management	Het management is de groep mensen die beslissen welke richting het bedrijf in de toekomst uitgaat.
Eigenaars	De eigenaars zijn de aandeelhouders van de onderneming, zij willen vooral dat de onderneming winst maakt.
Zakenpartners	De zakenpartners zijn de bedrijven die overeenkomsten hebben gesloten met de onderneming in kwestie.
Overheden	In de Belgische context moeten zowel NMBS als Infrabel verantwoording afleggen aan de federale overheid over de werking van hun onderneming. Verder moeten ze rekening houden met opgelegde wetten van de bevoegde minister(s) [29].
Gemeenschap	Grote ondernemingen hebben een grote invloed op de gemeenschap, de gemeenschap omvat alle mensen die niet per se gebruik maken van of meewerken aan de diensten van de onderneming. Een typisch voorbeeld hierbij zijn mensen die in de buurt wonen van een spoorweg.

Het doel van het dashboard is dat het management of specifieke medewerkers door gebruik te maken van de applicatie een beter beeld krijgt op de werking van het bedrijf. Het is ook mogelijk om de werking van bedrijven met elkaar vergelijken. Dit kan gebeuren door identieke KPI-bomen (KPI-bomen waarbij identieke KPI's eenzelfde gewicht krijgen) op te stellen en deze in te vullen met de meetwaarden van de verschillende bedrijven. Vanuit dat oogpunt kan het systeem dus ook gebruikt worden door overheden om de werking van de bedrijven te controleren en beter op elkaar af te stemmen.

Dit is in de eerste plaats gericht op stiptheid van de service maar in het dashboard kunnen ook extra scopes gedefinieerd worden zoals klanttevredenheid en kosten. Op deze manier kan vanuit de overheid gecontroleerd worden of de gemaakte kosten gezorgd hebben voor een betere algemene werking van de ondernemingen. Verder kan ook gekeken worden welke ondernemingen efficiënter te werk gaan en deze ondernemingen zouden dan bijvoorbeeld ook meer gesteund kunnen worden.

3.3.2. België

De eerste casestudie zal een deel van het Belgische treinverkeer analyseren. Om een correcte analyse te maken van de manier van werken in België is het in de eerste plaats belangrijk te kijken hoe de verantwoordelijkheden verdeeld zijn. De meest gekende bedrijven die de gewone man linkt met het treinverkeer zijn Infrabel en NMBS. Vroeger was dit één bedrijf maar omwille van de liberalisering van de spoorwegen werd het gesplitst. Het infrastructuurbeheer en de commerciële activiteiten mochten geen deel uitmaken van de verantwoordelijkheden van één onderneming [30].

Infrabel is de beheerder van de Belgische spoorweginfrastructuur. Het bedrijf staat in voor het onderhoud en uitbreiding van het spoorwegennet. Hiernaast beheren ze ook het treinverkeer en factureren ze het gebruik van de sporen. NMBS regelt alles in verband met het reizigersvervoer, gebruik makend van de infrastructuur die aangeboden wordt door Infrabel. Dit gaat van de organisatie van het reizigersvervoer tot en met het verzekeren van de veiligheid op de rijtuigen. Maar hiernaast regelen ze ook een deel van het goederenverkeer in België en beheren ze de stations. Kort samengevat is het dus zo dat Infrabel de infrastructuur verzorgt waar ondernemingen zoals NMBS gebruik van maken om diensten aan te bieden. Er is tevens nog een derde belangrijke onderneming genaamd HR Rail, dit is de werkgever van het personeel van zowel Infrabel als NMBS [31].

Beide bedrijven hebben verschillende doelstellingen omwille van hun verschillende scope. Infrabel heeft vijf grote prioriteiten. In de eerste plaats willen ze de veiligheid in, op en rond de sporen verzekeren. Verder is het ook van belang dat de treinen stipt kunnen rijden en dat er steeds geïnnoveerd wordt om de ontwikkeling van het land te ondersteunen. Als vierde prioriteit willen ze, zoals de meeste bedrijven, financieel gezond zijn. Als laatste willen ze er ook voor zorgen dat ze op één lijn staan met de samenleving [16]. Bij NMBS worden de prioriteiten niet expliciet vermeld, zij willen voornamelijk de klant centraal zetten. Hierbij zorgen ze voor veilige, betrouwbare en efficiënte mobiliteitsoplossingen en een relevante en rendabele service. Ze focussen op drie P's: profit, people en planet. Ze willen financieel gezond zijn, de maatschappij een dienst bewijzen en ook rekening houden met het milieu [32].

Uit de jaarverslagen van NMBS kunnen verschillende relevante cijfers gehaald worden die als input kunnen dienen voor de KPI's. Er is een heel spectrum aan meetwaarden die worden gepresenteerd in deze verslagen. Een aantal voorbeelden hiervan zijn stiptheidspercentage, klanttevredenheid en aantal reizigers. Ook cijfers in verband met inkomsten en uitgaven kunnen hier handmatig uit geëxtraheerd worden. In het jaarverslag van 2014 specificeert NMBS ook een aantal KPI's, deze verdelen ze over de verschillende directies van het bedrijf. In volgende tabel worden deze directies en hun bijhorende KPI's opgesomd [33].

NMBS Transport	Optimale veiligheid Betere Stiptheid Kwaliteit dienstverlening aan de klant
NMBS Marketing & Sales	Reizigersgroei Klanttevredenheid Hogere omzet Multimodale oplossingen
NMBS Stations	Kwaliteit stationsprojecten Klanttevredenheid Toegankelijkheid Multimodale oplossingen
NMBS Technics	Betrouwbaarheid en beschikbaarheid treinen Netheid treinen Kwaliteit modernisering en nieuwe treinen
NMBS Finance	Operationeel resultaat Schuldafbouw

Vertragingen

Een belangrijk gegeven bij vertragingen is het toewijzen welke partij het veroorzaakt heeft. Zoals vermeld in sectie 3.1.3 bij de onderverdeling Verkeersbeheer zijn er verschillende partijen die vertragingen kunnen veroorzaken. Bij Infrabel splitst men dit op in de infrastructuurbeheerder zelf, de spoorwegondernemingen of derden. Deze toewijzing is belangrijk voor het bepalen van de bonus/malus.

Voor de toewijzing van de vertragingen neemt de infrastructuurbeheerder (Infrabel) alleen de storingen in rekening die:

- 20' vertraging of meer veroorzaakt aan een reizigerstrein, door dezelfde aanleiding.
- in totaal 40' vertraging of meer veroorzaakt aan verscheidene reizigerstreinen. Hierbij worden secundaire vertragingen ook in rekening gebracht.
- 60' vertraging of meer veroorzaakt aan een snelle goederentrein;
- 120' vertraging of meer veroorzaakt aan een trage goederentrein;
- 90' vertraging of meer aan verscheidene goederentreinen veroorzaakt;
- oorzaak waren van de volledige of gedeeltelijke afschaffing van een trein of verscheidene reizigers- of goederentreinen.

De maximale vertragingen die in rekening gebracht worden voor het berekenen van de bonus/malus is 60' voor een reizigerstrein en 360' voor een goederentrein. Indien dus een reizigerstrein afgeschaft wordt zal het beschouwd worden als een reizigerstrein met een vertraging van 60'.

Natuurlijk is dit maar een leidraad en kunnen alle toewijzingen betwist worden. De dienst Arbitration staat in voor het onderzoeken van elke betwisting. Het globale maximumbedrag van de prestatieregeling bedraagt 6 miljoen euro. Elk jaar wordt het maximumaandeel per partij in de globale bonus en malus als volgt bepaald:

1. voor de infrastructuurbeheerder, 30 % van het maximale bedrag.
2. na aftrek van het maximaal toewijsbare aan de infrastructuurbeheerder van het maximale bedrag wordt het maximumaandeel voor elke andere partij proportioneel berekend volgens hun procentueel aandeel in het totale aantal gereden treinkilometer tijdens het afgelopen kalenderjaar.

Indien de partij niet voldoet aan de verwachte stiptheid zal hij een malus moeten uitbetalen. Echter als de eigenlijke stiptheid beter is dan verwachte zal hij een bonus krijgen. [34]

iRail

Realtime data kan worden opgehaald aan de hand van iRail, een open-source project onder leiding van Pieter Colpaert [35]. Sinds 2016 is de overheid verplicht om alle gegevens die ze verzamelen in verband met hun taken openbaar te maken naar de buitenwereld [36]. Deze data wordt dus beschikbaar en herbruikbaar voor initiatieven zoals bijvoorbeeld iRail.

Via de beschikbare API van iRail kunnen alle Belgische stations en verbindingen opgehaald worden. Verder kan zowel historische als live informatie betreffende de vertragingen op verbindingen opgehaald worden. iRail houdt voor elke verbinding de vertraging bij, hierbij horend houdt het ook bij of de trein afgelast werd. Bij de vertraging kan per trein ook exact gekeken worden wat de geplande aankomst-/vertrektijd is en wat de werkelijke aankomst-/vertrektijd is. Spoorwijzigingen kunnen ook uit de data gehaald worden en tegenwoordig bestaat ook de mogelijkheid om de graad van bezetting in de trein op te halen.

Voor dit laatste feit wordt geen objectieve meting uitgevoerd maar wel een feedbacksysteem met de treingebruiker. Reizigers op de trein kunnen een score geven (laag, gemiddeld of hoog) over de bezetting en de algoritmes van iRail zullen op die manier de bezetting kunnen voorspellen in de toekomst [37]. Deze data kan gebruikt worden in deze casestudie maar dan wel als subjectieve indicator van de klanttevredenheid met betrekking tot de drukte op de trein.

Naast de data die rechtstreeks verkregen wordt via iRail kan er ook data afgeleid worden. Zo kan het aantal treinen dat per uur door een station of langs een traject passeert geteld worden. Een mogelijk struikelblok hierbij zijn stations waar treinen passeren maar niet stoppen, deze worden niet vermeld. Door de link te leggen tussen ritten en spoorlijnen kan echter wel nagegaan worden hoeveel treinen in deze stations passeren. Via het twitteraccount van NMBS kan bovendien informatie gevonden worden over de oorzaak van de vertragingen.

Een extra meerwaarde kan gevonden worden aan de hand van weerservices. Door gebruik te maken van weerinformatie kan nagekeken worden wat de mogelijke invloed is van het weer op de vertraging. De weerinformatie kan gelinkt worden aan locaties van stations, deze locaties zijn beschikbaar via de API van iRail.

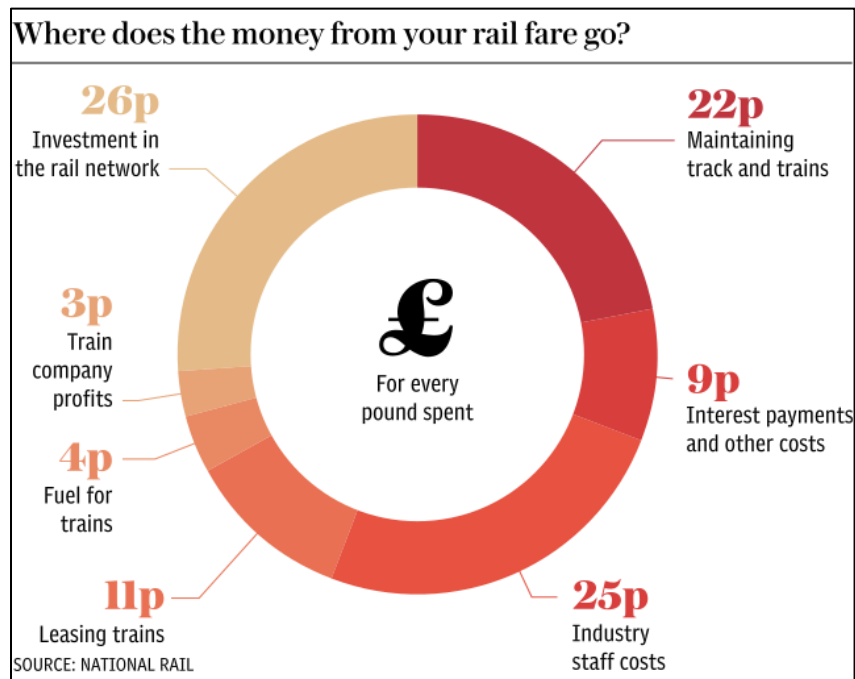
In deze casestudie zullen we onderzoek doen naar de betrouwbaarheid van de service van NMBS. Aan de hand van de gegevens over vertragingen, het twitteraccount van NMBS en een weerservice kan een oorzaak-gevolg structuur opgebouwd worden. Zo zullen de werkpunten duidelijk worden.

Infrabel verdeelt België op in vijf zones waarvoor ze specifieke doelstellingen in verband met stiptheid definiëren [38]. In dit onderzoek zal België als geheel onderzocht worden.

3.3.3. Verenigd Koninkrijk

In tegenstelling tot België zijn er in het Verenigd Koninkrijk een groot aantal verschillende treinoperatoren. In totaal zijn er 28 verschillende operators [39]; elke operator verzorgt de service op een bepaald deel van het treinnetwerk maar deze delen overlappen met elkaar. Op één spoorverbinding kunnen dus treinen van verschillende operators passeren.

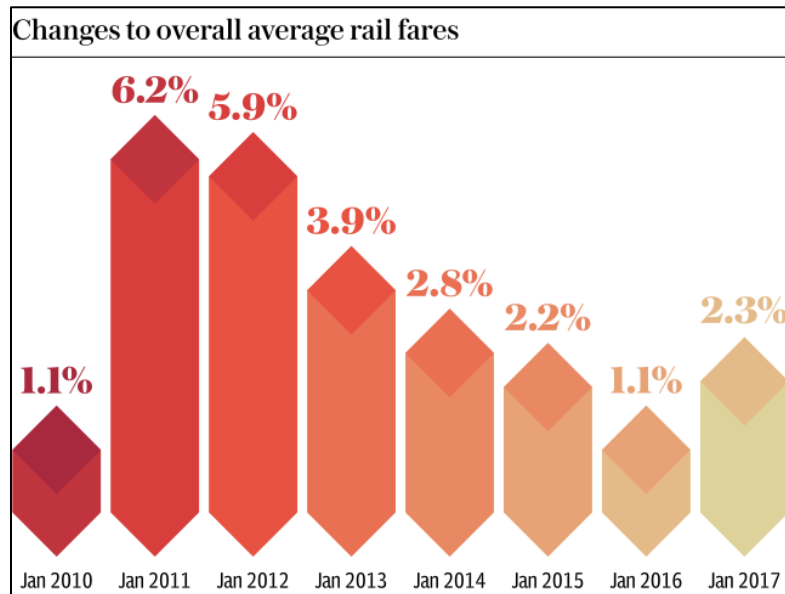
De infrastructuur wordt voor het grootste deel beheerd door Network Rail [40]. Network Rail ontstond in 2002 na het verdwijnen van Railtrack. Het bedrijf is volledig in handen van de overheid. Ze staan in voor het onderhoud en beheer van de infrastructuur van de Britse spoorwegen. Het zou kunnen dat in de toekomst de treinoperatoren zelf het beheer van de infrastructuur zullen verzorgen. Dit biedt een aantal voordelen want momenteel zijn ze volledig afhankelijk van Network Rail met betrekking tot de planning van infrastructuurwerken en dit levert soms conflicten op. Het zou ook problemen in verband met de verantwoordelijkheid van vertragingen en bijhorende terugbetaling aan reizigers moeten vereenvoudigen [41].



Figuur 11 : Waar gaat de ticketprijs naartoe?

Reizigers betalen liefst zo weinig mogelijk voor de service, echter moet iedere betrokken partij wel betaald worden. Op Figuur 11 is te zien dat 3% van de ticketprijs gerekend wordt als winst voor de ondernemingen. De rest van het bedrag wordt gebruikt om de treinservice te voorzien en te verbeteren. Bovenstaande waarden zijn gemiddelde waarden over alle treinoperatoren. Voor

reizigers is dit een interessante grafiek omdat mensen graag weten waar hun geld naartoe gaat. Op Figuur 12 is de jaarlijkse toename in ticketprijs te zien, gezien de prijzen steeds hoger worden willen de reizigers verantwoording hebben bij deze toenames. Hierbij rijst dus de vraag: wordt de service effectief beter?



Figuur 12 : Jaarlijkse toename in ticketprijs

In deze casestudie zal een vergelijking gemaakt worden tussen de werking van verschillende treinoperatoren. Aangezien er net als bij NMBS enkel informatie te vinden is over de vertragingen en niet over de verdere werking van het bedrijf zal de focus weer liggen op de betrouwbaarheid van de geleverde service.

Darwin

Darwin is de officiële treininformatieverzamelaar van Groot-Brittannië. Ze verzamelen de real-time informatie van de verschillende operatoren die gebruik maken van het Britse spoornetwerk [42]. Het systeem zorgt ervoor dat de berichten naar een bepaalde standaard worden omgezet waardoor de data-output voor alle treinoperatoren er identiek uit ziet op vlak van formattering.

In de context van Groot-Brittannië kan het principe van secundaire vertraging duidelijk toegepast worden, er zijn verschillende operatoren op eenzelfde netwerk en ze zullen bij vertragingen invloed ondervinden van elkaar. Het is echter zo dat de API niet specificeert van welke operator de vertraging afkomstig is.

3.4. Ontologieën

In deze thesis zullen ontologieën gebruikt worden om het domein te schetsen. Het is dus noodzakelijk om het concept en nut van ontologieën even toe te lichten. Om te beginnen zal er gedefinieerd worden wat nu precies een ontologie is. Hierna zal het nut van ontologieën verklaard worden en wordt er ook al toepassingsgericht gekeken wat precies mogelijk is.

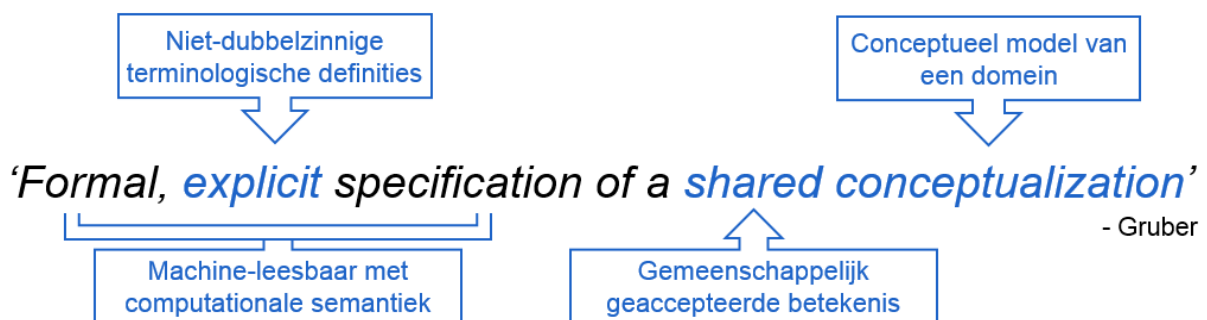
3.4.1. Definitie

Thomas Gruber, een welgekende pionier in dit vakgebied, gebruikte de volgende definitie [43]: “*An ontology is an explicit specification of a shared conceptualization*”. Deze definitie hangt natuurlijk af van hoe men *conceptualization* definieert. In deze definitie van een ontologie wordt *conceptualization* beschreven als een abstract, vereenvoudigd zicht op de wereld dat we willen representeren om een bepaalde reden.

Een ontologie bestaat uit een verzameling van concepten, samen met hun eigenschappen en onderlinge relaties. Het stelt een formeel en conceptueel model voor over een bepaald domein. De context waarin dit model zich bevindt wordt door het domein bepaald [44].

Ontologieën dienen als meta-dataschema's, die een gecontroleerd vocabularium van concepten aanbieden, elk met een expliciet gedefinieerde en machinaal-verwerkbare semantiek.

De Cambridge Dictionary beschrijft *vocabulary* als “*all the words known and used by a particular person*” [22]. Op het Semantic Web (sectie 3.4.3) kunnen *vocabulary* gedefinieerd worden als concepten en relaties die een bepaald domein beschrijven [45].



3.4.2. Waarom worden ontologieën gebruikt?

Ontologieën worden gebruikt om de kloof tussen menselijke en machinale redenering te verkleinen. Mensen gebruiken informatie in een bepaalde context voor een bepaalde gedachtegang. Deze context afleiden is voor artificiële intelligentie geen eenvoudige opgave, met vreemde resultaten als gevolg. Met behulp van ontologieën kan de computer zelf linken leggen en de context eruit afleiden.

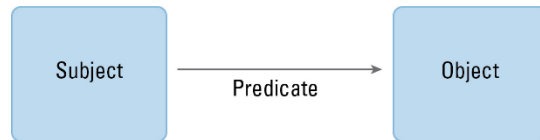
Een ontologie definieert een gemeenschappelijke woordenschat voor experts die informatie willen delen in een bepaald domein. Dit draagt bij tot een gemeenschappelijk inzicht in de structuur van informatie tussen mensen of *software-agents*. Hierdoor kan men makkelijk kennis delen en hergebruiken [46].

Door de opkomst van het Semantic Web kunnen zoekmachines veel gericht zoeken naar resultaten door de context van de query af te leiden in plaats van enkel op woorden te zoeken zonder er een betekenis van af te leiden. Wanneer men bijvoorbeeld “Armstrong eagle moon landing” meegeeft in een zoekmachine verwacht men niet informatie te vinden over Lance Armstrong of een arend. Door het gebruik van ontologieën kan men achterhalen dat men het heeft over Neill Armstrong die landt met zijn spaceshuttle op de maan.

3.4.3. The Semantic Web

The Semantic Web, of Web of Data, is de volgende grote revolutie in het connecteren van informatie. Het kan voorgesteld worden in een graaf, waar de knopen concepten representeren en de lijnen relaties tussen concepten. Op niveau van ontologieën drukken ze een vocabulaire uit die nuttig is voor mensen, maar nog steeds kan gebruikt worden voor *machine processing*. Hierdoor kunnen de computers de informatie interpreteren en connecteren.

Bij de voorstelling van een ontologie wordt er vaak gebruikt gemaakt van het Resource Description Framework (RDF). De onderliggende structuur van een expressie in RDF is een collectie van triples: elk bestaande uit een subject, een predicaat en een object/literal zoals te zien op Figuur 13 [47].



Figuur 13: Graph structure RDF

Een belangrijk bijkomend gegeven is dat RDF gebruik maakt van Uniform Resource Identifier (URI-)referenties om resources en eigenschappen uniek te identificeren.

Web Ontology Language (OWL) breidt dit principe nog verder uit door semantiek toe te voegen aan het schema. Het staat je toe om nog specifiekere eigenschappen en klassen en zo rijke en complexe kennis rond dingen, groepen van dingen en relaties tussen dingen voor te stellen. [48], [49]. Kort samengevat is OWL een vocabulaire uitbreiding van RDF.

Structuur van een OWL ontologie [45], [49]:

Classes: Een klasse is gedefinieerd als een collectie van entiteiten die bij elkaar horen omdat ze gemeenschappelijke eigenschappen hebben. Het wordt soms ook omschreven als een concept. De meest abstracte concepten die gedefinieerd worden in OWL zijn “Thing” en “Nothing”.

- Thing: Thing is de superklasse van alle klassen en de default klasse van alle individuen.
- Nothing: Nothing is een klasse die geen instanties heeft, tenzij bij inconsistenties, en is een subklasse van alle klassen.

Individu: Stelt een instantie van een klasse voor. Dit bevindt zich op het laagste level van objecten, ook wel *ground level* genoemd.

Property: Property is een binaire relatie tussen individuen onderling of tussen een individu en een datawaarde. OWL verdeelt deze onder in twee klassen:

- ObjectProperty is gedefinieerd als een relatie tussen instanties van twee klassen.
- DataProperty is gedefinieerd als een relatie tussen instanties van klassen en *literal values* zoals *string*, *number*, *date*, ...

Door de loop van de jaren zijn er twee versies van OWL uitgebracht. De eerste versie bestaat uit drie taalvarianten geordend op stijgende expressieve kracht: OWL Lite, OWL DL en OWL Full.

OWL Lite wordt vooral gebruikt voor het definiëren van een classificatiehiërarchie en eenvoudige beperkingen. Deze versie was vooral bedoeld om gebruikers leren om te gaan met OWL en een eenvoudige functionele subset aan te bieden.

OWL DL ondersteunt de gebruikers die maximale expressiviteit willen zonder het verlies van *computational completeness* en *decidability* van *reasoning systems*. Deze taal is gebaseerd op beschrijvende logica (*Eng.* Description Logics (DL)).

Een DL gebaseerde ontology kan conceptueel onderverdeeld worden in 3 componenten [50]: de TBox, de ABox en de RBox. Een TBox beschrijft abstracte concepten zoals is-een relaties ($\text{Man} \sqsubseteq \text{Person}$) of equivalentie ($\text{Man} \equiv \text{MaleHuman}$). Een ABox beschrijft concrete instanties en hun relaties tot andere instanties ($\text{hasChild}(\text{John}, \text{Mary})$) en/of abstracte concepten van TBox ($\text{John} : \text{Man}$). Als laatste hebben we een RBox dat relaties beschrijft ($\text{hasSon} \sqsubseteq \text{hasChild}$).

OWL Full is bedoeld voor gebruikers die maximale expressiviteit willen samen met syntactische vrijheid van RDF zonder *computational guarantees*.

OWL werd in 2004 een W3C-aanbeveling voor het modelleren van ontologieën. Ondanks, of zelfs vanwege zijn succes werden er hier en daar toch een paar problemen vastgesteld. Deze problemen werden door OWL 2.0, dat ook een W3C-aanbeveling werd in 2009, beantwoord.

OWL 2.0 breidt de W3C OWL Web Ontology Language uit met kleine maar bevorderlijke features die verzocht werden door de gebruikers. De nieuwe features gaan van extra *syntactic sugar*, *additional property and qualified cardinality constructors*, *extended datatypes support*, *simple meta-modelling* tot *extended annotations* [51].

Ook versie OWL 2.0 heeft verschillende taalvarianten (*Profiles*). Elk van deze taalvarianten is meer gelimiteerd dan OWL DL en biedt verschillende voordelen op vlak van berekeningen en implementatie [52].

OWL 2 (Expressive Language) EL: Dit profiel focust zich op terminologische expressiviteit en wordt gebruikt voor *light-weight* ontologieën.

OWL 2 (Query Language) QL: Deze is in staat om conjunctieve query's in LOGSPACE te beantwoorden gebruikmakende van standaard relationele database technologieën.

OWL 2 (Reasoning Language)

RL: Het OWL 2 RL-profiel is gericht naar applicaties die schaalbare *reasoning* vereisen zonder veel expressiviteit op te offeren.

3.4.4. Reasoning

Reasoning is het afleiden van relaties uit feiten wanneer deze relaties niet expliciet gedefinieerd zijn in de ontologie. Reasoning in ontologieën is één van de redenen waarom een specificatie formeel moet zijn, anders zou het niet verwerkbaar zijn. Het stukje software dat deze taak op zich neemt is de reasoner, zoals de naam al laat blijken. Een paar voorbeelden van taken die een reasoner moet uitvoeren komen aan bod in de volgende paragrafen [53].

Classification

Hierin worden de overervingsrelaties tussen verschillende subklassen berekend om zo een volledige hiërarchie te bekomen.

Query Answering

Na classificatie kan men beginnen met het uitvoeren van query's. Men kan bijvoorbeeld alle instanties van *Person* opvragen. Indien *Bob* een instantie zou zijn van *Man*, zou men verleid zijn om *Bob* niet bij de resultaten weer te geven. Maar *Man* is een subklasse van *Person* dus zou men *Bob* wel moeten teruggeven. Reasoners kunnen gebruikt worden om zo'n query's te beantwoorden en zijn dus niet enkel beperkt tot klassennamen.

Satisfiability and Consistency [54]

Wanneer er een onmogelijkheid in de ontologie zit, spreekt men van een inconsistente ontologie, maar soms verwacht men dit met *satisfiability*. Deze termen zijn makkelijk te onderscheiden aan de hand van een voorbeeld.

```

EukaryoticCell SubClassOf: Cell,
    hasPart some Nucleus

RedBloodCell
    SubClassOf: EukaryoticCell,
        (hasPart only (not Nucleus))

Blood
    SubClassOf: hasPart some RedBloodCell

```

In dit voorbeeld heeft een rode bloedcel een nucleus omdat hij behoort tot de eukaryote cel-klasse, maar er is gedefinieerd dat een rode bloedcel geen nucleus heeft. Momenteel is de ontologie nog steeds consistent, maar heeft *unsatisfiable* klassen. Een ontologie die *unsatisfiable* klassen bevat – zoals in dit voorbeeld – noemt men incoherent. Maar indien we nu een individu willen instantiëren van zo’n klasse, wordt de ontologie wel inconsistent, wat nog iets erger is.

```

Individual: MyRBC
    type: RedBloodCell

```

3.4.5. Semantic Web Rule Language

Semantic Web Rule Language (SWRL) wordt, net zoals de naam al laat blijken, gebruikt als rule language voor het Semantic web [55]. Ze kunnen gebruikt worden om extra feiten te extraheren uit de ontologie die anders niet mogelijk zijn.

Een regel bevat twee delen, een antecedent en een consequent. Indien de voorwaarden in de antecedent voldoen is het consequent ook waar.

Dit simplistisch voorbeeld geeft de regel weer om iemands nonkel te zijn:

```

hasParent(?child, ?parent) ∧ hasBrother(?parent, ?uncle) ⇒
hasUncle(?child, ?uncle)

```

Indien de reasoner SWRL-rule *inferencing* ondersteunt kunnen deze dus gebruikt worden als soort van extensie van OWL.

3.5. Conclusie

Wat zal nu net het nut van ontologieën in dit werk? Allereerst zal het ons domeinmodel schetsen zodanig dat bepaalde begrippen eenduidig zijn. Dit model zal ook gebruikt worden als opslagmodel voor feiten in de databank (DB). Het model zal gebruik maken van OWL DL, enerzijds door de expressiviteit door OWL en anderzijds door de computationele zekerheid door het gebruik van Description Logics lijkt dit de beste keuze. Ook het gebruik van SWRL Rules zal zijn nut terugvinden in dit werk. Dit maakt het mogelijk om extra feiten te extraheren wat anders niet mogelijk zal zijn.

Hoofdstuk 4: Implementatie

Dit hoofdstuk bouwt verder op de technische omschrijving in sectie 2.2. De implementatie bestaat uit verschillende aspecten. Ten eerste worden de gebruikte frameworks en technologieën besproken, hierna volgt de implementatie van de databronnen. Het package diagram wordt daarna besproken met een korte verklaring van de functie van elk package, gevolgd door een aantal secties waarin bepaalde packages nog verder toegelicht worden.

4.1. Frameworks en technologieën

In deze sectie worden de gebruikte frameworks en technologieën vermeld en er wordt kort aangehaald waarom precies deze keuzes gemaakt worden.

OWL API

OWL API¹ is een *high level* Application Programming Interface (API) die de creatie en manipulatie van OWL ontologieën ondersteunt. Sleutelaspecten van OWL API includeren een axioma-centrische abstractie, *first class change support*, *general purpose reasoner interfaces*, *validators* voor verschillende OWL 2 profielen en ondersteuning voor het parsen en serialiseren van ontologieën in verschillende syntaxen [56].

Vanwege de keuze voor een semantische basis, was het gebruik van OWL API voor de hand liggend. Door zijn OWL-gecentreerde implementatie is dit framework meer geschikt dan RDF-gecentreerde frameworks zoals Jena API.

JSON-LD

JSON-LD², of JavaScript Object Notation for Linked Data, is een manier om Linked Data voor te stellen in JSON. Door het immens succes van JSON voor het serialiseren en het versturen van data over een netwerkconnectie wordt deze voorstelling van Linked Data dan ook vaak gebruikt. JSON-LD is ontworpen rond het concept van een “context” om extra mapping van JSON naar een RDF-model te ondersteunen. De context linkt object properties in een JSON-document aan een ontologie. De sleutel-waarde notatie die gebruikt wordt in JSON maakt het ook geschikt om opgeslagen te worden in MongoDB aangezien dat een documentgebaseerd model volgt [57].

¹ <http://owlapi.sourceforge.net/>

² <http://json-ld.org/>

Door intens gebruik van LimeDS, dat vooral JSON gericht is, lijkt JSON-LD de beste keuze als serialisatiesyntax voor RDF. De eenvoudige leesbaarheid van JSON-LD is een extra pluspunt.

LimeDS

Het LimeDS (Lightweight Modular Environment for Data-oriented Services) framework biedt een visuele toolset aan die het mogelijk maakt voor developers om snelle dataworkflows te definiëren met een verscheidenheid aan databronnen. Door zijn built-in support voor *service reliability*, *scalability* en *caching* maakt LimeDS het eenvoudig om data-driven software te ontwerpen. [58] De grote pijlers van LimeDS zijn [59]:

- Snelle en gebruiksvriendelijke HyperText Transfer Protocol (HTTP) web exposure
- Agile Dependency Injection: elke component kan aangesproken worden vanuit een andere component zonder *boilerplate* code te gebruiken
- Scalability: load-balancing and data caching
- Polyglot omgeving: ondersteunt verschillende programmeertalen: Java, Javascript
- Powerful Management User Interface (UI)

Het LimeDS framework is gebouwd op het OSGi (Open Services Gateway initiative) framework. Dit framework is een Java framework dat gebruikt wordt voor en ontwikkelen en deployeren van softwareprogramma's en bibliotheken. Elke bundel is een *tightly coupled*, dynamisch laadbare collectie van classes, jars en configuratie bestanden dat expliciet hun uitwendige afhankelijkheden declareren.

GraphDB

GraphDB (eerder OWLIM) is een schaalbare semantische RDF en graafdatabank. GraphDB is de enige triplestore die semantische redenering op schaal kan uitvoeren [60]. Hierdoor laat het de gebruikers toe om nieuwe semantische feiten te extraheren van bestaande feiten onafhankelijk van de load op de databank. Wat ook de voornaamste reden is waarom deze RDF-triplestore gekozen wordt. SWRL Rules kan deze databank echter niet redeneren, dit wordt met OWL API opgevangen.

Om met de databank te communiceren wordt er gebruik gemaakt van RDF4J³ (vroeger Sesame).

³ <http://rdf4j.org/>

Dit is een open-source Java framework voor het verwerken van RDF-data, gaande van het parsen tot het uitvoeren van query's.

Sparkjava

Sparkjava⁴ is een microframework voor het creëren van webapplicaties in Java 8 en wordt voornamelijk gebruikt voor het ontwikkelen van microservices. Spark is gebouwd rond het concept van Java 8's lambdafilosofie waardoor snelle en eenvoudige ontwikkeling mogelijk wordt. Omwille hiervan is het eenvoudig om microservices op te zetten zonder te veel overhead door configuratie.

JEval

JEval is een bibliotheek die instaat voor het parsen en evalueren van mathematische expressies. De JEval expressieve evaluator ondersteunt op maat gemaakte functies, geneste functies en het gebruik van variabelen. Vanwege het feit dat deze bibliotheek niet steunt op andere bibliotheken kan deze eenvoudig geïntegreerd worden in het LimeDS-framework.

Twitter4j

Twitter4j is een onofficiële Java-bibliotheek voor de Twitter REST API⁵. Deze bibliotheek geeft ons de mogelijkheid om eenvoudig tweets binnen te halen en deze te verwerken. Door de afwezigheid van afhankelijkheden is deze bibliotheek ook eenvoudig te integreren in het LimeDS-framework.

Conclusie

Aangezien alles gebouwd is rond LimeDS is het aan te raden om zoveel mogelijk JSON gebruikende libraries te aan te wenden. Voorbeelden hiervan zijn OWL API en GraphDB, deze bibliotheken ondersteunen zowel JSON als JSON-LD wat verwerken van data in LimeDS eenvoudig maakt. Indien men bibliotheken wilt gebruiken in LimeDS zelf is het ook aangewezen om bibliotheken/bundles te gebruiken die geen afhankelijkheden hebben op andere bibliotheken. Aangezien het LimeDS framework gebouwd op het OSGi framework kan het toevoegen van bibliotheken heel moeizaam verlopen. Er een microservice van maken is een eventuele oplossing hiervoor, net zoals wij gedaan hebben in dit project (zie sectie 4.3.1).

⁴ <http://sparkjava.com/>

⁵ <https://dev.twitter.com/docs>

4.2. Databronnen

Hieronder worden de voornaamste databronnen vermeld waar de ruwe data wordt uit opgehaald. Door gebruik te maken van deze publieke data kunnen er aan de hand van analyses correlaties waargenomen worden. Indien men zelf een databron wil toevoegen aan het platform moet men een segment aanmaken met een link naar de databank om de data op te slaan. Deze data moet opgeslagen worden op een manier die voldoet aan de ontologieën die beschreven worden in sectie 4.4.

iRail

Zoals reeds vermeld in de literatuurstudie biedt iRail data aan van o.a. NMBS. Uit de tweets van NMBS kunnen de oorzaken van vertragingen tussen twee stations geëxtraheerd worden. Uit deze tweets kunnen dus twee stations en een tijdstip gehaald worden. Aan de hand van deze gegevens wordt een aanvraag uitgevoerd bij iRail. Op deze manier kan gekeken worden hoeveel minuten vertraging er effectief is voor een trein op een bepaald traject.

Endpoint

Bij iRail wordt een endpoint voorzien om de stations, gemapt op de ontologie, in de databank op te slaan. Dit endpoint dient één keer aangesproken te worden, bij voorkeur wanneer de databank aangemaakt wordt.

Google API: Geocode

De Geocode API van Google zorgt ervoor dat in deze applicatie aan de hand van de naam van een station de exacte locatie met lengtegraad en breedtegraad kan worden teruggevonden. Deze gegevens zijn nodig om daarna het weer op te halen zoals beschreven in de volgende sectie.

World Weather Online

World Weather Online wordt gebruikt om gegevens over het weer op te halen zoals temperatuur, zichtbaarheid en de naam die ze geven aan de weersituatie. Een aantal voorbeelden van weersituaties zijn: ‘partly cloudy’, ‘sunny’ en ‘light rain’. De API biedt de historische weerdata aan per uur van de dag op een bepaalde plaats. De parameters die gebruikt kunnen worden om deze API aan te spreken zijn een locatie (lengtegraad en breedtegraad) en een datum. Als resultaat wordt dan een JSON-Object teruggegeven met de weersituatie op elk uur van de dag op die plaats.

Segment

In dit segment wordt er tussen 6u 's ochtends en 24u 's avonds om de 3 uur weerdata opgehaald in het station Brussel-Centraal. De bedoeling hiervan is om in te kunnen schatten wat het meest voorkomende weer is in België. Wanneer dit geweten is kan dit vergeleken worden met het weer op momenten wanneer er vertragingen zijn.

Twitter

Door gebruik te maken van Twitter's REST API kan men niet enkel tweets ophalen maar ook andere gegevens zoals aantal retweets en aantal reacties. In de implementatie is een adapter geschreven die de tweets omzet naar het JSON-formaat gebruik makend van de twitter4j-bibliotheek. Hierbij is er keuze tussen een methode die aan de hand van een bepaalde query tweets ophaalt en een methode die de laatste tweets van een gebruiker ophaalt.

Aan de hand van deze adapter wordt NMBS-data opgehaald van twitter en richting databank verzonden. Het gaat hier om twee segmenten die enerzijds aan de hand van hashtags data per lijn ophalen en anderzijds om de reacties van NMBS op klachten.

Bij het zoeken aan de hand van hashtags is er geen officiële limiet op het aantal tweets dat je kan terugkrijgen. Het blijkt echter dat de limiet ligt op ongeveer tien dagen terug in de tijd bij het zoeken via hashtags.

A screenshot of a tweet from the account #NMBS. The text of the tweet is: "#L58 #GentSintPieters #GentDampoort Storing aan de seininrichting Situatie hersteld Normaal verkeer herneemt #NMBS". The hashtags are in blue, and the text is in black. The background of the tweet is white with a light blue border.

#L58 #GentSintPieters #GentDampoort
Storing aan de seininrichting
Situatie hersteld
Normaal verkeer herneemt
#NMBS

Figuur 14 : Voorbeeld van een tweet van NMBS.

Segment

Er werd voor de tweets van NMBS een segment gecreëerd waarin dagelijks tweets worden opgehaald van alle lijnen van NMBS, van zo'n tweet is een voorbeeld te zien op Figuur 14. Aan de

hand van deze tweets wordt ook data van iRail en data in verband met het weer via Geocode en World Weather Online opgehaald.

Uit de tweets van NMBS worden de namen van de stations afgeleid, dit zijn echter hashtags en voor sommige stations (bv. #GentSintPieters) is dit niet ideaal. Deze problemen worden deels omzeild door tussen de verschillende delen streepjes te plaatsen (bv. Gent-Sint-Pieters), maar er zijn stations waarbij dit niet werkt (bv. #SGravenbrakel). Deze stationsnamen moeten doorgegeven worden aan iRail om data op te halen maar dit lukt door bovenstaand probleem niet altijd. Een mogelijke oplossing hiervoor zou zijn om een Java-bibliotheek die gelijkenis tussen twee woorden kan detecteren te gebruiken. De id's van de stations gemapt op de naam van de stations worden via iRail opgehaald en opgeslagen in een .csv bestand. Op deze manier kunnen de id's van de stations gebruikt worden binnen het Twittersegment om data te pushen naar de databank. Het zou beter zijn indien deze id's uit de databank gehaald worden in plaats van uit een bestand. Vanwege het feit dat deze query's extra tijd in beslag zouden nemen werd dit voorlopig echter niet zo geïmplementeerd.

Darwin

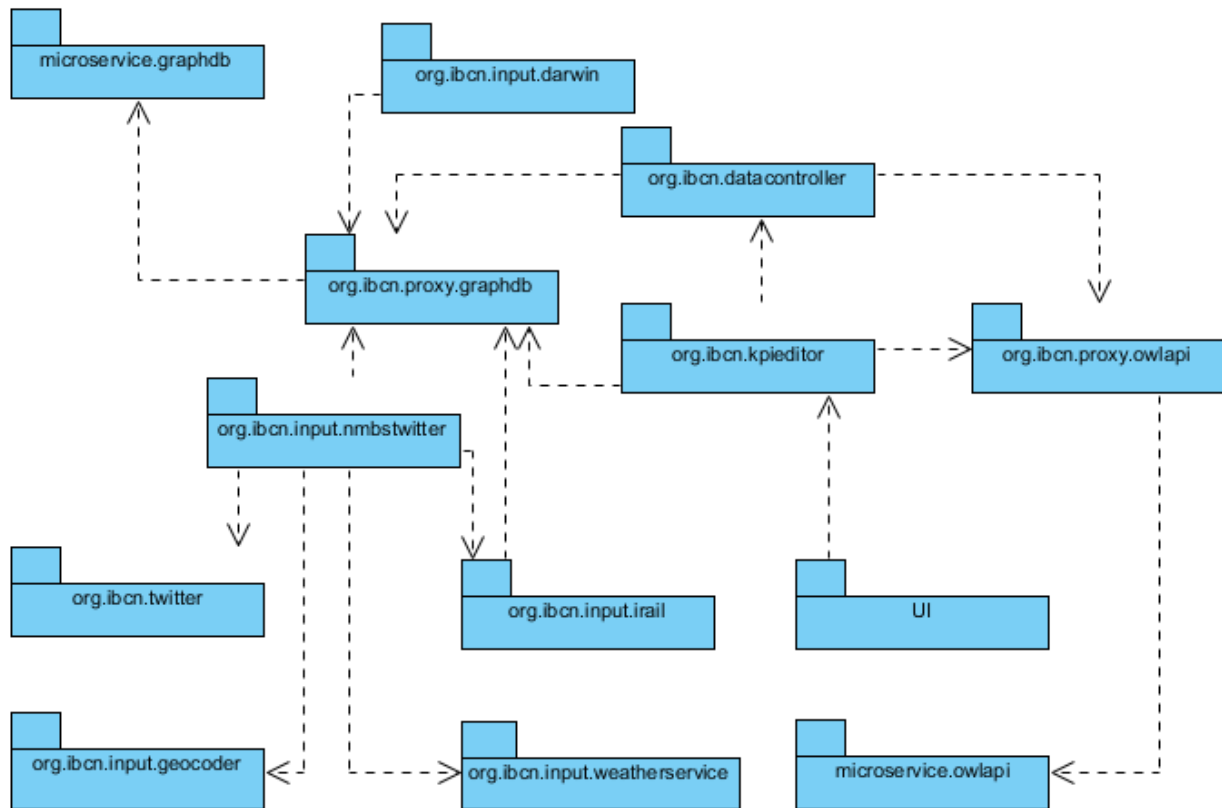
Darwin is de gecentraliseerde databron voor alle informatie van de Britse spoorwegen. Ze biedt een aantal verschillende datafeeds aan, één daarvan is Historic Service Performance (HSP). Bij deze datafeed wordt het resultaat teruggegeven in JSON-formaat; wat in LimeDS een groot pluspunt is. Via HSP kan enerzijds opgehaald worden welke treinen rijden tussen twee stations op een bepaald moment en anderzijds wat de oorzaak is en hoeveel de vertraging bedraagt per station.

Endpoint

Er zijn twee endpoints gecreëerd voor Darwin. In het eerste endpoint (DarwinEndpoint) wordt periodiek data opgehaald via HSP van Darwin. De klasse is zo geïmplementeerd dat er elk uur een aanvraag wordt verstuurd naar Darwin om alle data in de periode van het voorbije uur op te halen. Dit wordt zo geïmplementeerd opdat het platform niet constant aanvragen moet versturen.

Het tweede endpoint moet één keer aangesproken worden, bij voorkeur wanneer de databank aangemaakt wordt. Hierin worden alle Darwinstations gemapt op de ontologie en in de databank opgeslagen.

4.3. Package diagram



Figuur 15 : Overzicht van de packages binnen het platform.

In dit onderdeel zullen alle packages besproken worden die terug te vinden zijn in het overzicht op Figuur 15. Eerst zullen de microservices overlopen worden, daarna zullen ook de gecreëerde packages binnen het Lime DS-framework behandeld worden. Eerst zal de naam van het package vermeld staan gevolgd door een korte beschrijving van de functie.

4.3.1. Microservices

Microservices is een variant op de service-oriented architecture (SOA) stijl dat applicaties onderverdeeld in een collectie *loosely coupled services*. Voor communicatie tussen microservices wordt er vaak gebruik gemaakt van *HTTP resource API*.

Door problemen met afhankelijkheden werd gekozen om bepaalde functionaliteiten buiten het LimeDS-framework te implementeren als een microservice, elk met een REST-service dat dient als communicatiekanaal tussen andere microservices. Een bijkomend voordeel door het gebruik van microservices is dat deze eenvoudig, door hun *loosely coupled* principe, kunnen vervangen worden.

microservice.owlapi: Met behulp van OWL API is het mogelijk om semantische query's (Simple Protocol And RDF Query Language Description Language (SPARQL-DL)) en regels (SWRL Rules) uit te voeren. De verschillende operaties die deze micro-service ondersteunt staan vermeld in onderstaande tabel.

Methode	endpoint	Params	Body	Accept header	Content-type
DELETE	/destroy	instance	Undef	Undef	Undef
DELETE	/destroy/all	Undef	Undef	Undef	Undef
GET	/sparql	query, instance	Undef	Undef	Undef
GET	/create	instance	Undef	Undef	Undef
POST	/add/knowledge	instance	JSON-LD	Undef	Undef

microservice.graphdb: Alle code omtrent het initialiseren van de databank, Create-Read-Update-Delete (CRUD-)operaties en query's bevindt zich hier. De verschillende operaties die deze micro-service ondersteunt staan vermeld in onderstaande tabel.

Methode	endpoint	Params	Body	Accept header	Content-type
GET	/query/construct	query	Undef	Application/ld+json	Undef
GET	/query/construct	query	Undef	Application/turtle	Undef
GET	/query/select	query	Undef	Undef	Undef
POST	/add	Undef	JSON-LD	Undef	Application/ld+json

4.3.2. LimeDS

org.ibcn.input.twitter: Het gebruiken van Twitter als databron in LimeDS. Gebruikt achterliggend de bibliotheek van Twitter4j.

org.ibcn.input.nmbstwitter: Het package NMBSTwitter dient om de tweets van NMBS op te halen, deze tweets semantisch te verwerken en uiteindelijk te pushen naar de database via de GraphDBProxy.

org.ibcn.kpieditor: Het toevoegen en het berekenen van KPI's wordt hier verwezenlijkt.

org.ibcn.input.darwin: Het ophalen, semantisch verwerken van gegevens van de Darwin databron. Eveneens worden deze verwerkte gegevens verzonden naar de databank.

org.ibcn.proxy.owlapi: LimeDS-proxy van de OWL API-microservice. De communicatie tussen beiden gebeurt op basis van REST.

org.ibcn.proxy.graphdb : LimeDS-proxy van de microservice GraphDB-microservice. De communicatie tussen beiden gebeurt op basis van REST.

org.ibcn.input.geocode: Data ophalen van de Google Api: Geocode-databron.

org.ibcn.input.weather: Data ophalen van de World Weather Online-databron.

org.ibcn.datacontroller : Communicatiemiddel tussen database en OWL API via de proxy's. Na het uitvoeren van query's in de databank worden deze gegevens ingeladen in OWL API voor verdere semantische verwerking.

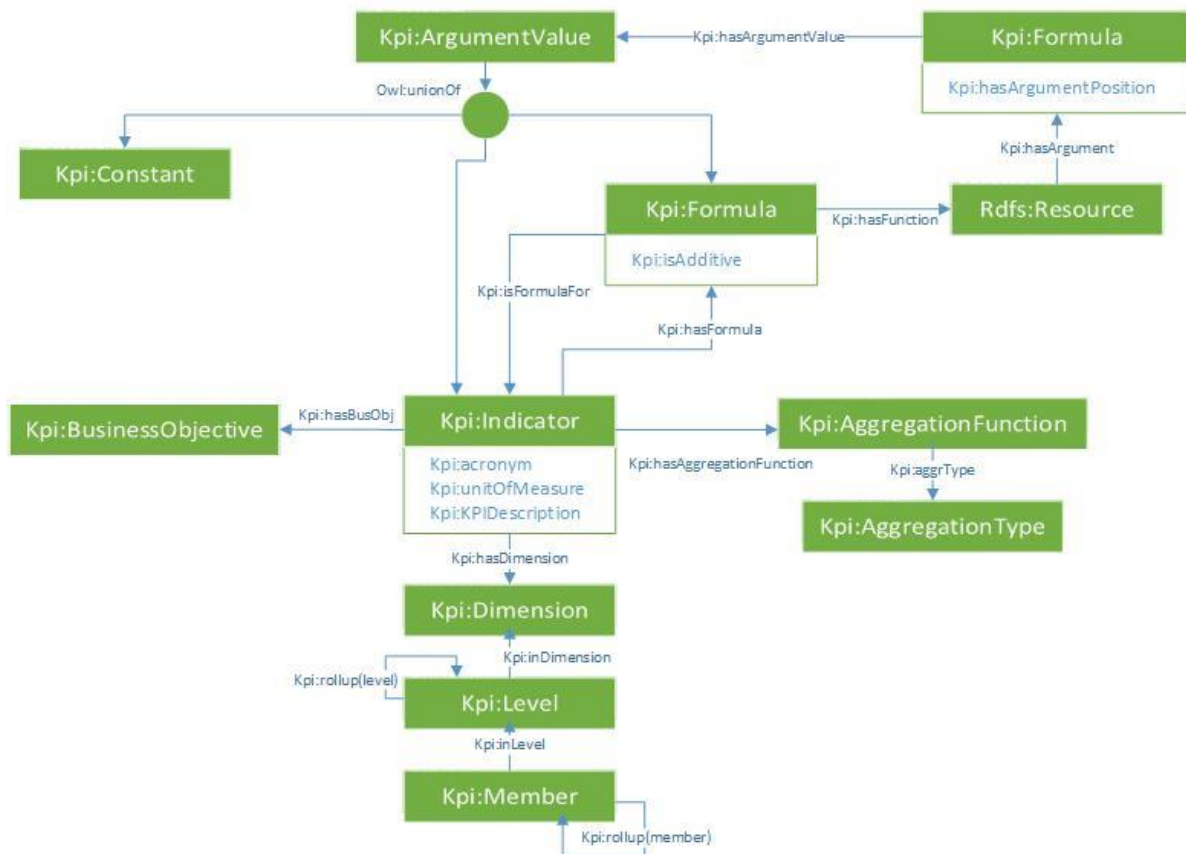
org.ibcn.input.irail: Data ophalen van de iRail-databron.

4.4. Ontologieën

Ontologieën bieden de mogelijkheid om verschillende domeinen te modelleren, maar vindt vooral zijn kracht in de mogelijkheid verder te kunnen redeneren (zie sectie 3.4.4). Een goed startpunt is om alle bestaande ontologieën omtrent de te behandelen domeinen (treinmobiliteit en KPI's) te onderzoeken zodat er een goede basis kan gelegd worden. Hieronder worden twee ontologieën besproken waarop de uiteindelijke ontologie gebaseerd is.

De finale ontologie zal gebruik maken van OWL DL. Enerzijds door de expressiviteit door OWL en anderzijds door de computationele zekerheid van Description Logics lijkt dit de beste keuze.

4.4.1. Key Performance Indicators

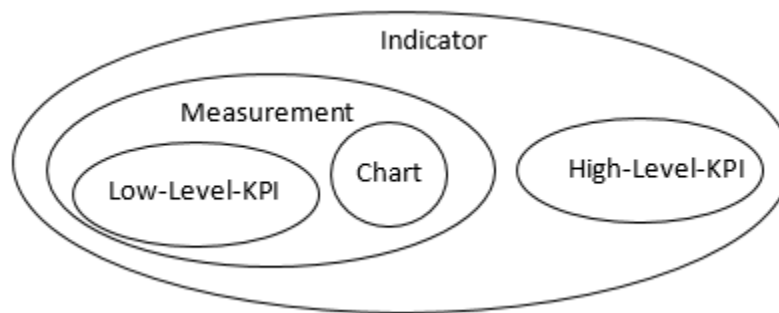


Figuur 16 : KPIOnto

KPIOnto⁶ is een logisch model voor de representatie van indicatoren en hun formules en kan gezien worden op Figuur 16. Deze ontologie gebruikt een multidimensionaal model zoals gekend bij data *warehousing*. De dimensies worden gebruikt om de context van een bepaalde KPI te

⁶ <http://kdmg.dii.univpm.it/kpionto/specification/>

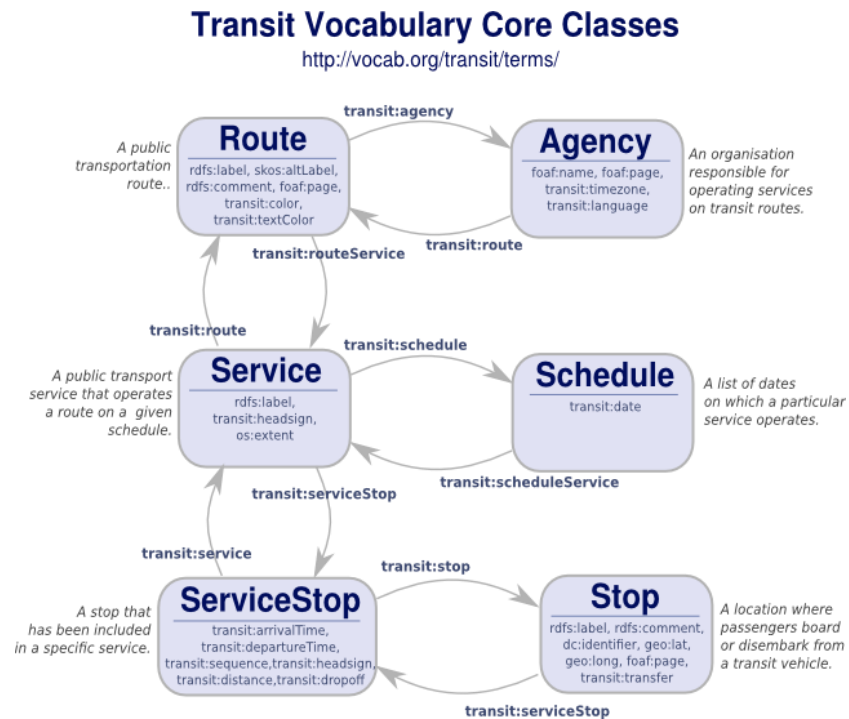
verduidelijken. Verder worden deze dimensies ook nog verder opgedeeld in een hiërarchie van levels. Elk level stelt een andere manier voor om elementen te groeperen binnen dezelfde dimensie. Een KPI kan een atomische KPI (*Low Level KPI*) zijn of opgebouwd zijn uit atomische KPI's (*High Level KPI*), een atomische KPI is een KPI op het laagste level. Afhankelijkheden van een samengestelde KPI zijn gedefinieerd door gebruik van een mathematische expressie, deze wordt voorgesteld in het concept *Formula*. Dit concept is in staat om de semantiek van een indicator uit te drukken door te definiëren hoe de indicator wordt berekend en hoe deze opgebouwd is [14], [61]. Door de manier waarop deze ontologie is opgebouwd kan er handig op geredeneerd worden zodat formulemanipulatie en -consistentie checken mogelijk wordt. Dit valt echter buiten de scope van deze thesis en hier wordt dus ook niet verder op ingegaan.



Figuur 17: Indicatorverzameling.

Voor dit platform is het begrip *Indicator* licht uitgebreid. *Indicator* heeft een extra subklasse genaamd *Measurement* gekregen (zie Figuur 17). *Measurements* zijn atomische indicatoren (in deze applicatie ook *Low Level KPI's* genoemd) of grafieken (*Chart*). Het verschil tussen beide is dat een KPI één waarde voorstelt in tegenstelling tot een *Chart* dat een multidimensionale waarde heeft. De uitbreiding van *Measurement* t.o.v. *Indicator* is dat een *Measurement* een SPARQL en SPARQL-DL-query bevat. Dit heeft zijn nut bij het berekenen van de waarde van KPI's, meer uitleg hieromtrent volgt in sectie 4.7.2.

4.4.2. Railway

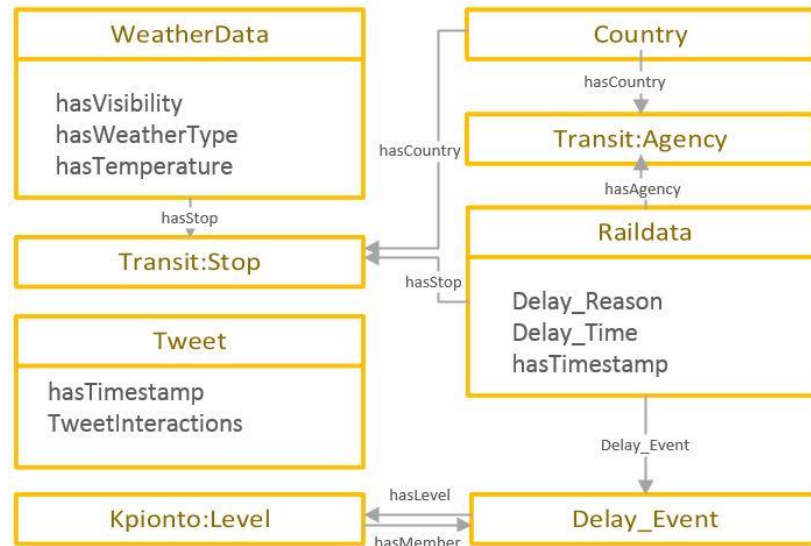


Figuur 18: Het transit schema.

Het Transit schema⁷, te zien op Figuur 18, wordt gebruikt voor het beschrijven van vervoerssystemen, routes, tussenstops en de dienstregelingen. Er zijn veel uitgebreidere ontologieën in dit domein, maar deze ontologie beschrijft de basis op voldoende wijze voor de scope van dit platform, vandaar de keuze voor het gebruiken van deze ontologie.

⁷ <http://vocab.org/transit/>

4.4.3. Twitter Delay Events



Figuur 19 : Ontwerp van de zelf ontwikkelde ontologie.

Op Figuur 19 is de ontologie die we zelf ontwikkeld hebben te zien, de namespace van deze ontologie is: <http://www.semanticweb.org/cedri/syvermee/thesis#>. Aangezien Twitter een belangrijke databron is zal dit ook op een manier moeten geïntegreerd worden in de ontologie. We hebben er voor gekozen om voor dit deel zelf een ontologie te ontwikkelen en proberen te linken aan de andere ontologieën.

Bij het ophalen van tweets wordt een instantie aangemaakt die zowel van het type *RailData* als van het type *Tweet* is. Dit vanwege het feit dat eigenschappen van beide concepten worden opgehaald hierbij. Het is de bedoeling dat een instantie van het type *RailData* altijd minstens een *Delay_Event* of een *Delay_Time* bevat, samen met een timestamp en minstens één station (*Stop*). Idealiter bevat deze instantie ook de vervoersmaatschappij (*Agency*) die de dienst verzorgt. Dit staat niet expliciet gespecificeerd in de ontologie, dit zou in de toekomst beter wel zo gedaan worden.

Het verschil tussen een *Delay_Event* en een *Delay_Reason* is dat het eerste een objecteigenschap is en het andere een dataeigenschap. Toegepast op een voorbeeld zal uit een Tweet de reden van vertraging (bv. “Personen op de sporen”) geëxtraheerd worden, dit is de *Delay_Reason*. In de programmacode wordt dit gemapt op een *Delay_Event*-instantie uit de ontologie, in dit geval:

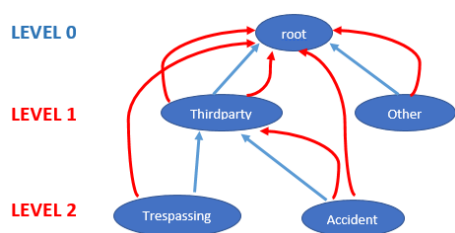
Trespassing. De *Delay_Reason* wordt bijgehouden zodanig dat binnen alle *Trespassing*-instanties een overzicht gemaakt kan worden van de meest voorkomende *Trespassing*-vertragingen.

Wanneer data van Darwin wordt opgehaald zullen enkel instanties van het type *Raildata* gecreëerd worden.

Deze ontologie gebruikt het principe van *Levels* naar analogie met KPIOnto. Men zou kunnen opteren voor het gebruik van subklassen, aangezien een “hogere” *Delay_Event* een collectie is van “lagere”. Maar omdat wij dit niet zien als een abstracter concept en eerder ook als een instantie, ondanks dat het een soort collectie is, gebruiken wij hier *Levels* in plaats van subklassen. Hierbij worden drie SWRL-regels gebruikt die instaan voor het automatisch bepalen van het Level.

SWRL RULES

- `Rollup(?x,?y) -> above(?x,?y)`
- `Rollup(?x,?y) ^ above(?y,?z) -> above(?x,?z)`
- `Delay_Event(?x) ^ Rollup(?y,?x) ^ hasLevel(?x,?lvl) ^ Rollup(?newlvl,?lvl) -> hasLevel(?y,?newlvl)`



Figuur 20 : Delay Event boom

Door de eerste twee regels is het mogelijk om alle voorgangers te achterhalen via de *above*-relatie waar men *rollup* definieert. Een specifiek voorbeeld is het *Delay_Event Trespassing*, bij deze instantie wordt één eigenschap gedefinieerd: *rollup:Thirdparty*.

Bij *Thirdparty* wordt op analoge wijze de eigenschap *rollup:root* gedefinieerd. Door de bovenstaande regels zal geredeneerd worden dat *Trespassing* ook de eigenschappen *above:Thirdparty* en *above:root* bevat.

De laatste regel bepaalt voor elk *Delay_Event* zijn bijhorende Level. De enige voorwaarde hierbij is dat de root-instantie gedefinieerd wordt met eigenschap *hasLevel:Level_0* en dat er voldoende *Level_x* instanties aangemaakt zijn om diep genoeg in de boomstructuur te gaan. Deze *Level_x* instanties hebben één eigenschap: *rollup:Level_(x-1)*, *Level_1* heeft zo de eigenschap *rollup:Level_0*. In dit voorbeeld zal hierdoor *Thirdparty* de eigenschap *hasLevel:Level_1* krijgen vanwege de *rollup:root*-eigenschap in de *Thirdparty*-instantie.

Door enkel al wat blauw is op Figuur 20 te specificëren redeneert de reasoner de overige relaties. Door dit toe te passen kan op een eenvoudige manier snel achterhaald worden wat bijvoorbeeld de Level_1 oorzaak is van een bepaalde vertraging.

4.5. Database

Door de potentieel grote hoeveelheid data, werd gekozen om een database te gebruiken. Het gebruik van databasesystemen biedt ook verschillende voordelen voor de functionaliteit van de applicatie zoals schaalbaarheid, transacties, revisies en *multi-user* functionaliteit.

In deze applicatie wordt gekozen voor een RDF triplestore. Een RDF triplestore is een type graaf-databank waarin semantische feiten worden opgeslagen. Een RDF-triplestore heeft een aantal grote voordelen ten opzichte van een Relationale Databank (RDB). Door de hoge interconnectiviteit van data is de voorstelling in triples de beste keuze, hierdoor ontwijkt men onnodige joins. Bijkomend is er een grote schema flexibiliteit waardoor het toevoegen van nieuwe data eenvoudig is. De query language die gebruikt wordt bij triplestores (SPARQL) is speciaal ontworpen om connectieve data te queryen, dit in tegenstelling tot Structured Query Language (SQL) bij relationele databanken. De meeste RDF-triplestores kunnen ook gequeryed worden over http (Hypertext Transfer Protocol), dus kunnen deze geïntegreerd worden in een service architecture.

Net zoals eerder vermeld wordt er gekozen voor een RDF-triplestore, met name GraphDB.

Voor communicatie tot deze databank maakt het platform achterliggend gebruik van RDF4J.

Bij het opstarten van de microservice wordt een repository geïnitieerd met alle nodige triples van de totale ontologie, indien de gespecificeerde repository nog niet bestaat.

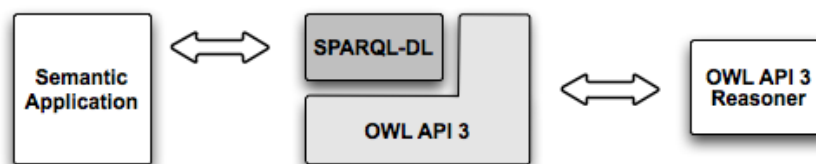
Nadien is de databank klaar om nieuwe data te ontvangen en query's te beantwoorden die binnen komen via de REST-service. Alle aanvragen aan de REST-service worden beantwoord met een JSON-object in de vorm van `{result: String}`, of `{result: JsonObject (JSON-LD)}` indien het een antwoord op een query zou zijn.

Eerder in de implementatie waren alle responses in de vorm van `{result: String}`, omwille van consistentie, maar het parsen van String naar Json met behulp de parse-methode van Json in LimeDS bracht onnodige vertragingen. De reden hiervoor ligt bij de JSON-bibliotheek binnen LimeDS. Om deze reden gebeurt dit nu in de microservice met de parser van Google wat een grote tijds winst met zich meebrengt.

4.6. Semantic Reasoning

Omwille van de grote hoeveelheid data in de databank werd er gekozen om rule-based semantisch redeneren uit te voeren op een kleine subset van de databank. Deze subset bevat data die nodig is bij het berekenen van die bepaalde KPI. Door op een kleine subset van de databank te redeneren worden overbodige bewerkingen vermeden, bijgevolg zal de performantie stijgen. Het rule-based semantisch redeneren is een tijd opslopende bewerking (zie Hoofdstuk 6:Testing) dus elke performantieverbetering is gewenst. Natuurlijk hangt de tijd af van een aantal factoren zoals de hoeveelheid data waarop men redeneert (vandaar de keuze voor het redeneren op een subset) en de complexiteit van de data en query. Voor het semantisch redeneren wordt er gebruik gemaakt van het OWL API-framework en de Hermit Reasoner.

Allereerst moet van deze subset een OWL-ontologie gemaakt worden, hierna kan het verwerkt worden door OWL API. Het is dus niet voldoende om enkel de instanties op te halen, ook de ontologie moet eraan gekoppeld worden. De datacontroller zorgt voor deze koppeling door een import te specificeren van de gebruikte ontologie in het JSON-LD-resultaat van de databank.



Figuur 21 : SPARQL-DL query engine [62].

Nadien kan de subset ingeladen worden in de microservice en kunnen er query's op afgevuurd worden. OWL API voorziet standaard geen methodiek om query's zoals SPARQL te verwerken. Daarom werd er gekozen om de SPARQL-DL-API⁸, een SPARQL-DL engine zoals te zien op Figuur 21, te gebruiken om deze functionaliteit te implementeren.

⁸ <http://derivo.de/en/resources/SPARQL-DL-API/>

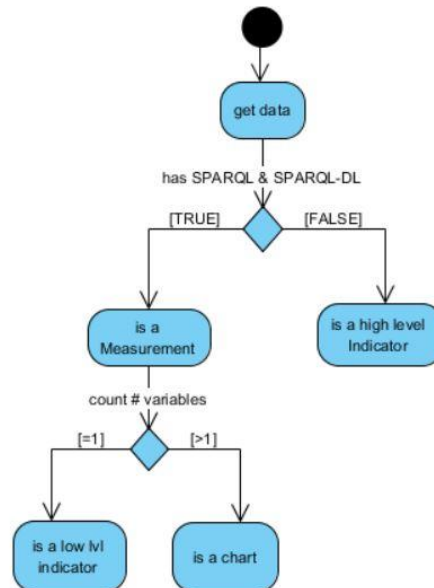
SPARQL is ontworpen voor het uitvoeren van query's op RDF-data, deze querytaal wordt door de meeste RDF-databanken dan ook ondersteunt. SPARQL-DL is een expressievere variant van deze querytaal, het is ontworpen met Description Logics in het achterhoofd. Deze variant is volledig afgestemd op de OWL 2 standard [63]. In het verdere verslag zullen SPARQL-DL query's verwezen worden als semantische query's.

Er werd ook gezocht naar een mogelijkheid om op een eenvoudige manier dynamisch SWRL-rules toe te voegen. Er werd getracht om de SWRLAPI⁹ van Protégé te gebruiken, maar omwille van versieproblemen (OWL 3 t.o.v. OWL 4) was dit niet mogelijk.

Eens de ontologie is ingeladen wordt deze bijgehouden in een sleutel-waarde-structuur (KPI-instantie als sleutel, ontologie als waarde) zodanig dat men niet meerdere keren dezelfde ontologie moet inladen. Natuurlijk moet er ook rekening gehouden worden met nieuwe data. Omwille van deze reden is er een flusher geïmplementeerd, gebruik makend van de functionaliteit van LimeDS, die om de 10 minuten alle KPI-instanties verwijdt zodanig dat men de nieuwe ontologie kan inladen. Een mogelijk probleem met deze aanpak is dat tussen de tijd van controleren of de instantie al bestaat en het uitvoeren van de query's de instantie al verwijderd kan zijn, met `{result: "failed"}` tot gevolg. Dit vanwege het feit dat dit twee aparte aanvragen naar de microservice zijn. Deze raceconditie zou kunnen opgelost worden met een soort van locking-systeem, maar enerzijds door tijdgebrek en anderzijds de kleine kans op deze raceconditie is dit niet verder geïmplementeerd.

⁹ <https://github.com/protegeproject/swrlapi>

4.7. KPIEditor



Figuur 22 : Classificatie KPI

In de KPIEditor wordt de mogelijkheid gegeven KPI's aan te maken. De classificatie gebeurt volgens het schema op Figuur 22. In de editor wordt de optie gegeven om een High Level of een *Measurement* te maken. Bij een *Measurement* wordt de gebruiker gevraagd om een SPARQL en SPARQL-DL query op te geven. Bij deze *Measurements* is er een verschil tussen KPI's met slechts één variabele (één outputwaarde), deze worden geclassificeerd als Low Level KPI's. *Measurements* met meerdere variabelen worden geclassificeerd als Charts. Deze eerste soort vraagt een aggregaatsfunctie als extra parameter en zal bruikbaar zijn in het opbouwen van de boomstructuur door gebruik in High Level KPI's. De soort met meerdere variabelen zal bruikbaar zijn in het dashboard als grafiek. De High Level KPI's worden opgebouwd aan de hand van Low Level KPI's en andere High Level KPI's. Wanneer geopteerd wordt voor een High Level KPI zal de gebruiker een formule moeten definiëren die opgebouwd kan worden uit constanten en KPI's.

4.7.1. Inloggen

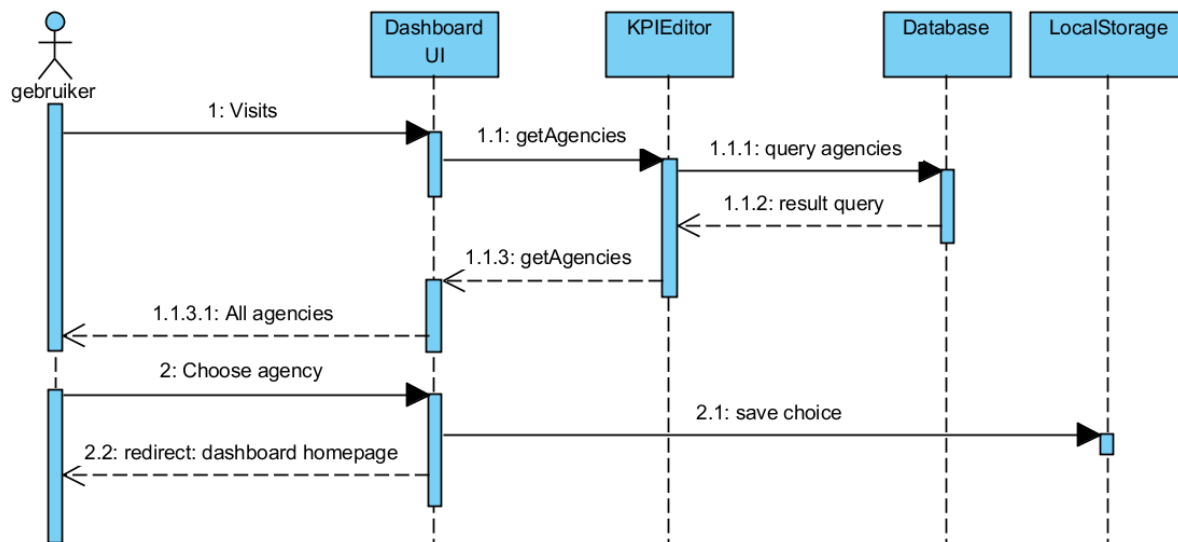
Het eerste wat de gebruiker te zien zal krijgen als hij gebruik maakt van het dashboard is de index pagina waarvan een deel te zien is op Figuur 23. Hier kan de gebruiker een keuze maken om zich aan te melden als een bepaalde organisatie.

- ☐ Agency: NMBS - Country: Belgium
- ☐ Agency: DarwinAgency_SE - Country: UK
- ☐ Agency: DarwinAgency_LM - Country: UK
- ☐ Agency: DarwinAgency_VT - Country: UK
- ☐ Agency: DarwinAgency_SN - Country: UK

BEVESTIG UW KEUZE

Figuur 23 : Voorbeeld van keuzemogelijkheden bij inloggen op het dashboard.

Het sequentiediagram op Figuur 24 geeft weer hoe dit achterliggend in zijn werk gaat.



Figuur 24 : Sequentiediagram in verband met het inloggen.

Deze keuze wordt client-side in Local Storage opgeslagen, zoals af te lezen is. De keuze om Local Storage te gebruiken in plaats van server-side een sessie bij te houden was voor simpliciteit. In een volledige versie van het platform zou er gekozen worden om een sessie te bewaren maar in deze thesis hebben we anders gehandeld.

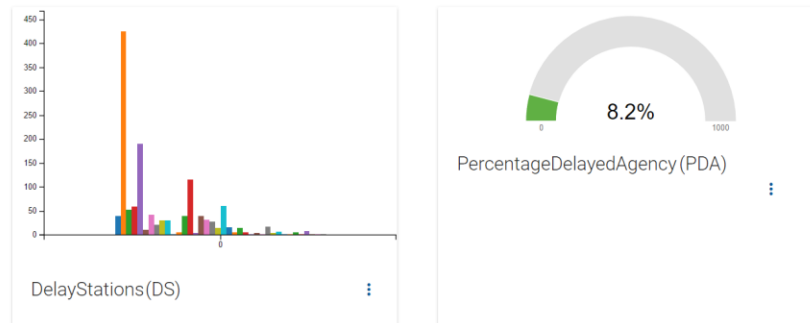
Het inloggen vindt vooral zijn nut terug in gepersonaliseerde SPARQL-query's, waarover later meer (zie sectie Creator).

Indien het inloggen succesvol was wordt de gebruiker doorverwezen naar het overzicht van het dashboard waar het echte werk kan beginnen.

4.7.2. Dashboard

KPI Dashboard

Beleidsimpacterend dashboard op basis van mobiliteitsdata.

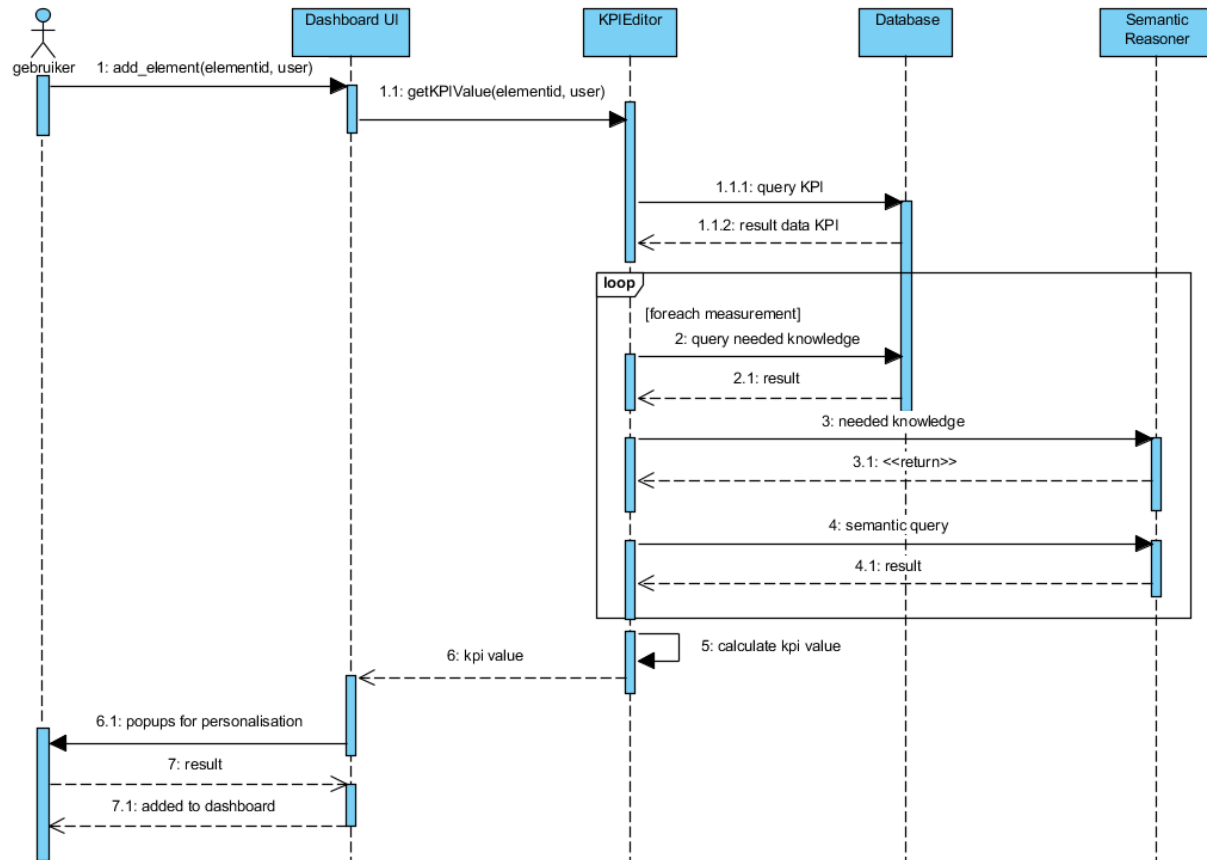


Figuur 25 : Overzicht van het dashboard na inloggen op de website.

De functionaliteit van het dashboard zal in de volgende secties besproken worden, bij het inloggen op het dashboard krijgt de gebruiker een overzicht zoals te zien is op Figuur 25.

KPI/Chart toevoegen aan het dashboard

Eén van de belangrijkste zaken is natuurlijk een reeds gedefinieerd element weergeven op het dashboard. Links bovenaan op de website staat een menu-knop, wanneer hierop geklikt wordt komt een lijst met alle Charts en KPI's tevoorschijn. De gebruiker kan hier bij de gewenste KPI de schakelaar aanklikken. Dit brengt een heel proces teweeg, het proces is geschetst in Figuur 26.

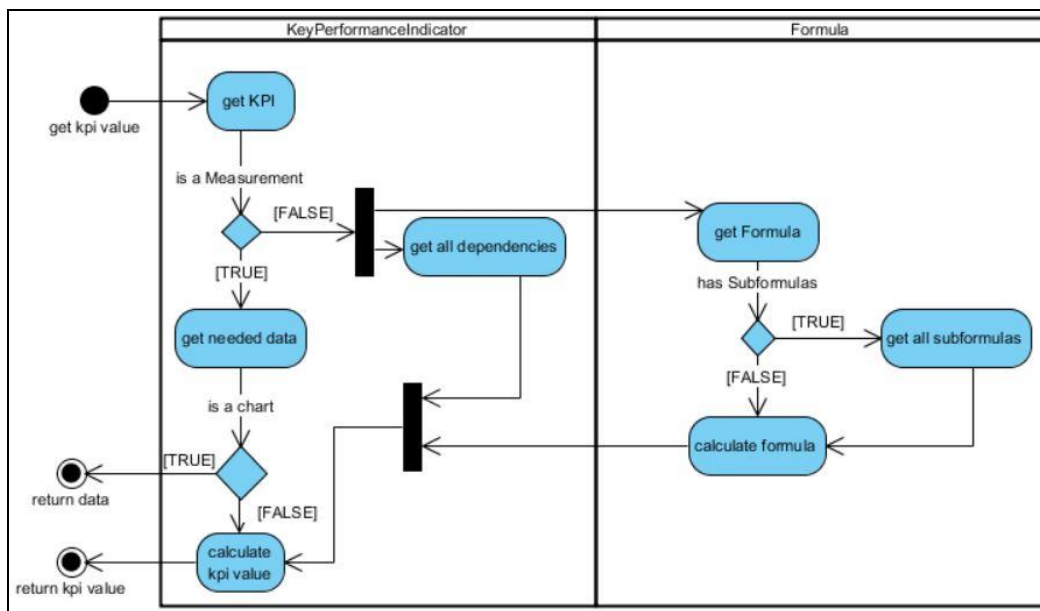


Figuur 26 : Sequentiediagram in verband met het toevoegen element aan dashboard.

Allereerst, na het ontvangen van de aanvraag voor het berekenen van een KPI, zullen alle gegevens omtrent deze KPI uit de databank gehaald worden en de afhankelijkheden op andere KPI's worden recursief, diepte-eerst, opgehaald. Indien zou blijken dat een afhankelijkheid een Measurement is, berekent hij de waarde voor deze afhankelijkheid aan de hand van de gedefinieerde query's (SPARQL en SPARQL-DL). De SPARQL-query wordt gebruikt om de KPI-relevante data uit de databank te halen en als een subset door te sturen naar OWL API. OWL API zal dan semantische query's uitvoeren op deze subset met behulp van de meegegeven SPARQL-DL-query.

Nadat dit proces voor elke afhankelijkheid gebeurd is, kan men hieruit de gevraagde KPI-waarde berekenen.

Berekenen van KPI waarde



Figuur 27: Activiteitendiagram voor het berekenen van een KPI.

Het berekenen van een KPI gebeurt in een aantal stappen zoals getoond wordt op Figuur 27.

De eerste stap voor het berekenen van de waarde is de nodige data ophalen over deze KPI. Uit deze data kunnen we dan de KPI classificeren, dit is belangrijk voor het verdere verloop van het berekenen.

De waarde van een High Level KPI wordt berekend aan de hand van zijn formule. Alle variabelen in deze formule komen overeen met een afhankelijkheid van de bijhorende KPI. Nadat alle waarden van de afhankelijkheden berekend werden en de formule bepaald werd kan de formule uitgerekend worden met behulp van JEval. Op deze manier wordt het resultaat van de KPI bekomen.

Voor een Low Level KPI verloopt dit proces anders, aangezien hij geen afhankelijkheden heeft. De waarde voor dit soort KPI wordt berekend op basis van de semantische query (SPARQL-DL) en geaggregeerd om het finale resultaat te bepalen. SPARQL-DL ondersteunt standaard geen aggregatie, bijgevolg wordt dit na het terugkrijgen van het resultaat pas verwezenlijkt.

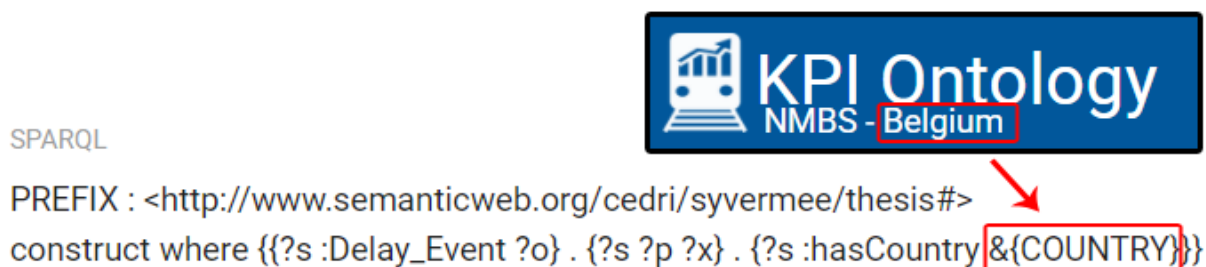
Indien de KPI een Chart is wordt het resultaat van de query teruggegeven en wordt er pas client-side geaggregeerd. Door het resultaat terug te geven krijgt de gebruiker meer vrijheid over hoe hij de data wilt verwerken. Deze vrijheid kan voor een Low Level KPI niet toegelaten worden, een High Level KPI steunt namelijk op het feit dat een Low Level *dependency* één bepaalde waarde teruggeeft.

Bij het gebruiken van het dashboard is geen specifieke kennis van ontologieën of andere informaticabegrippen nodig, het dashboard kan met andere woorden gebruikt worden door het management.

4.7.3. Creator

Op het platform bestaat ook de mogelijkheid om nieuwe KPI's aan te maken. Hierbij kan een keuze gemaakt worden tussen High en Low Level KPI's. De gebruiker van het platform, bij voorkeur iemand met kennis van SPARQL en SPARQL-DL, moet een naam en afkorting definiëren voor de nieuwe KPI. Hiernaast wordt ook de mogelijkheid gegeven om een korte beschrijving van de KPI toe te voegen, deze beschrijving is dan ook zichtbaar op het dashboard.

Wanneer geopteerd wordt voor een Low Level KPI zal de gebruiker een SPARQL en SPARQL-DL query moeten definiëren zoals te zien is op Figuur 29. In de SPARQL-query kunnen country en agency als variabelen meegegeven worden. Dit zorgt ervoor dat de KPI's gebruikt kunnen worden door verschillende ondernemingen, de waarde van de variabele zal namelijk afhangen van de gebruiker. Op Figuur 28 staat een voorbeeld waarin het gebruik wordt, in dit geval wordt `&{COUNTRY}` vervangen door `'Belgium'`.



Figuur 28 : Voorbeeld van gebruik van variabelen in een SPARQL-query.

Wanneer in de SPARQL-DL query slechts één variabele wordt geselecteerd zal ook een aggregaatsfunctie moeten gedefinieerd worden. Hierbij kan de keuze gemaakt worden tussen een

aantal basisfuncties zoals minimum, maximum, som, aantal en gemiddelde van de geleverde dataset. Hierbij moet wel gezorgd worden dat bij functies zoals som en gemiddelde de geselecteerde data bestaat uit getallen en geen tekst. KPI's met één variabele kunnen gebruikt worden in High Level KPI's, wanneer er meer dan één variabele geselecteerd wordt zal de KPI geclassificeerd worden als grafiek (Chart). Verdere configuratie van de grafieken gebeurt op het dashboard.

Aggregatiefunctie

Maak uw keuze ▼

SPARQL

SPARQL-DL

VOEG TOE

Figuur 29 : Opties bij het aanmaken van een Low Level KPI.

Bij de keuze voor een High Level KPI zal in plaats van query's een formule gedefinieerd moeten worden zoals te zien is op Figuur 30. In deze formule kan de gebruiker Low Level KPI's (met één variabele) toevoegen, naast de KPI's kunnen er ook constanten gebruikt worden. De formule accepteert de basisoperatoren: optellen, aftrekken, vermenigvuldigen en delen. Verder is ook het gebruik van haakjes in de formule mogelijk, in de ontologie is het mogelijk om haakjes in haakjes te gebruiken maar op de website is dit voor de eenvoud beperkt tot één paar haakjes per keer.

Formule

+ - * / () 0 1 2 3 4 5 6 7 8 9
 CLEAR

CountThirdparty / CountDelays * 100

CountThirdparty	+
CountDelays	+

Figuur 30 : Opties bij het aanmaken van een High Level KPI.

De gebruiker van het dashboard dient kennis te hebben van SPARQL en SPARQL-DL om nieuwe Measurements te maken. Het maken van een High Level KPI's kan zowel door informatici als het management gebeuren.

4.7.4. Analyzer

In de analyzer bevindt zich een overzicht van alle KPI's en de manier waarop deze met elkaar verbonden zijn. In het voorbeeld op Figuur 31 staat de KPI ProblemDelayAgency die een formule heeft die acht Low Level KPI's bevat.



Figuur 31 : Overzicht van KPI's in Analyzer.

4.8. Conclusie

Achteraf bekeken is de implementatie kon er hier en daar anders gehandeld worden. Een aantal voorbeelden hiervan zullen nu volgen. Het opmaken van de ontologie kon misschien beter, dit werd in een vrij vroeg stadium gecreëerd en met de kennis die we nu gaandeweg hebben vergaard zouden we het model anders geschetst. Dit is vooral van toepassing op de klasse Tweet waarin beter de property hasRailData zou zitten in plaats van hasDelayEvent en dergelijke ter plekke in te vullen in een Tweet-instantie. Verder was er ook het principe van Charts dat in een later stadium werd toegevoegd en daarom niet perfect geïntegreerd is. Een extraatje waaraan gedacht werd is het dynamisch toevoegen van databronnen, vanwege tijdsgebrek werd hier niet verder op ingegaan.

Hoofdstuk 5: Resultaten

Na de implementatie moet de vraag gesteld worden wat er geleerd kan worden via het dashboard, zal het dashboard een impact hebben op het beleid van de onderneming die er van gebruik maakt? Eén van de doelen van deze thesis is onderzoeken of het mogelijk is om pijnpunten van spoorondernemingen bloot te leggen aan de hand van open data.

Om dit te doen zal ten eerste gekeken worden naar de situatie in België: haalt het dashboard dezelfde pijnpunten aan als NMBS en Infrabel? Hierna zal de situatie in het Verenigd Koninkrijk onder de loep genomen worden. Om af te sluiten zal ook een vergelijkende studie uitgevoerd worden tussen België en het Verenigd Koninkrijk aan de hand van gemeenschappelijke KPI's.

Een opmerking voorafgaand aan dit hoofdstuk is dat alle grafieken die zullen volgen rechtstreeks afkomstig zijn uit het dashboard tenzij anders vermeld door middel van een referentie.

5.1. België

5.1.1. Inleiding in verband met de opgehaalde data

De gegevens die in deze thesis opgehaald worden met betrekking tot NMBS zijn afkomstig van Geocode, iRail, Twitter en World Weather Online.

Van het twitteraccount van NMBS worden alle tweets opgehaald die voldoen aan het formaat in Figuur 32. Dit is het standaardformaat dat NMBS momenteel gebruikt op hun twitteraccount om melding te maken van vertragingen. In de meeste gevallen zal er na onderstaande tweet ook een tweet volgen die melding maakt dat het normaal verkeer herneemt op deze lijn. Deze tweets bevatten altijd een lijnnummer (1) en de hashtag NMBS (4), ook de stations (2) waartussen de problemen plaatsvinden worden vermeld. In sommige gevallen staat hier slechts één station in plaats van twee, waarschijnlijk omwille van het feit dat het probleem zich vlakbij dat station situeert. De reden van de problemen (3) wordt ook meegegeven in elke tweet op de tweede regel. Deze tweet wordt elk uur dat het probleem nog bestaat opnieuw de wereld ingestuurd.

1 #L59 #Lokeren #SintNiklaas 2
3 Storing aan een overweg
Verstoord verkeer
4 #NMBS

Figuur 32 : Een voorbeeldtweet van NMBS.

Op basis van deze informatie kunnen tweets van alle lijnen opgehaald worden en de oorzaken van problemen opgeslagen worden. Deze oorzaken worden gemapt op de ontologie die zelf ontwikkeld werd. Deze data wordt opgeslagen samen met het tijdstip van de tweet en de stations die invloed ondervinden van deze problemen. Op ditzelfde moment wordt ook data in verband met de vertraging (in seconden) opgehaald tussen deze twee stations op dit tijdstip via iRail. Om precies te zijn wordt de vertraging in beide richtingen bepaald. Toegepast op Figuur 32 wordt bijvoorbeeld in één richting de vertraging bepaald bij aankomst in Sint-Niklaas bij de trein die het traject tussen Lokeren en Sint-Niklaas heeft afgelegd, hierna wordt ook de vertraging bepaald in de tegengestelde richting. Bij de tweets wordt ook steeds de weerdata opgehaald via Geocode en World Weather Online (zie sectie 4.2) in de stations op het uur van de tweet. Eventueel kan aan de hand van de weerdata ook nog een verband met vertragingen gevonden worden.

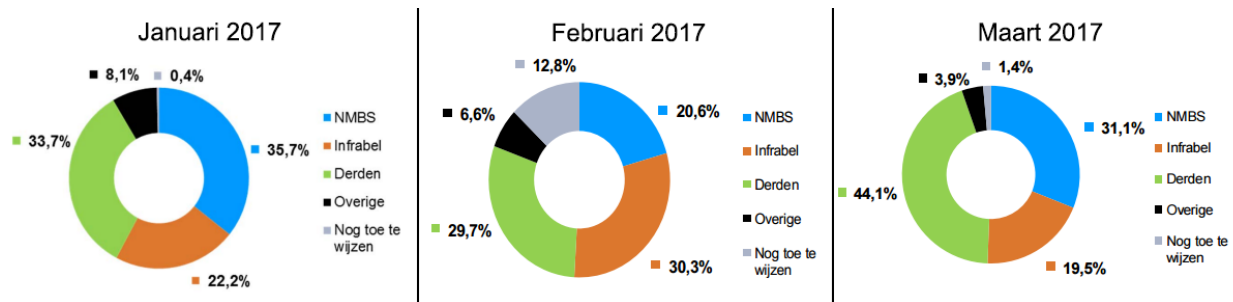
In de volgende paragrafen zal beschreven worden wat geleerd kan worden uit deze gegevens.

5.1.2. Analyse van de resultaten

Meest voorkomende vertragingsoorzaken

Om te beginnen zal gekeken worden welke oorzaak van vertraging er het vaakst voorkomt. Om dit te analyseren wordt data in een bepaalde tijdsperiode opgehaald. Hieruit kan nuttige informatie geëxtraheerd worden zoals welke oorzaak komt het vaakst voor? Verder kan ook nagekeken worden of dit vergelijkbaar is met de data die NMBS vrijgeeft in hun maandelijkse stiptheidsverslagen. In deze verslagen wordt bijgehouden welk percentage van de vertragingen afkomstig zijn van NMBS, Infrabel, derden en overige. Deze percentages kunnen ook afgeleid worden uit het dashboard.

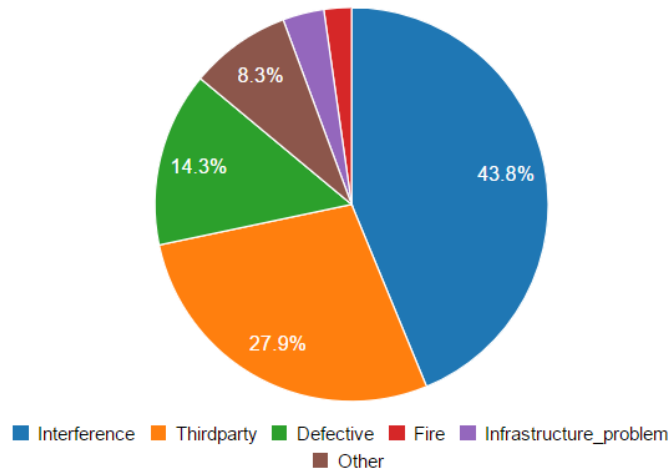
Aangezien de data die in het dashboard gebruikt wordt afkomstig is van de maand mei en het laatste maandverslag in maart werd opgesteld kan er geen exacte vergelijking gemaakt worden.



Figuur 33 : Oorzaak van vertraging in het eerste trimester van het jaar. [64]–[66]

Uit de gegevens op Figuur 33 blijkt dat ongeveer een derde van alle vertragingen veroorzaakt wordt door derden.

Om te kijken of deze gegevens overeenkomen met de tweets van NMBS wordt er gekeken naar de ontologie. Uit alle tweets van NMBS wordt de oorzaak van het probleem gelinkt aan de ontologie. Het resultaat van onze query is te zien op Figuur 34. Op deze manier kan een vergelijkende studie uitgevoerd worden. Ondanks het feit dat de steekproef slechts uit 265 tweets bestaat is het opvallend om te zien dat ongeveer een derde van alle oorzaken te wijten zijn aan derden. Dit komt overeen met de gegevens van NMBS van de voorbije maanden. Verder is het zo dat alle storingen (*Interference*) te maken hebben met de zaken die Infrabel verzorgt (overwegen, signalisering en spoorwissels) net als alles wat te maken heeft met andere infrastructuurproblemen (*Infrastructure_problem*). De som van deze twee bedraagt 47,2% wat beduidend meer is dan de percentages van de maanden januari tot en met maart. De vertragingen ten gevolge van defecten (*Defective*) aan de treinstellen bedraagt 14,3%, daar waar het in de maandverslagen ongeveer 30% is.



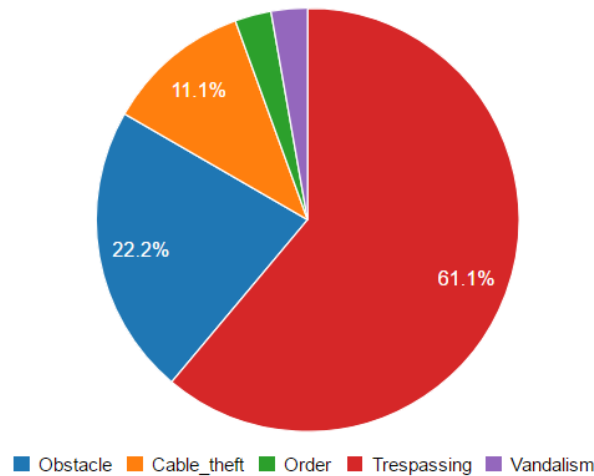
Figuur 34 : Oorzaak van vertraging volgens tweets van NMBS in de periode van 8 tot 17 mei. (aantal tweets: 265)

De tweets van NMBS die gecategoriseerd worden als *Thirdparty* voldoen aan één van volgende voorwaarden:

- De oorzaak bevat de zinsnede: ‘de sporen’; NMBS gebruikt deze zinsnede bij personen, vee en dergelijke meer zoals bijvoorbeeld bij ‘Personen op de sporen’.
- De oorzaak bevat het woord: ‘ongeval’; de NMBS gebruikt dit woord voornamelijk bij ‘Persoonsongeval’.
- De oorzaak bevat het woord: ‘vandalisme’.
- De oorzaak bevat het woord: ‘bevel’; de NMBS gebruikt dit woord bijvoorbeeld bij ‘Bevel politie’.
- De oorzaak bevat het woord: ‘kabeldiefstal’.

Uiteraard kan het zijn dat in de categorie *Other* nog tweets zitten die gecategoriseerd kunnen worden als *Thirdparty*, dit brengt ons naadloos bij één van de problemen bij deze manier van werken. Bij spoorbedrijven wordt niet op een éénduidige manier gecommuniceerd over de oorzaak van vertraging, bijgevolg is het een lastige opgave om een ontologie te creëren waarin alle mogelijke oorzaken te categoriseren vallen. Eventueel zouden de verschillende termen lexicaal vergeleken kunnen worden met elkaar. Indien ze een grote gelijkenis vertonen worden ze dan in dezelfde categorie geclassificeerd.

Het meest opvallende feit is dat een derde van de vertragingen afkomstig is van derden, maar dit is een algemene term. Waar komt deze vertraging nu exact vandaan? Dit kan opgevraagd worden door een niveau dieper te kijken zoals te zien op Figuur 35.



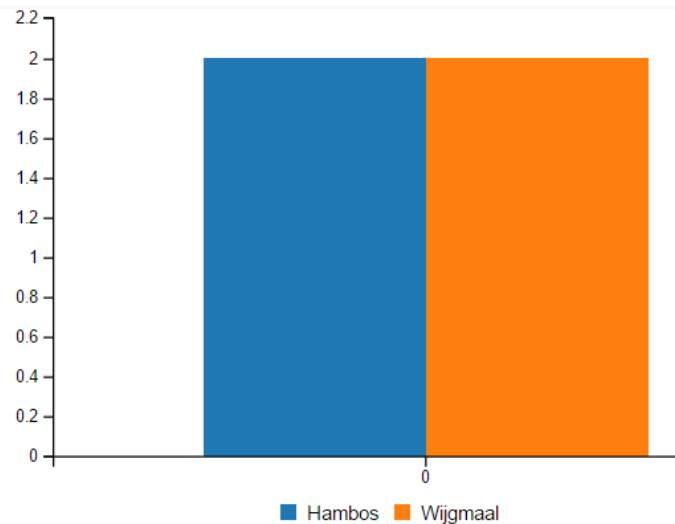
Figuur 35 : Onderverdeling van oorzaken binnen Thirdparty.

Het is meteen duidelijk dat het grootste deel van de vertraging door derden (*Thirdparty*) afkomstig is van personen op de sporen (*Trespassing*). Voor alle duidelijkheid omvat dit geen persoonsongevallen, dit is nog een aparte categorie waar in deze periode geen melding van is geweest. Beleidsmakers kunnen op dit resultaat inspelen. Zo is Infrabel al een paar jaar bezig met het installeren van struikelmatten op de Belgische sporen [67].

Oorzaken in de buurt van bepaalde stations

Uiteraard kan er ook op nog een andere manier omgesprongen worden met deze data, bijvoorbeeld: in de buurt van welke stations komen de meeste vertragingen voor? Wanneer dit bijvoorbeeld bij 'Fire' bekeken wordt kan gezien worden op Figuur 36 dat er een brand is geweest in de buurt van Wijnmaal. Uit verder onderzoek kan ontleed worden dat *Fire* komt van de term 'Bermbrand' die in de tweet werd gebruikt en tevens in de databank opgeslagen werd. Dit is een interessant gegeven gezien in 2011 al eens een dergelijke melding was, toen werd dit veroorzaakt

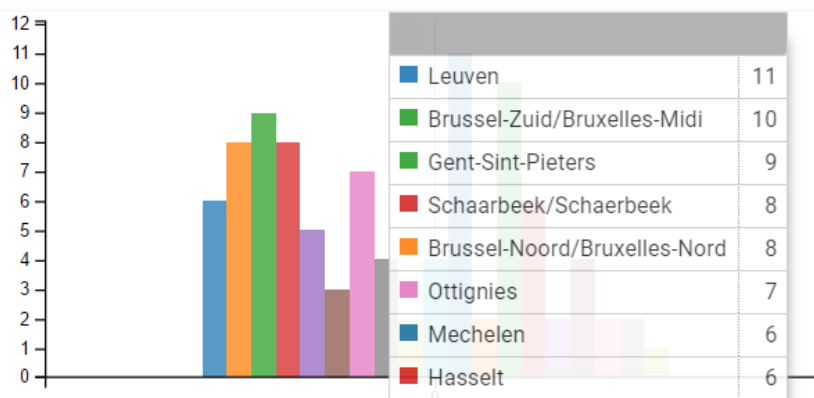
door vonken van een voorbijrijdende trein [68]. Hieruit kan dus blijken dat dit een probleem is dat zich herhaalt en dat misschien wel verder onderzoek vereist.



Figuur 36 : Stations waarbij Fire als oorzaak van vertraging wordt vermeld.

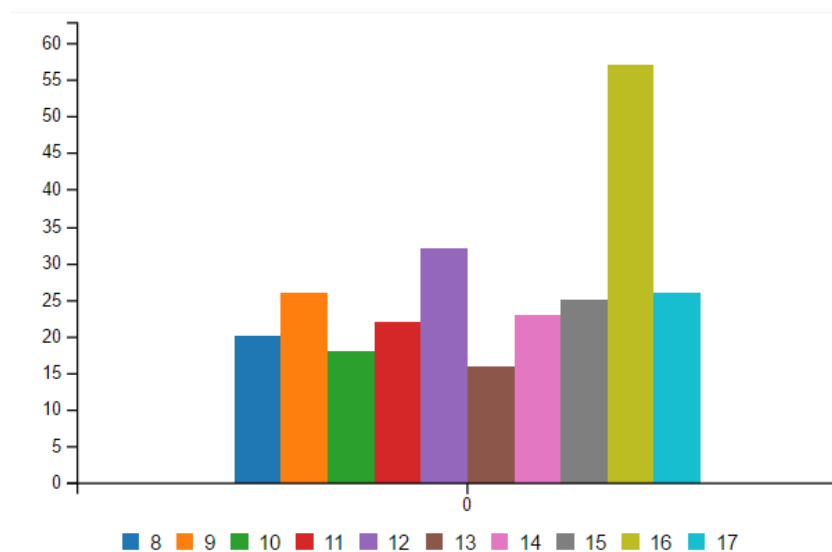
Vertragingen in de drukste stations

iRail biedt ook de gegevens aan van het aantal treinen die per dag een stop maken in elk station, op deze manier kunnen de drukstbezochte stations gevonden worden. Interessant is om te kijken naar de stations met bijvoorbeeld meer dan 100 stops per dag en hierbij te kijken naar het aantal vertragingen. Op Figuur 37 kan gezien worden dat de problemen zich concentreren in Leuven, Brussel-Zuid en Gent-Sint-Pieters. Opvallend is de afwezigheid van grote stations als Antwerpen-Berchem en Antwerpen-Centraal. Een reden hiervan kan zijn dat Antwerpen-Centraal (op twee sporen na) dient doet als een kopstation, bijgevolg kan er slechts vanuit één richting vertraging gecreëerd worden. Dit is in tegenstelling tot bijvoorbeeld Brussel-Zuid of Gent-Sint-Pieters waar er veel treinen passeren.

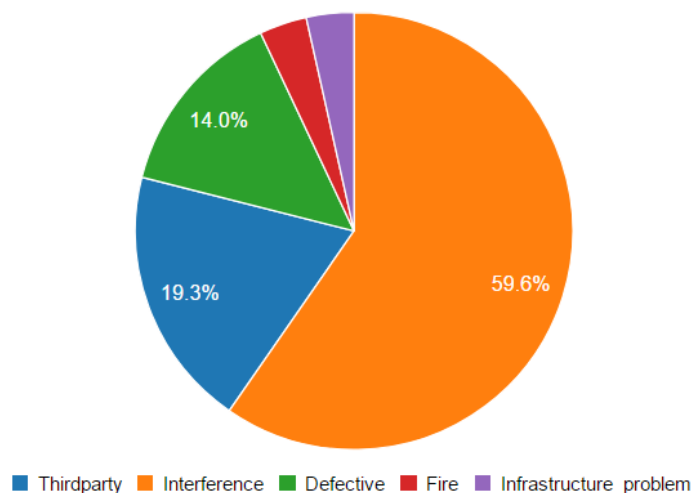


Figuur 37 : Het aantal vertragingen in de belangrijkste knooppunten (>100 stops per dag).

Een verdere analyse kan ook gebeuren door te kijken naar Figuur 38 waarbij de hoeveelheid vertragingen per dag getoond wordt. Op deze manier kan gevonden worden dat binnen de gebruikte dataset er een piek aan vertragingen was op 16 mei. Uit verder onderzoek, te zien op Figuur 39, blijkt dat deze dag licht afwijkt van het gemiddelde op vlak van soort vertragingen. Er is voornamelijk sprake van vertraging door *Interference*, dit zijn dus alle types storingen zoals bijvoorbeeld storing aan de seininrichting. Dit is een veel groter percentage dan op een gemiddelde dag dus zou het nuttig kunnen zijn om hier als beleidsmaker verder onderzoek naar te doen.



Figuur 38 : Aantal vertragingen per dag (van 8 tot en met 17 mei).



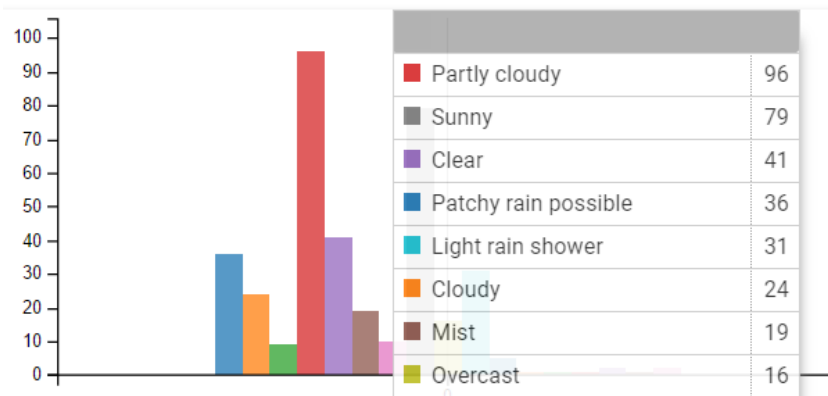
Figuur 39 : Oorzaak van vertraging op 16 mei.

Verband tussen tijd en oorzaak

Aangezien bij elke tweet ook data van iRail opgehaald wordt kan gekeken worden welke type oorzaken de meeste vertragingstijd genereren. Van alle tweets bevat het merendeel twee stations, bijgevolg kan er data van opgehaald worden. Er is echter gebleken dat iRail de historische data niet meer opslaat en zodoende kan er geen resultaat van getoond worden. Het zou bijgevolg handig zijn indien NMBS deze historische data zelf zou willen aanbieden in de toekomst.

Verband tussen vertraging en het weer

Aangezien bij elke tweet ook weerdata wordt opgehaald van de stations kan een verband gezocht worden tussen weersomstandigheden en vertragingen. Een opmerking hierbij is echter wel dat er geen rekening gehouden wordt met welke weersomstandigheid het vaakst voorkomt op deze locaties. Zo zou het bijvoorbeeld kunnen zijn dat in België 50% van de tijd het weer omschreven wordt als ‘Partly cloudy’ en slechts 10% als ‘Light rain’.

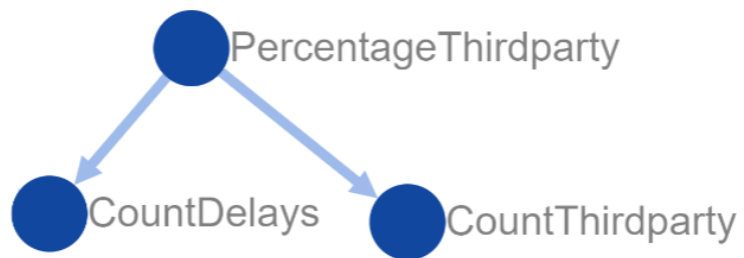


Figuur 40 : Het aantal vertragingen per weersomstandigheid.

Er is geen verband te vinden tussen weerdata en vertragingen op Figuur 40. Wanneer gekeken wordt naar het meest voorkomende weertype in de periode van de metingen, dan staan ‘Sunny’ en ‘Partly cloudy’ ook aan kop. Er kan niet meteen een verband gevonden worden tussen de vertragingen en het weer. Uiteraard is het mogelijk dat binnen een andere dataset – bijvoorbeeld in een periode met sneeuw en ijsel – hier wel een verband kan gevonden worden.

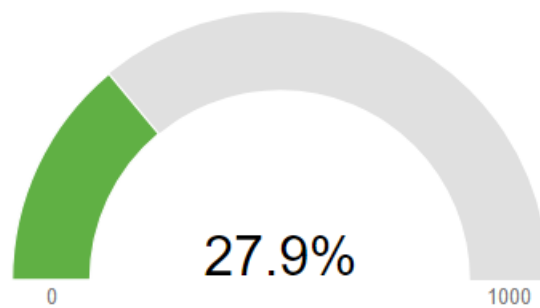
Score van de onderneming

Intussen zijn er al een aantal voorbeelden gepasseerd van grafieken die aan de hand van de opgehaalde data gegenereerd kunnen worden. Dit onderzoek baseert zich echter niet alleen op grafieken, het doel is om een KPI-boom op te bouwen. Een eerste eenvoudig voorbeeld hiervan is te zien op Figuur 41.



Figuur 41 : Eenvoudig voorbeeld van een KPI boom.

In *PercentageThirdparty* wordt een formule gecreëerd waarbij het aantal vertragingen door derden (*CountThirdparty*) gedeeld wordt door het aantal vertragingen in totaal (*CountDelays*). Deze boom bestaat uit twee Low Level KPI's en één High Level KPI (*PercentageThirdparty*). In *CountDelays* en *CountThirdparty* worden query's uitgevoerd om het aantal totale vertragingen en het aantal vertragingen door derden te tellen.



Figuur 42 : Het resultaat van de KPI *PercentageThirdparty*.

Het zou interessant zijn indien ondernemingen een bepaalde score zouden krijgen afhankelijk van de oorzaken van hun vertragingen. De manier waarop hiervoor te werk gegaan wordt is als volgt.

Indien van alle oorzaken het percentage berekend wordt zoals bij *PercentageThirdParty* op Figuur 42, dan kunnen al deze percentages gebruikt worden in een nieuwe KPI. In dit voorbeeld wordt deze KPI *ProblemDelay* genoemd. Deze KPI wordt berekend door de som te nemen van alles percentages, elk vermenigvuldigd met een gewichtsfactor.

Oorzaak	Gewicht
Defective	100%
Fire	15%
Infrastructure_Problem	85%

Interference	95%
Thirdparty	15%
Weather	15%
Other	25%

In dit geval worden vertragingsredenen die dichter aanleunen bij de onderneming (bijvoorbeeld *Interference*) meer in rekening gebracht dan redenen waar de onderneming minder aan kan doen (bijvoorbeeld *Thirdparty*). Een hoge waarde op deze KPI betekent dat de schuld van de vertragingen voornamelijk bij de onderneming ligt, het moet bijgevolg het objectief zijn van de onderneming om deze score naar beneden te krijgen. Het is ook een goede manier om ondernemingen met elkaar te vergelijken. Dit zal verder aan bod komen in sectie 5.2 en sectie 5.3.

Het resultaat voor deze KPI voor de Belgische spoorwegen is terug te vinden op Figuur 55 in sectie 5.3 bij de vergelijking tussen België en het Verenigd Koninkrijk.

5.1.3. Conclusie

De verdeling van de oorzaken van de vertraging komt niet volledig overeen met deze in de voorbije maandverslagen. Althans, het percentage vertragingen dat toegewezen wordt aan derden stemt wel overeen maar de verdeling Infrabel/NMBS verschilt. Een mogelijke reden hiervoor is dat niet alle storingen de schuld zijn van Infrabel en niet alle defecten de schuld van NMBS. De bepaling van de verantwoordelijkheid van de vertraging is een complexe zaak en hiervoor worden experts ingeschakeld. Het gebrek aan exhaustieve data is hier een grote spelbreker. Tweets zijn beperkt tot 140 karakters en reizigers hebben geen boodschap aan het feit of een vertraging veroorzaakt werd door Infrabel of NMBS. Het is wel reeds duidelijk geworden dat het dashboard bepaalde pijnpunten kan blootleggen, dat bewees het voorbeeld van de brand in Wijgmaal en het aantal vertragingen in de drukste stations. Helaas kan er geen conclusie getrokken worden uit het verband tussen vertragingsoorzaak en vertragingstijd aangezien de data hiervoor ontbreekt.

5.2. Verenigd Koninkrijk

5.2.1. Inleiding in verband met de opgehaalde data

Om het treinverkeer in het Verenigd Koninkrijk te analyseren, wordt tussen de acht grootste stations onderling, gebaseerd op het aantal mensen die in- en uitstappen van de periode 2011-2012 [69], data opgehaald via Darwin tussen 8 en 17 mei 2017 op weekdays. Het gaat om onderstaande stations:

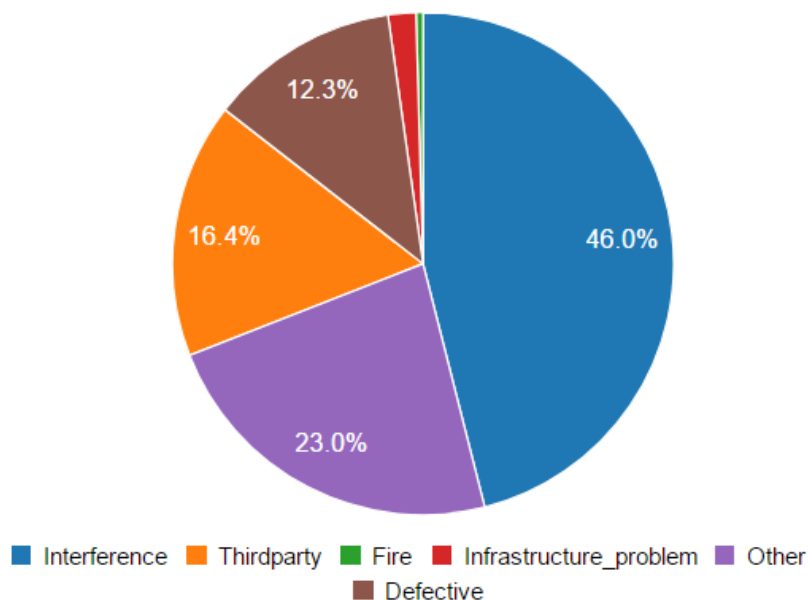
Station (Afkorting)	Aantal in- en uitstappen (2011-2012)
London Waterloo (WAT)	94 045 510
London Victoria (VIC)	76 231 290
London Liverpool Street (LST)	57 106 502
London Bridge (LBG)	52 634 024
London Charing Cross (CHX)	38 004 790
London Euston (EUS)	36 608 546
London Paddington (PAD)	33 736 546
Birmingham New Street (BHM)	31 213 842

Op deze trajecten zijn verschillende treinoperators aan het werk. Bijgevolg zal het mogelijk zijn om deze operators kwalitatief met elkaar te vergelijken op vlak van stiptheid en type oorzaken van hun vertragingen. De vier operators die deze verbindingen voorzien zijn London Midland (LM), Southeastern (SE), Southern (SN) en Virgin Trains (VT).

5.2.2. Analyse van de resultaten

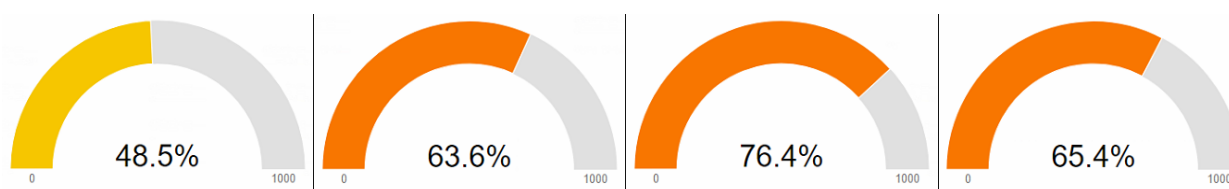
Meest voorkomende vertragingsoorzaken

Om te beginnen wordt de algemene verdeling van vertragingsoorzaken in het Verenigd Koninkrijk getoond. Dit is om een algemeen beeld te geven waarmee de operators vergeleken kunnen worden. De vergelijking met België op vlak van oorzaken is terug te vinden in sectie 5.3. Volgens Network Rail, de railinfrastructuurbeheerder, wordt 20% van de vertragingen veroorzaakt door derden [70]. Uit Figuur 43 blijkt dat dit in de voorbije periode 16,4% was, dit ligt dus in lijn met de verwachtingen.



Figuur 43 : De verdeling van oorzaken in Groot-Brittannië. (totaal aantal: 669)

Vergelijking van scores/vertragingsoorzaken van ondernemingen

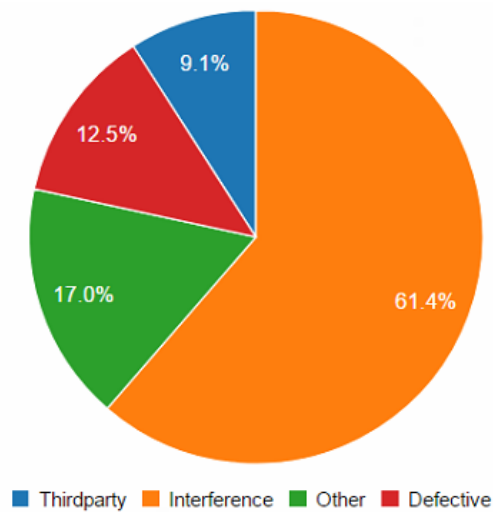


Figuur 44 : Scores voor KPI ProblemDelay (v.l.n.r. Southern, Southeastern, London Midland, Virgin Trains)

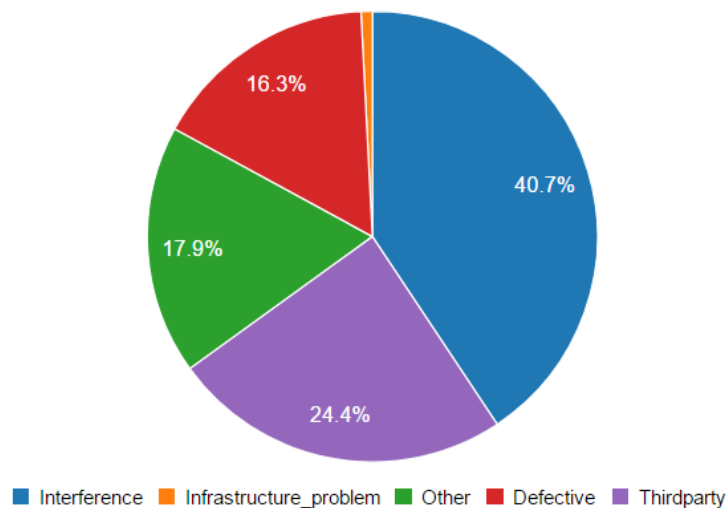
De berekening van deze score voor *ProblemDelay* werd eerder toegelicht in sectie 5.1. Uit deze berekeningen van de ondernemingen, zoals te zien op Figuur 44, is duidelijk af te leiden dat Southern de beste werking heeft. Het is echter zo dat Southern de kleinste steekproef (30 vertragingen) heeft, bijgevolg is hun resultaat het minst betrouwbaar. Opvallend is ook de veel hogere score van London Midland ten opzichte van Southeastern en Virgin trains. Voor de beleidsmakers van London Midland zou dit een waarschuwing moeten zijn om in de toekomst beter te werken.

Bij het bestuderen van de cirkeldiagrammen (zie Figuur 45-47) valt op dat London Midland inderdaad veel meer last heeft van storingen (*Interference*) en dat slechts 9,1% van alle vertragingen toegewezen wordt aan derden. Dit is opvallend omdat in het algemene beeld van het Verenigd Koninkrijk op het dashboard (zie Figuur 43) en volgens de cijfers van Network Rail dit

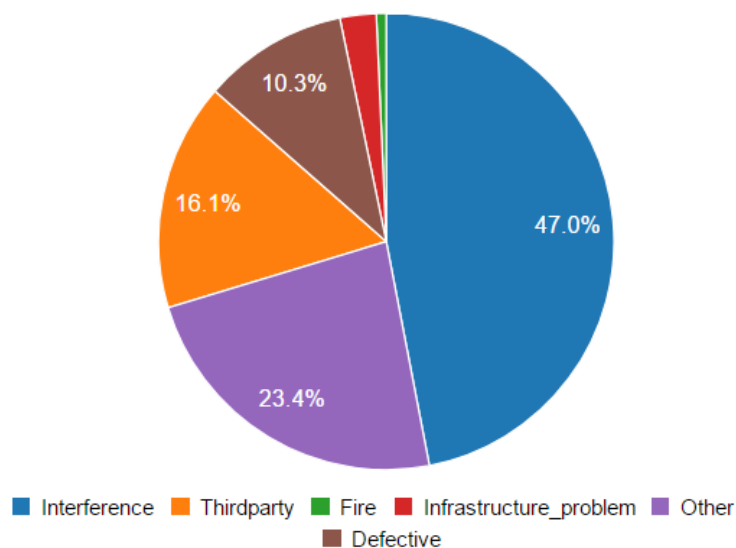
15 à 20% bedraagt. Na overleg met onze begeleiders is beslist dat de reden hiervoor is dat de lijnen waar London Midland gebruik van maakt relatief nieuw zijn en deze bijgevolg beter beveiligd zijn tegen bijvoorbeeld spoorlopers.



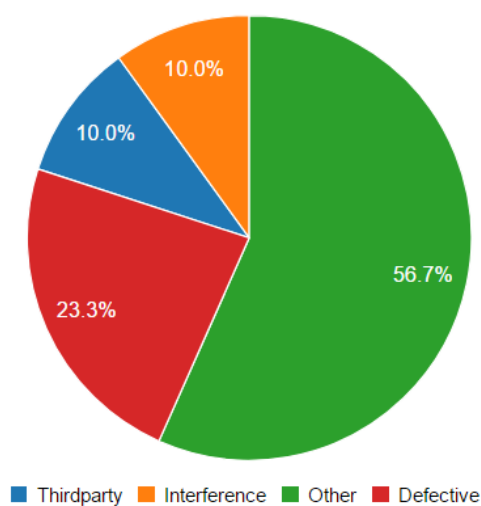
Figuur 45 : London Midland (aantal: 88).



Figuur 46 : Virgin Trains (aantal: 123).



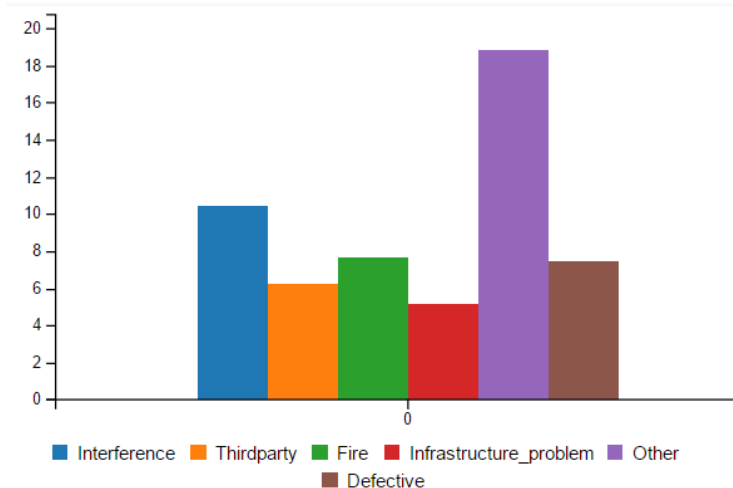
Figuur 47 : Southeastern (aantal: 428).



Figuur 48 : Southern (aantal: 30).

Verband tussen tijd en oorzaak

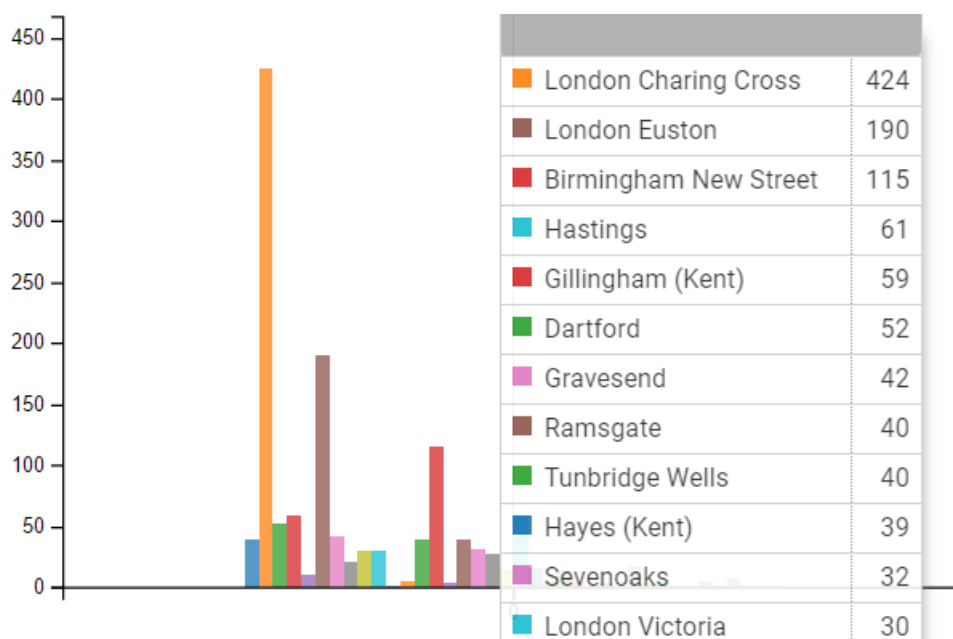
Naast de oorzaak van vertraging wordt ook de vertraging in tijd bijgehouden zoals te zien op Figuur 49. Dit is de vertraging ten opzichte van het vooropgestelde schema in het laatste station op het traject. Hieruit blijkt dat vertraging veroorzaakt door *Other* het meeste invloed heeft.



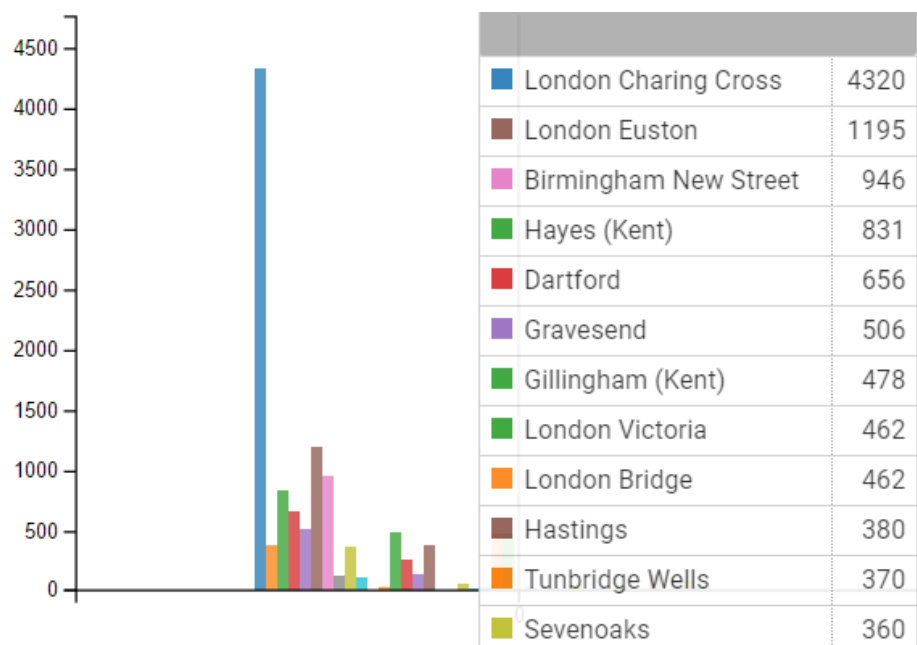
Figuur 49 : Gemiddelde vertraging in minuten per Level 1 oorzaak.

Een voorbeeld van oorzaken die gelinkt worden aan *Other* zijn vertragingen die veroorzaakt worden door problemen bij een vorige trein. Een ander voorbeeld is overbevolking van de trein of veiligheidsredenen.

Stations met de meeste vertragingen



Figuur 50 : Stations met de meeste vertragingen.

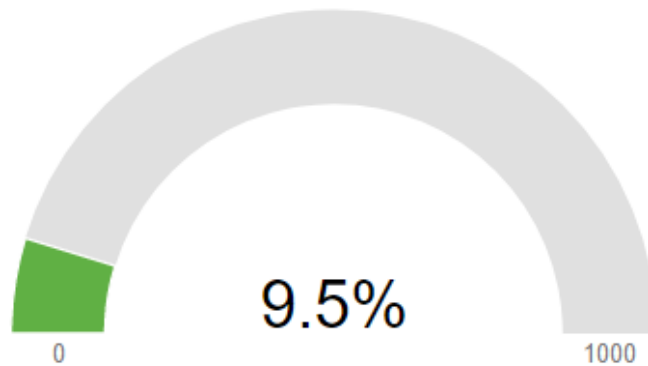


Figuur 51 : Stations waar de meeste treinen starten of stoppen.

Het meest opvallende wat kan afgeleid worden uit Figuur 50 en Figuur 51 is dat Hayes veel lager staat op de lijst met aantal vertragingen dan op de lijst van aantal treinen. Dit betekent dus dat Hayes één van de grotere stations is met de minste vertragingen.

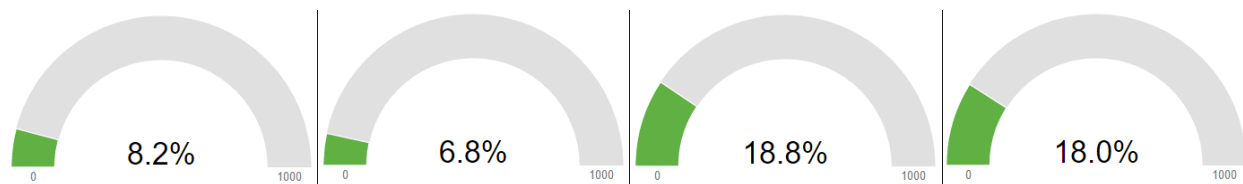
Stiptheid van de treinen

Uit de cijfers van de regering in het Verenigd Koninkrijk blijkt dat ongeveer 85% van de treinen (gerekend over de spoorondernemingen die in dit onderzoek aan bod komen) op tijd reed in 2015/16 zoals te zien op Figuur 54 [71]. Een trein die op tijd rijdt heeft 4 minuten 59 seconden of minder vertraging. Uit de cijfers van de periode die gebruikt wordt in het dashboard blijkt dat 9,5% van de treinen last hadden van vertraging, zoals te zien op Figuur 52. Er zijn in totaal 6121 services geregistreerd binnen deze periode.



Figuur 52 : Percentage treinen met vertraging groter dan of gelijk aan 5 minuten.

Uit Figuur 53 kan afgeleid worden dat London Midland en Virgin Trains het meeste last hebben van vertragingen. Let wel op dat in dit onderzoek enkel rekening gehouden moet worden met Virgin Trains West Coast aangezien East Coast niet van toepassing is op de stations waartussen data opgehaald wordt. Southern, ook bekend als Govia Thameslink Railway, en Southeastern hebben het minste last van vertragingen. Wanneer dit vergeleken wordt met de cijfers van de regering (zie Figuur 54) kan geconcludeerd worden dat deze niet volledig overeenstemmen. Vooral het verschil bij Southern (Govia Thameslink Railway) is opvallend. De reden hiervoor kan zijn dat de steekproef van stations relatief klein is en vanwege de korte tijdsperiode.



Figuur 53 : Percentage treinen met vertraging groter dan of gelijk aan 5 minuten per onderneming. (v.l.n.r. Southern, Southeastern, London Midland, Virgin Trains)

c2c	96,7
Chiltern Railways	94,4
CrossCountry	89,4
East Midlands Trains	92,8
First TransPennine Express	87,7
Govia Thameslink Railway	81,5
Great Western Railway	89,5
Greater Anglia	89,3
London Midland	88,2
Northern Rail	90,7
South West Trains	90,1
Southeastern	86,8
Virgin Trains East Coast	85,2
Virgin Trains West Coast	86,0

Figuur 54 : Percentage treinen die op tijd aankomen volgens de regering (gemiddelde over 2015/16) [71].

5.2.3. Conclusie

Het is duidelijk geworden dat verschillende spoorondernemingen in het Verenigd Koninkrijk een opvallend verschillende werking hebben. Zo zijn de vertragingen afkomstig van verschillende types oorzaken. De grootste constante (in drie van de vier gevallen) is dat de meeste vertragingen afkomstig zijn van storingen (Interference). Een ander opvallend resultaat is dat verschillende oorzaken een verschillende vertragingstijd genereren, met name de zogenaamde overige (Other) oorzaken genereerden veel meer vertraging. Ook bepaalde stations hadden duidelijk relatief meer last van vertraging dan andere stations. De stiptheid van de treinen is in het algemeen beter is dan wat verwacht kan worden wanneer naar de voorbije jaarresultaten gekeken wordt. Er is echter wel een duidelijk verschil in stiptheid tussen de verschillende ondernemingen.

Een belangrijke voetnoot bij deze resultaten is dat de mapping van de vertragingsoorzaken op de ontologie door experts in dit vakgebied zou moeten worden uitgevoerd. In dit geval is dit gebeurd door de makers van deze thesis. Hierdoor worden een aantal oorzaken gekwalificeerd als overig (Other) wanneer mogelijks een andere categorie beter zou zijn.

5.3. Vergelijking tussen België en het Verenigd Koninkrijk

Uiteraard is het ook mogelijk om de werking van verschillende landen met elkaar te vergelijken, in dit geval zal de vergelijking gemaakt worden tussen België en het Verenigd Koninkrijk.

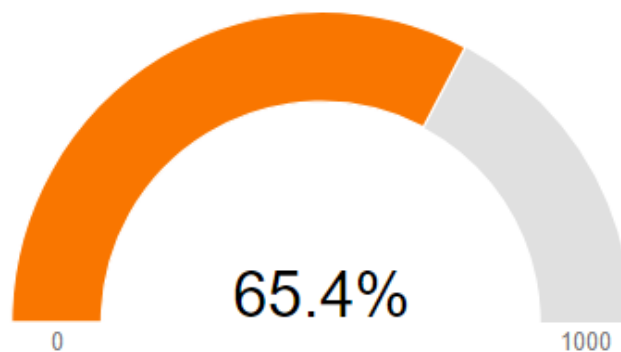
5.3.1. Analyse van de resultaten

Vergelijking oorzaak van vertraging

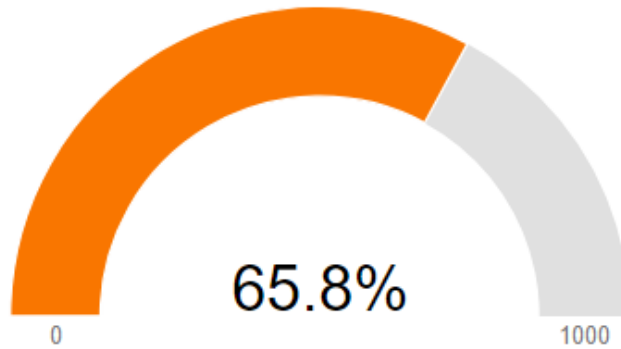
De eerste vergelijking zal gebeuren op vlak van de oorzaak van vertraging, voor België is dit terug te vinden op Figuur 34 en voor het Verenigd Koninkrijk op Figuur 43. Het grootste opvallende verschil is het aantal vertragingen veroorzaakt door derden. In België is dit 10% hoger, dit kan op verschillende manieren geïnterpreteerd worden. In België rijdt ongeveer 90% van de treinen op tijd [64]–[66], wat vergelijkbaar is met de treinen in het Verenigd Koninkrijk (zie sectie 5.2.2). Het verschil is dat in België treinen met minder dan 6 minuten vertraging gerekend worden als stipte treinen, in het VK ligt de grens op 5 minuten. Dit betekent dat de verdeling van de vertraging redelijk goed met elkaar te vergelijken valt.

Uit het feit dat er in België meer vertraging is door derden kan geconcludeerd worden dat de Belgische sporen minder goed beveiligd zijn tegen derden. Een andere manier om dit verschil te interpreteren is dat België eigenlijk een betere werking heeft dan het Verenigd Koninkrijk. De reden hiervoor is dat zowel in België als in het VK 10% van de treinen vertraging hebben maar in België ligt de schuld minder bij de spoorondernemingen zelf.

Deze redenering wordt deels verdergezet in het berekenen van de score van de KPI *ProblemDelay* (zie sectie 5.1). De score op de KPI is quasi identiek voor beide ondernemingen (zie Figuur 55 en 55). De reden hiervoor is het relatief grote aantal vertragingen door *Other* bij het VK.



Figuur 55 : Belgische score op KPI 'ProblemDelay'.



Figuur 56 : Britse score op KPI 'ProblemDelay'.

5.3.2. Conclusie

In België is de spoorindustrie genationaliseerd en in het Verenigd Koninkrijk is deze geprivatiseerd, desondanks is de totaalscore bijna identiek. Het is wel zo dat er merkbaar minder vertraging veroorzaakt wordt door derden in het Verenigd Koninkrijk. Bijgevolg kan er op het vlak van vertragingsoorzaken niet beslist worden of privatisering beter is dan nationalisering, of omgekeerd.

Hoofdstuk 6: Testing

6.1. Performance testing

Voor de performantie van de KPISetup te meten wordt Apache JMeter¹⁰ gebruikt. De Apache JMeter applicatie is een Java-applicatie ontworpen voor het testen van load op een applicatie en het meten van performantie. Dit zal dus getest worden in dit hoofdstuk: de http-endpoint (van KPISetup) wordt gebruikt om http-aanvragen naar toe te sturen en zo de antwoordtijd naargelang de aanvraag te evalueren.

De gebruikte set-up voor het uitvoeren van deze testen kan in Figuur 57 terug gevonden worden.



Figuur 57: Specificatie van gebruikte set-up voor testing.

6.1.1. Load/Stress testing

De belangrijkste functionaliteiten van het platform worden getest, namelijk gegevens ophalen uit de databank, rule-based semantisch redeneren en het berekenen van een KPI. Er zijn twee verschillende performance testen uitgevoerd: een met een statische load en een met een dynamische load. De resultaten van deze testen zullen in onderstaande paragrafen verklaard worden.

¹⁰ <http://jmeter.apache.org/>

In beide testen zullen er een totaal van 4 verschillende aanvragen gestuurd worden:

- **AllDelayEvent:** Een *Measurement* met semantische query. Verdere details:

Sparql =

```
PREFIX : <http://www.semanticweb.org/cedri/syvermee/thesis#>
construct where {
    ?s :Delay_Event ?o.
    ?s ?p ?x.
    ?s :hasCountry &{COUNTRY}.
}
```

Sparql-dl=

```
PREFIX : <http://www.semanticweb.org/cedri/syvermee/thesis#>
SELECT ?tweet ?x WHERE {
    PropertyValue(?tweet, :Delay_Event, ?event),
    PropertyValue(?event, :above, ?x),
    PropertyValue(?x, :hasLevel, :Level_1)}
OR WHERE {
    PropertyValue(?tweet, :Delay_Event, ?x),
    PropertyValue(?x, :hasLevel, :Level_1)
}
```

- **GetTree:** Dit haalt alle KPI's met bijhorende gegevens op, voornamelijk een databasequery.

- **Indicator3:** Een high level KPI die geen afhankelijkheden heeft met *Measurements*.

Verdere details:

Formula = (TestKPI/50)

TestKPI is een hardgecodeerde KPI met 250 als waarde.

- **Login:** Dit haalt alle *agency's* met bijhorende gegevens, voornamelijk een databasequery.

Static load

Allereerst de test met statische load, deze test begint met 10 threads en zal 5 loops maken.

De 10 threads weerspiegelen 10 organisaties die één bepaalde onderneming wilt vergelijken, wat gebruikt werd als leidraad.



Figuur 58: Resultaten van de load test. X-as: Tijd, Y-as: Response time.

Het eerste wat opvalt in bovenstaande grafiek is dat er grote fluctuaties zijn als het aankomt op rule-based semantisch redeneren. Er wordt ook duidelijk dat het rule-based semantisch redeneren veruit de meest tijd opslopende operatie is, wat wel te verwachten was aangezien dit meer complexe operaties bevat dan de andere aanvragen.

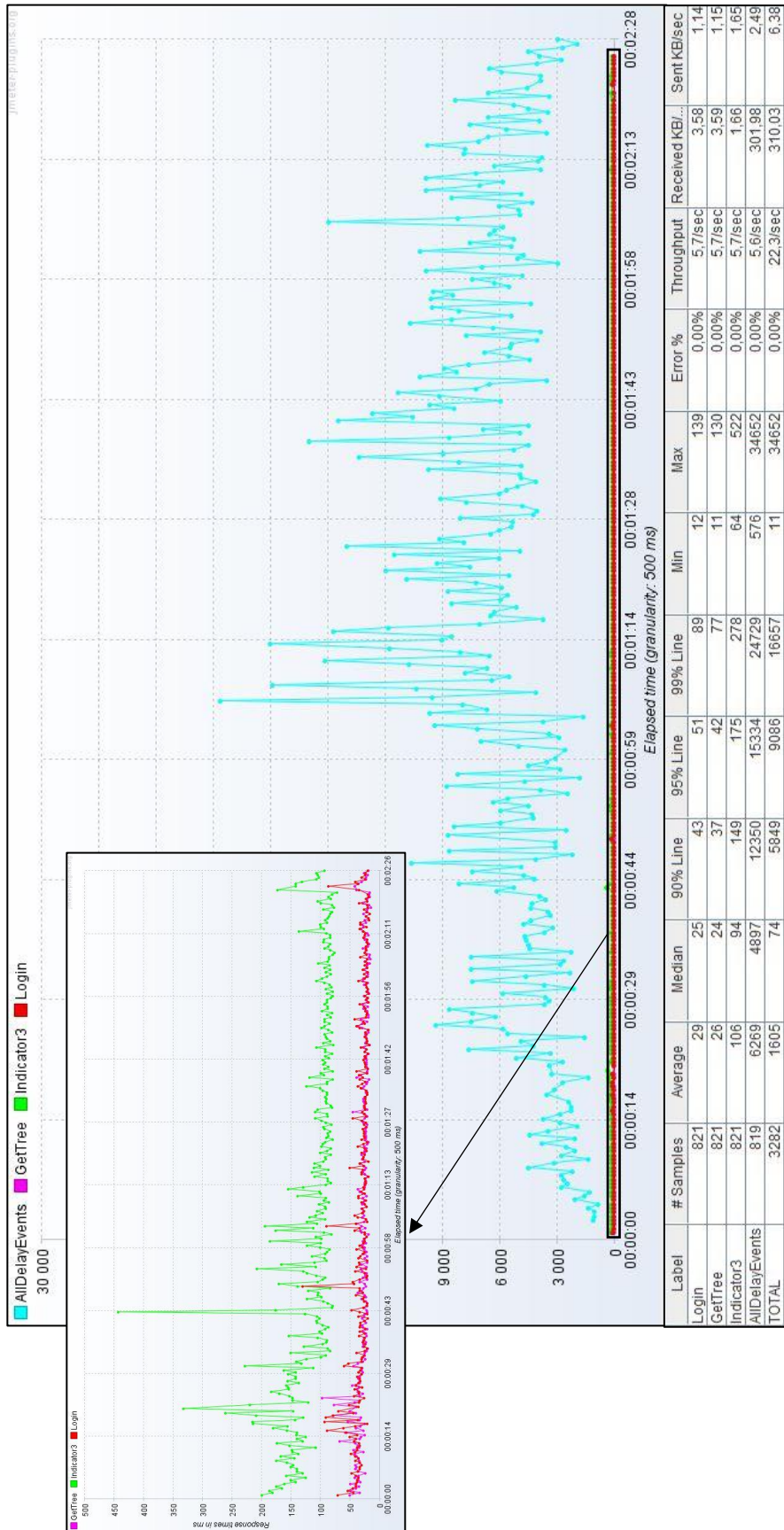
Dynamic load

In deze paragraaf wordt onderzocht hoe stressbestendig het platform nu uiteindelijk is. Uit voorafgaande test bleek het rule-based semantisch redeneren toch niet zo stabiel, omwille hiervan wordt deze stresstest uitgevoerd om de limieten te ontdekken. Om de test nog iets complexer te maken wordt er deze keer gebruik gemaakt van een dynamische load die schaalst in tijd (zie Figuur 59).



Figuur 59 : De geteste userload, om de 10 seconden starten er 10 nieuwe threads tot een maximum van 50 threads waarna het na een minuut snel terug daalt.

Uit Figuur 60 is af te leiden dat het rule-based semantisch redeneren niet echt schaalbaar is. De gemiddelde responsetijd bleef bij de andere operaties ongeveer hetzelfde ten opzichte van vorige test, dit in tegenstelling tot het rule-based semantisch redeneren.

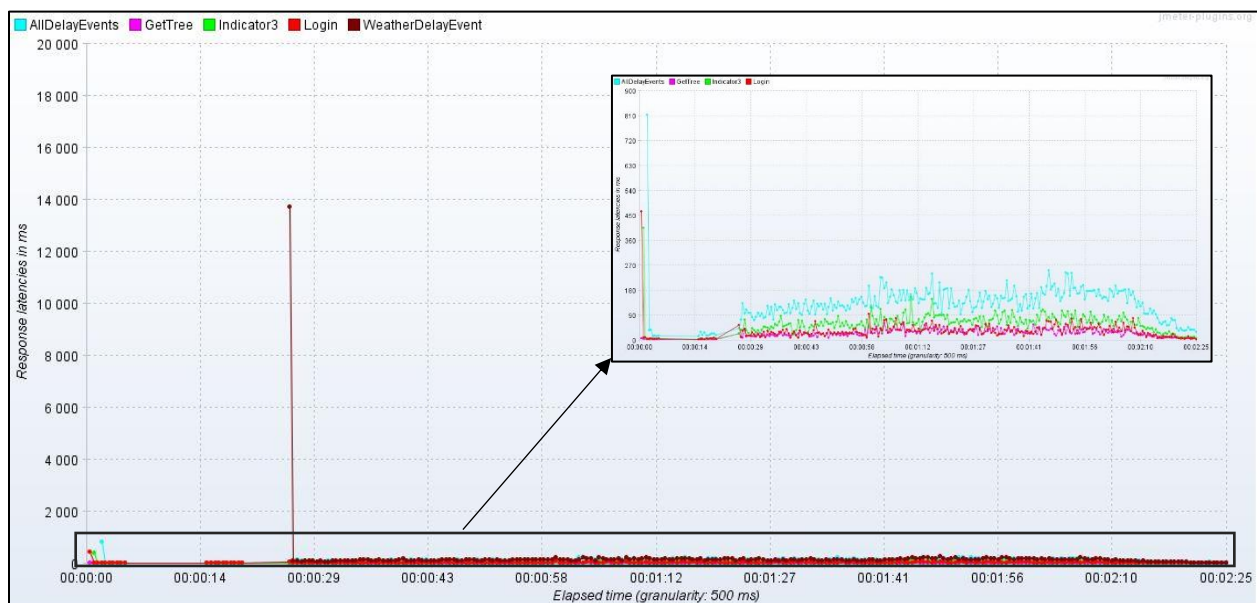


Figuur 60 : Resultaten stress test. X-as: Tijd, Y-as: Response time.

6.2. Conclusie

Het is duidelijk dat het rule-based semantisch redeneren een grote bottleneck is. Dit kan eventueel gecompenseerd worden door het gebruik van caching. Veel KPI's zullen vaak dezelfde low level afhankelijkheden hebben, hierdoor is het onnodig om telkens voor elke KPI dezelfde berekeningen uit te voeren. Uiteraard kunnen resultaten maar voor een beperkte tijd gecached worden zodat we niet telkens oude data terugkrijgen.

Door de functionaliteit van LimeDS kan deze caching eenvoudig geïmplementeerd worden met behulp van annotaties. In onderstaande figuur (Figuur 61) wordt gebruik gemaakt van een caching van 30 seconden. Wat heel opvallend is, is dat de caching schijnbaar langer duurt dan 2 min 25 seconden. Dit ondanks het feit dat gespecificeerd is dat het tot 30 seconden zou mogen cachen. Door extensieve testing is geconcludeerd dat er mogelijks een bug zit in LimeDS.



Figuur 61 : Stress test met caching

Door deze bug zal er dan toch geen caching worden gebruikt in de applicatie wat de applicatie gevoelig maakt aan een grote hoeveelheid (en overbodige) semantische query's.

Achteraf bekeken hadden we misschien meer gebruik moeten maken van de OWL 2 RL specificatie in plaats van ons te baseren op OWL 1 DL. De OWL 2 RL specificatie dient voor schaalbare reasoning dus zou dit ook eventueel invloed gehad hebben op de performantie.

Nog een bug die door het testen aan het licht is gekomen, is dat het gebruik van de methode `addFirst` van `JsonArray` in combinatie met caching ervoor zorgt dat een `ConcurrentModificationException` optreedt. Deze error komt wel vaker voor bij het itereren van lijsten indien men ondertussen modificaties doet. Deze bug wordt opgelost door het gebruik van andere methodes van deze klasse die tezamen dezelfde functionaliteit aanbieden. De echte reden waarom dit enkel bij caching gebeurde blijft een mysterie.

Hoofdstuk 7: The past and the future

Er zijn nog een aantal zaken die we hadden willen implementeren maar die we niet hebben uitgevoerd. De reden hiervoor is de beperkte tijd en het niet prioritair zijn van een aantal zaken.

Een eerste voorbeeld hiervan zijn CRUD-operaties. Momenteel is het reeds mogelijk om KPI's te creëren en de resultaten te lezen. Het is echter nog niet mogelijk om deze aan te passen of te verwijderen. Een tweede nuttige uitbreiding zou het dynamisch toevoegen van databronnen zijn. Op deze manier zouden via het dashboard databronnen gelinkt kunnen worden aan de ontologie, dit zou ook de mogelijkheid toevoegen om Comma-Separated Values (CSV)/ Extensible Markup Language (XML)/ JSON als databronnen te gebruiken. Over deze functionaliteit werd al eens nagedacht, RDF Mapping Language (RML)¹¹ zou een goed vertrekpunt zijn om deze functionaliteit te implementeren.

Het zou nuttig zijn indien SPARQL en SPARQL-DL query's getest worden op correctheid voordat ze worden opgeslagen in de databank. Hierbij aansluitend zou het ook handig zijn indien niet bij elk Measurement een SPARQL-query én een SPARQL-DL-query gedefinieerd moet worden. In sommige gevallen is een SPARQL-query voldoende om het resultaat te berekenen, dit zou een performantiewinst opleveren aangezien er niet meer overbodig geredeneerd moet worden.

Een volgende performantiewinst kan bereikt worden via caching. Zoals vermeld in Hoofdstuk 6: werkt de caching niet volledig zoals verwacht en werd dit bijgevolg voorlopig weggelaten. Ook het gebruik van OWL 2 RL had waarschijnlijk ook zijn invloed gehad op de performantie.

Bij terugblikken op de resultaten van de testing is er gebleken dat het rule-based semantisch redeneren niet al te performant is. Mogelijks door deze functionaliteit multithreaded te maken zou dit sterk verbeteren. Dit kan echter niet met zekerheid gezegd worden. Als we dit samen met het doodlopend spoor van dynamisch toevoegen van SWRL Rules in afweging nemen had het waarschijnlijk beter geweest indien voor een databank gekozen werd die vooral het semantisch redeneren met voorafgaande opgestelde rules optimaliseert. Dit had evenveel gekund met

¹¹ <http://rml.io/RMLmappingLanguage.html>

Stardog. De nieuwste versie van Stardog, namelijk Stardog 5.0-beta¹², is 8 Mei 2017 uitgekomen (lang na de keuze van GraphDB) en ondersteunt veel van de benodigde functionaliteiten.

¹² <http://www.stardog.com/>

Besluit

In de eerste plaats werd de doelstelling en de manier waarop deze doelstelling bereikt zou worden beschreven. Hierna volgde de literatuurstudie waarin een uitgebreide lijst van KPI's binnen het openbaar vervoer verzameld werd. Deze lijst van KPI's is veel uitgebreider dan de mogelijkheden aan KPI's in de uiteindelijke applicatie, maar het geeft een indicatie van waar dergelijke ondernemingen op focussen. Er werden ook verschillende soorten dashboards bekeken. De verschillende werking van de spoorondernemingen in België en het Verenigd Koninkrijk werd ook onder de loep genomen. Hiervoor werd zowel naar de algemene werking als de stakeholders van de bedrijven gekeken. Naast het socio-economische aspect van de literatuurstudie werd er ook nog een technologisch onderzoek gedaan in verband met ontologieën. Hierin volgde de eerste kennismaking met ontologieën en de verschillende technologieën.

Al dit voorbereidende werk werd gevolgd door de implementatie. Hierin werd de ontologie ontwikkeld en de data van verschillend databronnen opgehaald. Hiernaast werd ook een database opgezet om alle data op te slaan zodat hier query's op kunnen uitgevoerd worden. Als laatste stap volgde nog de implementatie van het dashboard waarop verschillende KPI's gecreëerd en getoond kunnen worden.

De belangrijkste vraag bij het afwerken van de thesis is uiteraard: is de doelstelling behaald? Na het analyseren van de resultaten blijkt dat het mogelijk is om de pijnpunten van een onderneming bloot te leggen. Een vergelijking tussen verschillende ondernemingen of landen kan ook gebeuren zoals aangetoond werd bij de verschillende ondernemingen in het Verenigd Koninkrijk en bij de vergelijking tussen België en het Verenigd Koninkrijk.

Een belangrijke opmerking bij alle resultaten is dat de data van verschillende ondernemingen niet eenduidig met elkaar kan gemapt worden. Het zou bijgevolg dus vooral interessant zijn indien er bijvoorbeeld vanuit een groot orgaan zoals de Europese Unie richtlijnen worden opgesteld met betrekking tot welke data verzameld moet worden en op welke manier dit moet gebeuren. Op deze manier zouden verschillende ondernemingen beter met elkaar vergeleken kunnen worden.

Referenties

- [1] S. Cardoen, “Koning auto verliest (een beetje) terrein,” 12-Feb-2016.
- [2] Z. Saerens and H. Beeckman, “Novemberfiles stijgen jaar na jaar,” 30-Nov-2016.
- [3] “Integrail,” 2015.
- [4] Z. Li and D. A. Hensher, “Crowding in public transport: A review of objective and subjective measures,” 2013.
- [5] I. Kaparias, M. Bell, and M. Tomassini, “CONDUITS: Key Performance Indicators for traffic management and Intelligent Transport Systems,” 2011.
- [6] Sydney Metropolitan Bus Service Contract, “Schedule 4: Key Performance Indicators.”
- [7] De Lijn, “Totale milieu-impact,” 2013. [Online]. Available: <https://www.delijn.be/nl/overdelijn/organisatie/zorgzaam-ondernemen/milieu/ecologische-voetafdruk.html>.
- [8] NMBS, “NMBS - Tevredenheidsenquête,” 2015.
- [9] NMBS, “Rapport over de kwaliteit van de dienstverlening 2013,” 2013.
- [10] E. Naleeza, G. Sun, E. Margaret, and A. Foo Jong, “Journeys - Comparison of Public Transport Operations Key Indicators in Comparison,” 2011.
- [11] G. Demydyuk, “Optimal financial key performance indicators: evidence from the airline industry,” 2011.
- [12] ServiceNow, “KPI Library,” 2017. [Online]. Available: <http://kpilibrary.com>.
- [13] J. Dehon, “Jaarlijks Veiligheidsverslag Infrabel,” 2015.
- [14] S. Mnif, S. Galoui, S. Elkosantini, S. Darmoul, and L. Ben Said, “Ontology based performance evaluation of public transport systems,” *2015 4th Int. Conf. Adv. Logist. Transp.*, no. May, pp. 205–210, 2015.
- [15] A. Enoma and S. Allen, “Developing key performance indicators for airport safety and security,” 2007.
- [16] Infrabel, “2015 in Vogelvlucht,” 2015.
- [17] Infrabel, “Duurzaamheids-verslag 2015,” 2015.
- [18] Geckoboard, “Geckoboard - Website,” 2017. [Online]. Available: <https://www.geckoboard.com/>.
- [19] KPI Fire, “KPI Fire - Website,” 2015. [Online]. Available: <https://www.kpifire.com/>.
- [20] Roebank Publishing Limited, “Target Dashboard - Website,” 2017. [Online]. Available: <https://www.targetdashboard.com/>.
- [21] A. a. Atkinson, J. H. Waterhouse, and R. B. Wells, “A Stakeholder Approach to Strategic Performance Measurement,” 1997.
- [22] Cambridge University Press, “Cambridge Dictionary,” 2017. [Online]. Available: <http://dictionary.cambridge.org>.
- [23] Infrabel, “Duurzaamheidsverslag,” 2014.
- [24] Infrabel, “Management en Structuur,” 2017.
- [25] Infrabel, “A FIRST FOR INFRABEL ! Five years of Traffic Control,” 2013.
- [26] NMBS, “NMBS - Website,” 2017. [Online]. Available: <http://www.belgianrail.be/nl/>.
- [27] Southern Railway, “Southern Railway,” 2017.
- [28] Arriva Mainland, “No Title,” 2011.
- [29] FOD Mobiliteit, “Bevoegdheidsverdeling,” 2017.
- [30] Infrabel, “Onze geschiedenis,” 2017. [Online]. Available: <https://www.infrabel.be/nl/over-ons/onze-onderneming/onze-geschiedenis>.
- [31] Infrabel, “Verskil tussen Infrabel en NMBS,” 2017. [Online]. Available: <https://www.infrabel.be/nl/over-ons/onze-onderneming/verschil-tussen-infrabel-en-nmbs>.

- [32] NMBS, “Visie en Missie,” 2017. [Online]. Available: http://www.belgianrail.be/nl/corporate/onderneming/visie_missie.aspx.
- [33] NMBS, “NMBS I Visie en activiteiten 2014,” 2014.
- [34] Infrabel, “Bijlage B.10 Prestatieregeling – van toepassing vanaf 1 januari 2017 1. Berekening van de minuten vertraging,” pp. 1–5, 2017.
- [35] Open Knowledge Belgium, “iRail,” 2017. [Online]. Available: <https://hello.irail.be/>.
- [36] Open VLD, “KAMER STEM T MORGEN AMBITIEUZE OPEN DATA-WET,” 2016. [Online]. Available: <http://www.openvld.be/?type=nieuws&id=1&pageid=86061>.
- [37] Open Knowledge Belgium, “iRail - Applicatie,” 2017. [Online]. Available: <https://irail.be/app>.
- [38] J. Decelle, “Key Performance Indicators ‘STIPTHEID,’” 2015.
- [39] rail.co.uk, “UK Rail Operators,” 2017. [Online]. Available: <http://www.rail.co.uk/our-partners/rail-operators/>.
- [40] Office of Rail and Road, “Mainline Network,” 2017. [Online]. Available: <http://orr.gov.uk/about-orr/who-we-work-with/rail-infrastructure/mainline-network>.
- [41] C. Hope and N. Harley, “Network Rail to be stripped of control over Britain’s train tracks as operators will win power to improve services,” *Telegraph*, 03-Dec-2016.
- [42] National Rail Enquiries, “Darwin Data Feeds,” 2017. [Online]. Available: <http://www.nationalrail.co.uk/100296.aspx>.
- [43] T. R. Gruber, “Toward principles for the design of ontologies used for knowledge sharing,” *Int. J. Hum. Comput. Stud.*, vol. 43, no. 5–6, pp. 907–928, 1995.
- [44] T. Gruber, “Ontology definition,” 2007. [Online]. Available: <http://web.dfc.unibo.it/buzzetti/IUcorso2007-08/mdidattici/ontology-definition-2007.htm>. [Accessed: 20-Feb-2017].
- [45] W3C, “VOCABULARIES,” W3C. [Online]. Available: <https://www.w3.org/standards/semanticweb/ontology>. [Accessed: 14-Jul-2017].
- [46] N. F. Noy and D. L. McGuinness, “Ontology Development 101: A Guide to Creating Your First Ontology.” [Online]. Available: http://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html. [Accessed: 16-Feb-2017].
- [47] R. Cyganiak, “RDF 1.1 Concepts and Abstract Syntax,” 2014. [Online]. Available: <https://www.w3.org/TR/rdf11-concepts/#data-model>. [Accessed: 15-Feb-2017].
- [48] M. Smith, C. Welty, and D. L. McGuinness, “OWL Web Ontology Language Guide,” 2004. [Online]. Available: <https://www.w3.org/TR/owl-guide/>. [Accessed: 17-Feb-2017].
- [49] R. Agrawal, “Difference between OWL Lite, DL, and Full,” 2007. [Online]. Available: <https://ragrawal.wordpress.com/2007/02/20/difference-between-owl-lite-dl-and-full/>. [Accessed: 16-Feb-2017].
- [50] A. Fokoue, A. Kershenbaum, L. Ma, and E. Schonberg, “The Summary Abox : Cutting Ontologies Down to Size.”
- [51] B. C. Grau, I. Horrocks, B. Motik, B. Parsia, P. F. Patel-Schneider, and U. Sattler, “OWL2: The next step for OWL,” *J. Web Semant.*, vol. 6, no. 4, pp. 309–322, 2008.
- [52] B. Motik, “OWL 2 Web Ontology Language Profiles (Second Edition),” W3C, 2012. [Online]. Available: <https://www.w3.org/TR/owl2-profiles/>.
- [53] R. Stevens and U. Sattler, “How does a reasoner work?,” 2014. [Online]. Available: <http://ontogenesis.knowledgeblog.org/1486#ITEM-1486-1>. [Accessed: 16-Mar-2017].
- [54] R. Stevens and U. Sattler, “(I can’t get no) satisfiability,” 2013. [Online]. Available: <http://ontogenesis.knowledgeblog.org/1329>.
- [55] I. Horrocks, H. Boley, and B. Grosz, “SWRL: A Semantic Web Rule Language.” [Online]. Available: <https://www.w3.org/Submission/SWRL/>.
- [56] M. Horridge and S. Bechhofer, “The OWLAPI: A Java API for OWL ontologies,” *Semant. Web*, vol. 2, no. 1, pp. 11–21, 2011.

- [57] Wikipedia, “JSON-LD.” .
- [58] “A Lightweight Modular Environment for Data-oriented Services.” [Online]. Available: <https://www.iminds.be/nl/technology/LimeDS>.
- [59] “Wiki LimeDS.” [Online]. Available: <https://bitbucket.org/ibcndevs/limeds-framework/wiki/Home>.
- [60] “LargeTripleStores.” [Online]. Available: <https://www.w3.org/wiki/LargeTripleStores>.
- [61] C. Diamantini, L. Genga, D. Potena, and E. Storti, “Collaborative building of an ontology of key performance indicators,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 8841, pp. 148–165, 2014.
- [62] “SPARQL-DL API.” [Online]. Available: <http://derivo.de/en/resources/sparql-dl-api/>.
- [63] A. Café, “SPARQL-DL - Theory & Practice,” 2013. [Online]. Available: <https://www.slideshare.net/adrielcafe/sparqldl-theory-practice>. [Accessed: 20-Mar-2017].
- [64] NMBS, “Reporting stiptheid – Januari 2017,” 2017.
- [65] NMBS, “Reporting stiptheid - Februari 2017,” 2017.
- [66] NMBS, “Reporting stiptheid - Maart 2017,” 2017.
- [67] Infrabel, “Infrabel voert strijd tegen spoorlopers,” 2017. [Online]. Available: <https://www.infrabel.be/nl/pers/infrabel-voert-strijd-tegen-spoorlopers>. [Accessed: 17-May-2017].
- [68] De Redactie, “Spoorwegberm Wijgmaal brandde een tijdje,” 21-Mar-2011.
- [69] The Guardian, “Every train station in Britain listed and mapped: find out how busy each one is,” 2013.
- [70] Network Rail, “Delays explained,” 2017. [Online]. Available: <https://www.networkrail.co.uk/running-the-railway/looking-after-the-railway/delays-explained/>. [Accessed: 14-May-2017].
- [71] Department for Transport, “Proportion of trains running on time,” 2015. [Online]. Available: <https://www.gov.uk/government/publications/proportion-of-trains-running-on-time>. [Accessed: 22-May-2017].

