Intelligent fire risk monitor based on Linked Open Data

Nicky van Oorschot
netage.nl
nicky_van_oorschot@netage.nl
The Netherlands

Bart van Leeuwen
netage.nl
bart_van_leeuwen@netage.nl
The Netherlands

ABSTRACT

Every day the Fire department of the Netherlands work hard to save people's lives. Therefore, they have been investing in Business Intelligence approaches for several years, to get more information for accident prevention and accident fighting. In this paper Linked Open Data has been used as a business intelligence approach for the creation of dwelling fire risk profiles based on demographic data. During the research a Proof of Concept showed the appliance of Linked Open Data for this purpose. However the data have some quality mismatches, such as: outdated, accuracy issues and not 100% completed. From evaluation sessions it turned out that the outcomes show similarities with a fire incident map and the gut feeling of several firefighters.

1. INTRODUCTION

The primary objective of the fire department is to save people's lives. Saving lives is about seconds. Complete information could make the difference. For instance, the maximum response time is determined by Dutch law (Binnenlandse Zaken, 2011). The factors which are taken into account to determine the maximum response time for a building are the type of the building and the year of construction. But the risk of a dwelling fire and the development of a dwelling fire depends on multiple factors. Some of these factors are not taken into account. Therefore, the use of "Business Intelligence" within Brandweer Nederland is an upcoming interest. Investments to harvest data and process data into valuable information have been made since the last couple of years. At the moment, Brandweer Nederland is working on business intelligence, loads of data are stored in data warehouses to get information on accident prevention and accident fighting (Nederland Brandweer, 2014). Besides the approach with data warehouses, Brandweer Nederland is highly interested in another approach: linked open data (LOD). "*Linked Open Data (LOD) is a growing movement for organisations to make their existing data available in a machine-readable format. This enables users to create and combine data sets and to make their own interpretations of the data available in digestible formats and applications.*" (Bauer & Kaltenböck, 2012). Open datasets have grown exponentially (Stellato, 2012). Combining various Datasets could provide more viable information needed for accident prevention and accident fighting. For example, research showed that residences which are more and better isolated (A+ Energy Label[1]) develop another type of fire, once the building is on fire (Schaap, 2013). Therefore, the level of isolation could be used to determine the risk of dwelling fire, and maybe even forecast the type of fire that will develop. The Linked Open Data approach differs from the warehouse approach. Data warehouses and their maintenance are expensive. Linked data gathers the information from the original source when it is needed. Less investments are necessary in the technical infrastructure. The datasets are maintained by the source itself and provided by an API or by files. Combining more datasets means that new outcomes and insights are created, or established outcomes and insights change. Moreover, change in outcomes could affect the determined response time or other decisions, which provide the extra seconds necessary to save somebody's life. This study examines whether linked open datasets could be used as a dynamic way to provide the fire department with quality intelligent information.

1.1 Main Research Question

"Does linked open data provide a qualitative and dynamic way to create a dynamic fire risk profile monitor for cities and neighbourhoods?"

1.1.1 Research Sub Questions

1. Which datasets are necessary and which datasets are available?
2. What is the quality of the data and is the quality sufficient?
3. To what extend is it possible to use linked data to get a solid demographic fire risk profile?
4. To what extend is it possible to create a dynamic linked data system, for the creation of fire risk profiles?

1.2 Research Design

Based on statistical relations from the Fire department Amsterdam ("Handreiking sociaal woningbrandrisicoprofiel"), matching open datasets were gathered from the World Wide Web.

These datasets were combined and linked to each other and used to develop a Proof of Concept, developed with C#.NET and the Dotspatial library for geographical features[2], the outcomes of the Proof of Concept were validated by four firefighters from the safety regions Rotterdam-Rijmond and Midden- en West Brabant. Validation was based on their gut feelings and based on an older map with fire incidents (2005-2008) and a fire risk map of the fire department.

2. PROOF OF CONCEPT

Through interviews with several firefighters, it has become clear that in this area little research has been done ("*We do not know a lot about it, therefore we do not use it in the risk profiles we construct*"). The safety region of Amsterdam-Amstelland has done a statistical research on the relations between demographic factors and the risk of a dwelling fire. With this information, they could conduct precise communication with the appropriate target risk groups. In the Netherlands, this is the only knowledge on this topic.

---

[1] https://data.overheid.nl/data/dataset/woningen-met-energielabels-eigendom

[2] https://dotspatial.codeplex.com/

2.1 Dwelling Fire Risk index

The relations were extracted from "Handreiking sociaal woningbrandrisicoprofiel". Table 1 shows the factors and the associated risks.

Table 1; Dwelling fire risk factors

Factor	Risk (%)
Western immigrant resident	15,7% lower risk
Native main resident	27% lower risk
One parent family	40% higher risk
More people registered household	7,6% higher risk
Single-person households	11,5% higher risk
0-11 year (s) in household	10,6% higher risk
12-17 year (s) in household	39,1% higher risk
Main resident 65 or older	21,2% lower risk
Construction year 1991 or younger	15,8% lower risk
Owner-occupied dwelling	27,4% lower risk
Multiple storey apartments	24,8% higher risk

Some factors influence the risk positively and some negatively. Since the factors are not independent, it is not possible to multiply the risks with each other as an independent probability calculation and therefore another method had to be used.

2.1.1 Dwelling Risk Index calculation

All found data (proportions) were multiplied with the probability and added to each other. Therefore, it was possible to create an dwelling risk index for every neighbourhood.

The dwelling fire risk is calculated using the following formula:

Equation 1; Dwelling Fire Risk Formula;

$$\text{RISK} = \sum (Fn * Pn)$$

The proportion of the factor, times the probability gives a sub score. The sub scores of all factors added up, give the risk index. For example, an area with 15% of one parental families (40% risk) results in the following sub score: F*P = 15*0.40 = 6.

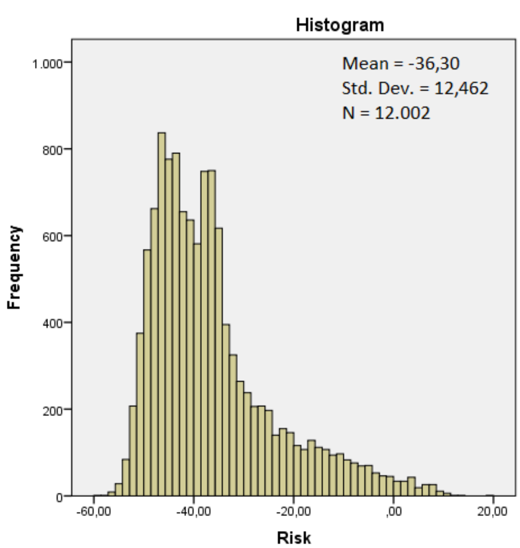

Figure 1; Distribution dwelling fire risks

By linking the datasets, in accordance with the statistical relations and formula, the risk indexes have been calculated. Missing results have been filled up with the mean of the dataset. This has resulted in the following distribution of risk indexes:

2.2 Necessary datasets

Datasets were gathered from the internet. Extra datasets were gathered to create analysis beyond only the dwelling fire risk assessment, such as: response times and fire department locations. Spatial information was needed to create geographic visualisations.

The following datasets which are used in the Proof of Concept:

Dataset	Provider	Period	Update frequencies	Quality Mechanism
Socio-Demographic data	Centraal bureau van de statistiek (CBS)	2013	Yearly	Yes
Spatial data neighbourhoods	Centraal bureau van de statistiek (CBS)	2013	Yearly	Yes
One parent family	Het Rijksinstituut voor Volksgezondheid en Milieu (RIVM)	2013	None	No
Measured Response Time, Fire department	Nu.nl/data	2010-2012	None	No
Locations of all fire departments in the Netherlands	Rijksuniversiteit Groningen	None	Monthly	Yes
Spatial data Safety regions	Kadaster	None	Monthly	Yes
Construction years	BAG - kadaster	None	Monthly	Yes

2.3 Dataset approaches

All datasets except the data about the construction years are provided as spreadsheet files. Therefore, one should go to the website of the provider and download a file with the data in it. The data about the construction years provided by "Het Kadaster" is provided through a SPARQL endpoint, which makes it possible to request ad hoc data through a web interface.

2.4 Proof of Concept Scenarios

Putting all datasets together with the help of C#.NET and the Dotspatial library, has resulted in the Proof of Concept. The Proof of Concept calculates a risk index for all neighbourhoods in the Netherlands and gives a colour to the neighbourhoods corresponding to the risk index.

2.4.1 Scenario's/Views

The Proof of Concept is able to provide insight into three different scenarios. Each scenario is called a view. In the future the Proof of Concept should provide more scenarios, or should be able to provide flexible scenarios.

2.4.1.1 Dwelling fire risk view

To determine whether a risk index is high or low, statistical percentiles were used. As one can see in **Fout! Verwijzingsbron niet gevonden.** the dwelling fire risks are normally distributed. The percentiles (and corresponding colours) used to determine the ratio in high or low areas:

Percentile	Risk Index	Colour
10th	-48,75	
20th	-46,344	
30th	-44,18	
40th	-41,88	
50th	-39,10	
60th	-36,77	
70th	-34,03	
80th	-28,19	
90th	-17,463	
100th	18,91	

Table 2; Risk Percentiles and corresponding colours

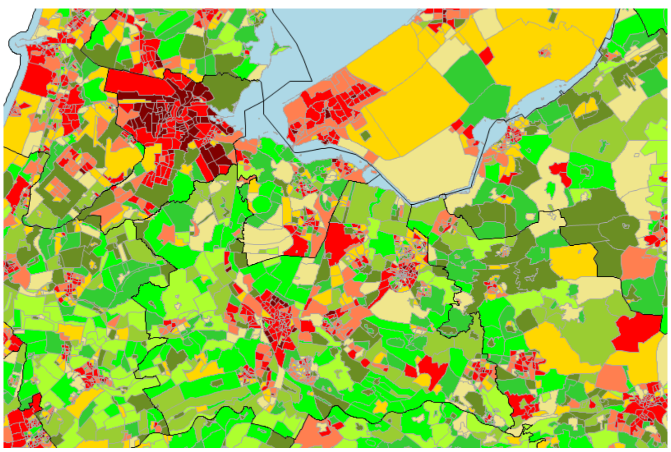

Figure 2; Dwelling fire risk Utrecht (Amsterdam red spot left).

2.4.1.2 Urbanity versus Fire department response time view

In the demographic dataset an attribute which represents the urbanity of the neighbourhood is present. A number between one and five is given. One is very strong urbanity and five is very weak urbanity. From moderate urbanity to very strong urbanity is used in this scenario. The maximum response time for structures with a residence function is 8 minutes. It might be interesting to know, in which regions the urbanity is moderate, strong or very strong and the response time is beyond 8 minutes.

Several neighbourhoods came up with this scenario. The red areas in Figure 3, have a moderate or strong urbanity (1000 up to 2500 addresses per km^2 and the response time of the fire department is over 8 minutes.

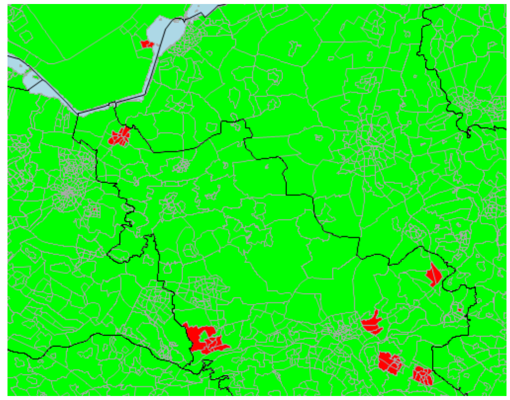

Figure 3; Urbanity vs. Response time Gelderland-Midden.

2.4.1.3 Dwelling fire risk index versus Fire department response time view

The risk index could be used for more scenarios beyond the risk index itself. In the third scenario, the dwelling fire risk index has been in combination with the fire department response time.

Neighbourhoods with a high dwelling fire risk index are reflected in the response time of the fire department. Therefore, the risk indexes from the 90th percentile are shown in red in Figure 4 (high dwelling fire risk) when the response time is beyond the 8 minutes statutory maximum time.

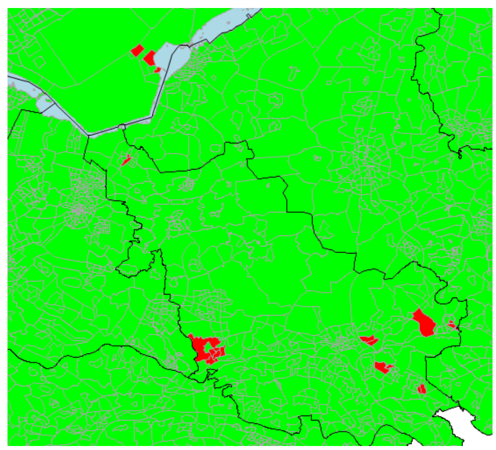

Figure 4; Risk index vs. response time Gelderland-Midden.

Both views (paragraph 2.4.1.2 & 2.4.1.3) show almost the same areas in red. This is mainly caused by the fire department response time. The areas shown in red in both views are problem areas, because in these areas urbanity is strong, fire risk is high and response time is beyond 8 minutes statutory maximum response time.

3. PROOF OF CONCEPT EVALUATION

The statistical relations used have not been validated with more research to help with the interpretation of the statistical relations. The most important question is: 'How do the outcomes of the Proof of Concept compare with reality?'.

3.1.1 Gut feel domain experts

The three views made with the Proof of Concept were verified. In these sessions has been shown that the dwelling fire risk index is the most valuable view, since the response time is needed for the other views and these response times are only available on

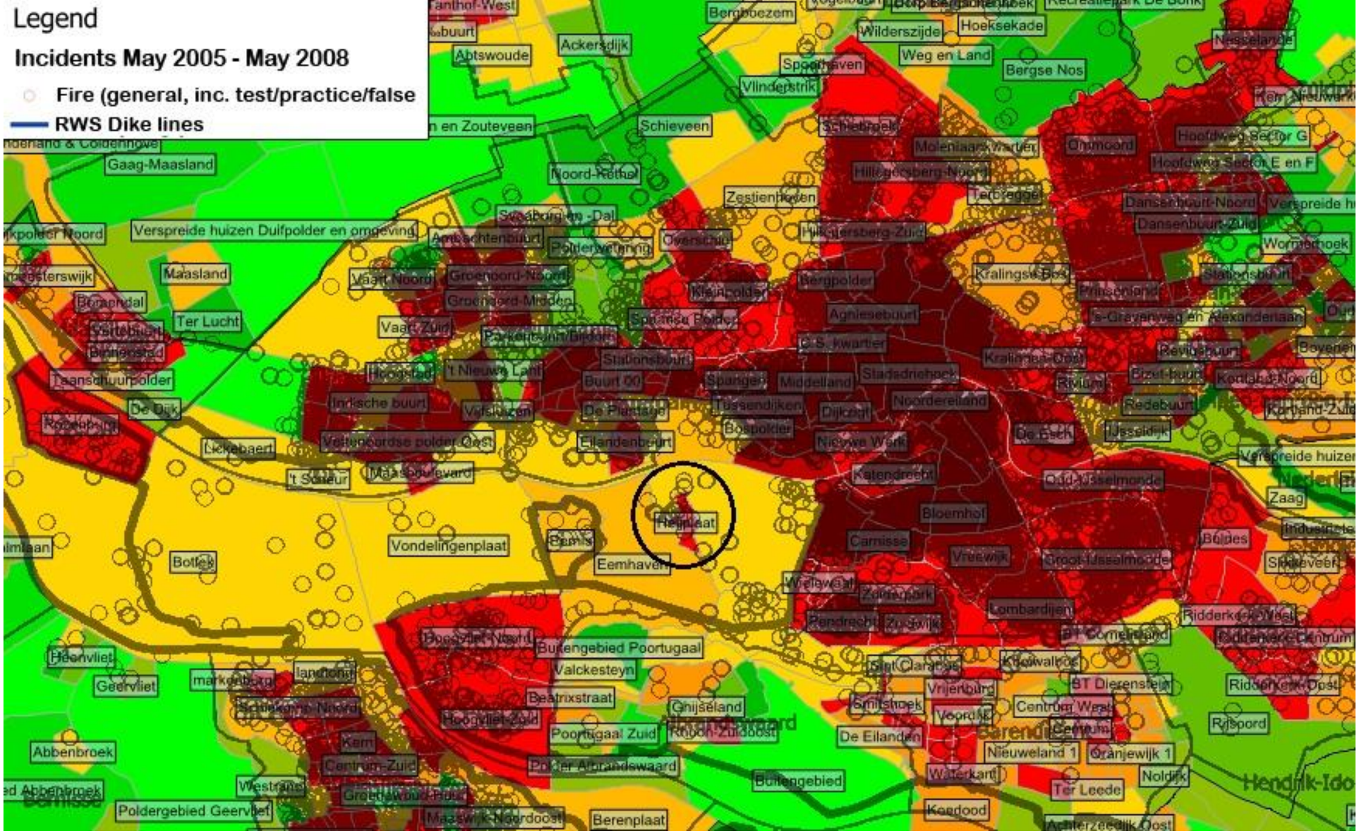

Figure 5; Figure 6; Map with fire incidents placed over the dwelling fire risk index outcomes from the research tool. The Safety region Rotterdam-Rijmond. The area of “Heijplaat” is black circled.

municipality level. In this dataset the median of the response times per municipality has been taken as the mean. Therefore the results are not satisfactory altogether since the interesting results lie in the deviation of response time within a municipality. With a high probability the fire department could tell that the results of the two other views deviate from the reality.

3.1.1.1 Fire risk index

The verification of the dwelling risk index is based on the gut feelings of the fire department in their own region. The fire department has compared the outcomes from the Proof of Concept of their own region with the gut feelings of the fire risks in their own region. From this evaluation it turned out that there were similarities between the gut feelings of the fire department and the Proof of Concept: "*It seems to fit well*". The high risk areas have indeed a higher risk. Although as a test during the evaluation a small high risk area “Heijplaat” (**Fout! Verwijzingsbron niet gevonden.**) was taken which has a high risk, but from their gut feelings the firefighters disagree with this high risk. According to the firefighters this area should have a lower risk. During the evaluation the evaluation panel did not succeed to explain the high risk (“*Probably the amount of immigrants is not true*”). The main statement is: “The risk indexes seem to fit well in general, probably there are some errors and deviations”.

3.1.2 Reference material

During the evaluations, multiple ways were invented to do a more thorough evaluation of the outcomes. In order to verify the outcomes two references have been used:

- Fire risk map created by the Safety region
- Fire incident map with all incidents from 2005 until 2008 (**Fout! Verwijzingsbron niet gevonden.**).

3.1.2.1 Dwelling fire risk map Safety Regions

The safety regions had already created a map to index fire risks themselves. Although, this map has been created in a totally different way than the Proof of concept of this research. The fire risk map of the fire department is based on: fire incidents in the past, neighbourhood scores and building types.

It was hard to compare the outcomes from the Proof of Concept with the fire index map from the fire department. Both maps work with a relative index score. Nevertheless, they were quite comparable with each other. In both maps the inner cities had a higher risk, however outside the inner cities there are deviations.

3.1.2.2 Fire incident map

The Fire Department of Rotterdam-Rijmond compared the outcomes with the incidents from 2005-2008 plotted on a map during evaluation.

In **Fout! Verwijzingsbron niet gevonden.** the incident map is placed over the outcomes from the Proof of Concept. As one can see, the high risk neighbourhoods had more fire incidents in the past than the low risk neighbourhoods. This overlay gives a better comparable view.

All participants in the evaluation agree, that the Proof of Concept provides a good fire risk index. This conclusion has purely been drawn based on the comparison between the incident map and the fire risk map from the Proof of Concept tool. Moreover, “Heijplaat”, the area from the gut feeling evaluation part, appears to have had more fire incidents in the past than the average low risk areas. The high risk assigned to “Heijplaat” appears to be connected to reality.

4. DISCUSSION

4.1 Datasets

4.1.1 Necessary data

All necessary datasets were found matching the factors from paragraph 2.1 to conduct a proper risk calculation. However, the statistical relations from the dwelling fire risk profile ("Handreiking sociaal woningbrandrisicoprofiel") are not validated by other research nor further investigated. Therefore, the statistical relations are not complete (probably there are more factors) and there is no confirmation of the correctness of these factors.

Since the "Handreiking sociaal woningbrandrisicoprofiel" is the only statistical knowledge that is available, no other options were available to use for this research. Improvement of statistical knowledge and improvement of the factors will probably improve the dwelling fire risk calculation in the future.

4.1.2 Individual files

The main idea of this study is to create information by combining open datasets. All datasets used in the Proof of Concept are open datasets available on the internet. Compared to a business intelligence approach in which data warehouses are used, a linked open data approach keeps the data on the side of the provider. In practice it appears that almost all datasets are provided as a spreadsheet file which one has to download from the internet instead of using a SPARQL interface to gather data from the provider. In the Proof of Concept only the data about construction years from BAG were provided through a SPARQL interface. Individual (spreadsheet) files means downloading and storing data yourself, in other words creating a data warehouse. Due to these static datasets, data alignment is harder, since static files contain different file formats which are stored locally. In a dynamic way (i.e. SPARQL) the output (standard data format, i.e. XML and JSON) will be linked and combined with each other on the spot.

4.1.3 Outdated data

Four out of seven datasets used in the Proof of Concept are from 2013 or older. Important to note is that the dataset which contains the demographic data and the spatial data of all neighbourhoods in the Netherlands date from 2013. During the evaluation sessions the outdated data were discussed. During both sessions the fire department participants stated independently that demographic characteristics do not change in a few years. The outdated demographic dataset would not be a problem. However, the government sometimes invests a lot of money in a neighbourhood which could change the demographic characteristics of a neighbourhood. During the evaluation sessions the participants mentioned that the number of times such investments happen are negligible. The demographic dataset is always one to two years outdated, and therefore the obsolescence of this dataset will not cause problems.

4.1.4 Dynamics of datasets

The dynamics of the Proof of Concept would be better, when all data could be requested, calculated and presented the outcomes, instead of downloaded, managed, calculated and presented the outcomes (which is currently the case).

5. CONCLUSION

The necessary datasets to conduct a dwelling fire risk index were found, in some cases the accuracy is a problem. Furthermore, there is a slight accuracy mismatch in the demographic dataset. Most datasets are two years old, this is not a problem. Based on gut feeling all domain experts could verify that the results seemed to be right on the first impression. The comparison of the outcomes with the fire incidents in the past plotted on a map from the fire department, was surprisingly similar. The fire incident map and the outcomes of the Proof of Concept have been compared successfully by a domain expert as well. With the used datasets and the used statistical relations, it seemed to be possible to do a successful dwelling fire risk assessment. The dynamics of the Proof of Concept was not as expected. Since almost all datasets are provided as individual files, it is not possible to gather data, calculate and present the results. The evaluation sessions with the domain experts have given an indication of the correctness of the outcomes. However, it would be a good thing to find more ways to verify the outcomes.

6. ACKNOWLEDGMENTS

I would like to acknowledge those who have helped me to achieve the goals of this study. I would like to thank Jan Wielemaker (VU University Amsterdam) for his guidance and support throughout the research as supervisor. I would like to thank Bart van Leeuwen as well. Bart fulfil the role as co-supervisor, Bart has helped me with his firefight and Linked Open data expertise. I would like to thank the domain experts (Vincent Oskam, Edward Seepers, Rogier Piek & Vincent Baartmans) which has contribute to this research during the evaluation sessions. Without their help it would be impossible to learn about the correctness of the outcomes.

7. REFERENCES

Bauer, F., & Kaltenböck, M. (2012). *Linked Open Data: The Essentials A Quick Start Guide for Decision Makers.* Retrieved from http://www.semantic-web.at/LOD-TheEssentials.pdf

Binnenlandse Zaken. (2011). Besluit veiligheidsregio's. Retrieved from http://wetten.overheid.nl/BWBR0027844/geldigheidsdatum_21-01-2015

Handreiking sociaal woningbrandrisicoprofiel. (n.d.).

Nederland Brandweer. (2014). Business Intelligence. Retrieved from http://www.brandweernederland.nl/service/bvim-0/brandweer-business

Schaap, L. (2013). Nieuwe isolatiematerialen erg brandgevaarlijk. Retrieved from https://www.nvm.nl/nl-NL/Actual/Juni_2013/Nieuwe_isolatiematerialen_erg_brandgevaarlijk.aspx

Stellato, A. (2012). Dictionary, Thesaurus or Ontology? Disentangling Our Choices in the Semantic Web Jungle. *Journal of Integrative Agriculture*, *11*(5), 710–719. http://doi.org/10.1016/S2095-3119(12)60060-4