



Data analysis in the public sector

To what extent do public organizations use data to draw up more efficient and more effective policies? How can data possibly be used to predict, for instance, behavior, crime or fires? Governmental organizations have, perhaps, an even greater responsibility to deal with data in a responsible and secure manner. Certainly when it concerns personal data. What about privacy, and who actually owns data? In this article we try to answer these questions based on findings from studies done by KPMG in 2015 and 2016.



E.N. Berfelo LLM
is a manager at KPMG Management Consulting.
berfelo.elin@kpmg.nl



J.W. Eikema LLM
is a consultant at KPMG Management Consulting.
eikema.jan-wieger@kpmg.nl



A.P.H. de Klerk MSc
is a consultant at KPMG Management Consulting.
deklerk.arien@kpmg.nl



M.H.N.M. Hermans BSc
is writing his dissertation at
KPMG Management Consulting.
hermans.machiel@kpmg.nl

HOW GOVERNMENTAL ORGANIZATIONS USE DATA: AN INTRODUCTION

The volume of data has grown exponentially in recent years and continues to grow each day by more than 2.5 quintillion (10^{30}) bytes ([IBM16]). Bearing in mind the technological trends in the area of the creation and storage of data, it seems that the speed at which the volume of the stored data grows is not going to come to a standstill anytime soon. This has led to higher investments in the capacity needed to analyze data and also in the development of predictive models. The latter is also known as 'data analytics'.

The concept of data analytics has also been around for a number of years and is characterized by great promises. For example, data, together with the potential to analyze and use it, was even described as the oil of the twentyfirst century, to point out its value. An increasing number of organizations seem be looking into its potential in an effective and efficient manner. This also applies to governmental organizations.

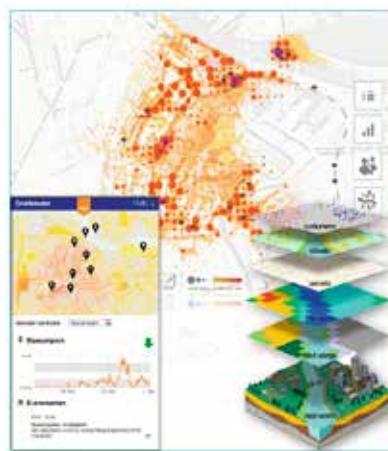
Governmental organizations have acquired more insight using data analyses. They can now operate in a more data-driven and information-led manner. In the first section we examine a number of possible applications of data in the public (security) sector. At the same time, this development also contains a challenge, because while the government appears to use data more effectively during its activities, there is also a need to guarantee the responsible and secure use of this data. In the second section of the article we will discuss the conditions for responsible data use.

A more secure society: from analysis to prediction

Following the meetings we had in 2015 with representatives of various public organizations, we have defined a number of ways for applying data analyses. Roughly speaking, we can see a division here. On the one hand it concerns the analysis and processing of data and on the other hand making predictions on the basis of this data. These applications are connected: when data is correctly processed and analyzed, it is possible to make predictions about behavior and risks based on this data. This study concerns an initial assessment. In 2016, KPMG carried out a broader research into the use of data and information by the government. The results of this investigation are presented separately in Figure 1.

An example of how data is analyzed and processed within the area of security is the Hansken system ([NFI17]). The Netherlands Forensic Institute (NFI) developed this system in close cooperation with the police to

Figure 1. Pressure gauge during the coronation of Willem-Alexander.



Governmental organizations have gained more insight on the basis of data analyses, and can now work in a more data-driven manner

make the ever-increasing volume of digital traces during criminal cases quickly and easily searchable. The system organizes digital information so that detectives can retrieve relevant information from the ever-increasing flow of digital data more quickly. This not only enables the efficient use of resources such as time and manpower, but it also positively influences the quality of work. Below we list three possibilities for which data analysis can be used.

1. Predicting behavior

Making predictions based on data is a careful balancing act and most predictions are mainly carried out in buildings or defined locations, such as the Amsterdam Arena ([Jeur17]). There are however also examples in which behavior was predicted on the basis of public space data. In Amsterdam, for example, during the coronation of

King Willem-Alexander, a number of data sources were combined to produce a sound data analysis (see Figure 1). This analysis was subsequently used to make predictions about the number of visitors. This resulted in the so-called 'crowd control' of the police. With the help of among other things data from WiFi spots, a real-time picture was created of crowds and their movements in the Dutch capital. Based on the constructed pressure gauges, the police could analyze the busy and less busy locations and make predictions about where any congestions could occur. Thanks to these predictions it was possible to anticipate where and to what extent additional police deployment was required.

2. Predicting crime

Another often-mentioned application of data in the area of security is the so-called predictive policing. An initiative in this area has also been started in the Netherlands: the Crime Anticipation System (CAS) ([Will16]), developed within the former Amsterdam-Amstelland police force. The essence of the system is that information about crimes, such as domestic burglaries or pickpocketing, is linked to around 250 characteristics such as social structure, number of gambling halls and that historical data on burglaries and break-ins are digitally mapped. Using an algorithm, it is possible to make predictions and adjust the deployment of the police accordingly. It turned out that 36 percent of the burglaries could have been predicted; a good reason to start a pilot in four base teams outside Amsterdam in the fall of 2015. This pilot has been proven successfully and now the system will be implemented nationally. The ambition is to have all 168 police teams participating by the end of 2017.

3. Predicting fire

Making predictions based on data is not just done by the police. The fire brigade also connects various data sources. This is how they can make statements about possible connections between prevention, incidents and the surroundings. One of the application possibilities of such predictions is, for instance risk assessments that can be improved by linking data about the issuance of permits to incidents that have already happened. In an international context, we like to mention FireCast, an algorithm that is used by the New York City Fire Department (FDNY) to make fighting and preventing fires more effective and efficient. The possibility of a fire is predicted on the basis of 7,500 risk factors, including specifications, complaints and violations. The buildings are then classified for inspections and assigned to a fire station on the basis of a risk score.

CONDITIONS FOR THE RESPONSIBLE USE OF DATA

The previous section mentioned a number of practical examples in which the use of data played a crucial role. However, if you take a closer look at data analyses and predictions, you will encounter a complexity of legislation combined with issues of quality and transparency.

Quality of data

Due to the advance of data analysis and because data analyses are becoming more complex, we see that - besides the possibilities - the risks and the impact of incorrect data analyses increase as well. It is therefore crucial that all data used is of high quality. When companies and governments act on incorrect information that is obtained from bad data, this can be extremely ineffective ([AUSG13]). When we consider data to be the oil for the performance of organizations, we see a striking analogy. When oil is extracted from the ground, this raw product can be disastrous for the engine and performance of a car. To prevent this, a large number of steps are taken to make this oil suitable as fuel. This oil can only be sold at the gas station when it has undergone sufficient processing. The same applies to data. Only when data is sufficiently refined, it can be used as fuel for organizations. This means among other things that data is filtered before it is used for carrying out analyses and making predictions.

Sufficient expertise

Along with the requirements that are laid down for the availability of good quality data, high requirements are also placed on the professionals who analyze data. It would be a misconception to assume that data analysis is nothing more than number crunching that, with the purchase of some tools, is easily implemented in the organization's procedures. After all: 'a fool with a tool is still a fool'. What is important, is that there are sufficient well-educated data scientists who operate these tools and who are able to interpret the results in the correct context. If this requirement is not met, the potential of data cannot be fully utilized.

Application of legislation on privacy and data ownership

In addition to the quality of data and the knowledge of data scientists, a lot of attention is paid to legislation regarding the use of data in the public domain. The Scientific Council for Government Policy (WRR) recognizes the importance of big data to promote security ([Hirs16]), but also mentions several important conditions that must be met when organizations experiment with big data in the area of security:

“Big Data offers many opportunities to promote security. To take advantage of these opportunities there must be sufficient room to further explore the added value of Big Data. Big Data also has risks. Experimenting and making use of Big Data must therefore always be accompanied by sufficient guarantees in which the protection of privacy and personal data, the prohibition of discrimination, transparency and the reliability of both the data as well as the analytical methods are central. This way, sufficient public trust can be achieved among citizens in the way the government uses Big Data possibilities in the area of security.” ([Hirs16]).

The conclusion of the report corresponds with our observations that are related to privacy and data ownership. First, the privacy of citizens must be sufficiently guaranteed ([Koor15]). The regulations for this are laid down in the Dutch Personal Data Protection Act (Wbp), a framework of the European Privacy Directive. The principles of the Wbp are the legitimacy of processing data and providing information to those whose personal data are used.

However, the European Privacy Directive has been revised, because it was first drawn up at the time when the Internet was still relatively underdeveloped ([APG17]). As of May 25, 2018, there will be a General Data Protection Regulation (GDPR) ([APG16-1]) with a separate directive for data protection during investigation and prosecution ([APG16-2]). This directive includes regulations that are processed for the prevention, investigation and prosecution of criminal offences or the execution of criminal penalties. This directive is, among other things, in the interest of improved cooperation between the investigation services ([APG16-2]).

The GDPR also requires and facilitates the implementation of national legislation on various points. On the one hand, national legislation is responsible for implementing the provisions regarding the national supervisory body and, on the other hand, for using the room provided by the regulation for further interpretation of the provision of the regulation under national law. Policy neutrality is the leading principle, which means that the existing law is maintained, unless the General Data Protection Regulation makes this impossible ([Over16]).

The objective of this revision is that citizens get more possibilities to better manage their personal data. It is especially important for governmental organizations to be transparent about the use of personal data and to provide citizens with a choice to make their personal data available. Obviously this freedom of choice is restricted when it concerns investigation and prosecution.

Strengthening supervision

Along with the application of legislation within the framework of privacy and data ownership, it is necessary that the supervision within this framework is strengthened. The Ministry of Security and Justice and the Personal Data Authority have therefore started a process to ascertain the consequences of strengthening powers for the capacity and budget of the Personal Data Authority. A significant expansion of the Personal Data Authority is expected due to the strengthening of supervision. The government also wants to investigate whether it is desirable to extend the possibilities to have applications in the field of data analysis examined by a court. How supervisors and judges can be given sufficient insight into the applied algorithms and analytical methods for decision-making on the basis of data analysis, where it has a significant impact on citizens, is included in this investigation ([MVJ16]). KPMG has argued earlier for independent supervision of the application of data in systems ([Röel17]).

Transparency about the use of data

Furthermore, we found that it is especially important for organizations to be transparent about what they do with the data and whether the privacy of citizens is taken seriously enough. It may also occur that organizations make use of insufficiently thought out applications of data analyses. For example, one of the pitfalls is that during the analysis of data the difference between correlation on the one hand and causality on the other is not recognized. An example of this misconception is observed in New York, where a crime wave suddenly came to an end. This development was initially ascribed to the zero-tolerance policy that the police carried out in the area of crime, with heavier jail sentences and the resulting reduction in the illegal possession of weapons. It later turned out however that it was not the zero-tolerance policy, but the fact that abortion was made available for parents living in poverty that was the primary factor explaining the decrease in crime figures ([Bre16]).

**In the end, everyone
wants that predictive
crystal ball**

For every organization, but especially for organizations in the public sector it is necessary to be transparent about the use of personal data and to enable citizens to - as far as possible - make their own choice regarding the availability of their personal data. It is also required to pay special attention to open communication and to have a very active attitude in forming public opinion, especially now that data analysis is playing a larger role. The WRR report about big data in a free and secure society is a valuable contribution to this social discussion.

CONCLUSION

Despite the presence of national and international possibilities to apply data analyses and the exponentially increasing volume of data, the public sector does not seem to be able to keep up effectively. The lack of a sufficient number of data scientists and available resources, among other things, contribute to this. The use of good quality data analyses offers the possibility to make predictions about, for instance, behavior. This in turn can be used to carry out specific tasks more efficiently and effectively. The role of government is, however, not just restricted to data usage, it also takes on the role of supervisor and regulator regarding the use of data. This is to secure the privacy of the citizens and guarantee the quality of the available data. This makes it appear as though the government is imprisoned in a web of conflicting interests and does not have the potential to fully utilize data.

How data-driven is the Dutch government in 2016?

In 2016, KPMG carried out research into the status of governmental organizations regarding the use of Big Data to be able to work on the basis of intelligence. Working in a data-driven manner, whereby sources of information are combined and analyzed to come to improved decision-making, is slowly gaining ground with government organizations. This way of working enables the public sector to identify trends and patterns. The insights gained from this lead to a different way of management, better substantiated choices in policy and an improvement of the services provided to citizens. Achieved performances can also be evaluated on the basis of these insights.

The respondents of this research came from thirty different governmental organizations. The conclusion of the research was that the government has sufficient ambition in the area of data-driven working, but lacks expertise and budget.

The KPMG study shows that around 50% of governmental organizations is already experimenting with this way of working. And 40% is aware of the possibilities that this way of working offers, although they haven't applied it yet. Of the organizations that have discovered the path to data-driven working, only 20%

indicate that this occurs on a structural basis. Almost half of the investigated organizations indicate that within a period of one to two years they will implement (elements of) data-driven working to be able to better achieve their strategic objectives.

For many governmental organizations the lack of qualified personnel appears to be an obstacle in getting data-driven working off the ground. More than 60% indicate having insufficient data scientists available to make data-driven working possible. Slightly more than 10% of the investigated organizations indicate that they have enough data scientists available. Almost 20% of the respondents indicate that at this time they have nobody available that can carry out data analyses on a professional basis.

The availability of adequate resources also forms an important bottleneck. Only 30% of the respondents indicate to have sufficient budget available to be able to work completely on the basis of data and data analyses. Almost half of the investigated organizations think that they have insufficient resources available to take the necessary step. In order to still be able to make use of the potential of all the information, these organizations should look for partnerships with other governmental organizations and external parties.

Figure 2. Results of the KPMG study.

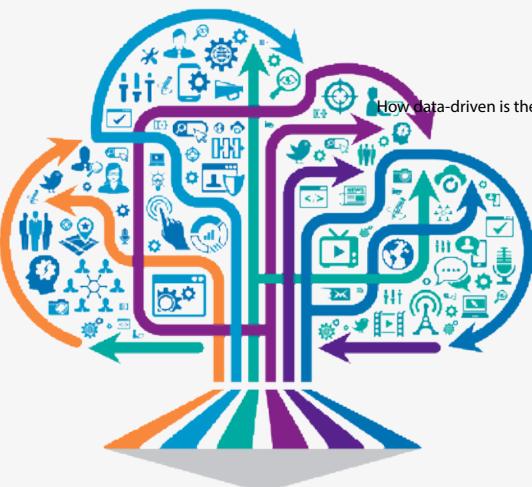
From information to insight

The amount of data has increased exponentially in recent years. Data is the new raw material of organizations. More and more organizations are therefore investing in analyzing ever-expanding amounts of data and the development of predictive models. This provides for better-substantiated policy choices, a greatly improved service offering, higher efficiency and improved insight. The change that the availability of information entails has great consequences for the management and organization of governmental organizations. It also gives the government a responsibility to deal with this in a proper and safe way. It is not easy to identify the extent to which the Dutch government operates in a data-driven manner. To gain more insight, KPMG conducted a study in November 2016.

Huge increase in amount of data

By 2015, a total of more than 3.4 Zetabyte (one Zetabyte is a 1 with 21 zeros) of data is collected. By way of illustration: in 2003 we all produced 'only' five Exabyte (one Exabyte is a 1 with 18 Zeros).

The expectation is that the amount of data will still double every 18 months. Data production will increase to around 44 Zetabyte by 2020.



The Hague

Data is available to the government to gain insight, make better-substantiated choices and improve services to citizens

47% wants to work in a data-driven manner to achieve strategic goals within 1 to 2 years.

To achieve strategic goals, my organization should work in a data-driven manner within:

- 1 to 2 years: 47%
- 3 to 5 years: 43%
- Not necessary: 7%
- Other: 3%

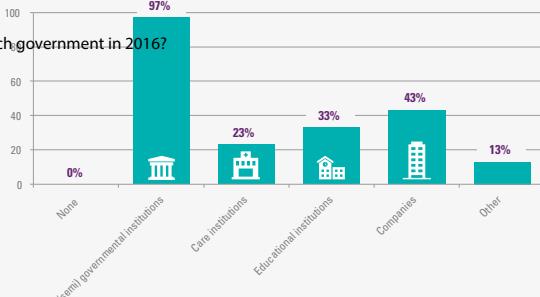
What does data-driven mean? It means we combine and analyze different sources of information in a coherent manner. This is how trends and patterns can be identified. The insights gained lead to different ways of management, better-substantiated policy choices and an improvement in services towards citizens. Performance can also be evaluated based on these insights.

KPMG Study

To gain more insight into the data-driven way of operating, KPMG conducted a study in the public sector:

- What?** To what extent does the public sector operate in a data-driven manner?
- How?** Survey sent to organizations within the national government
- Who?** Respondents from 30 different organizations: from policy advisor to Secretary General

Which external parties does your organization exchange data with?

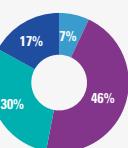


External Party	Percentage
None	0%
(semi) governmental institutions	97%
Care institutions	23%
Educational institutions	33%
Companies	43%
Other	13%

97% exchanges data with other (semi) governmental organizations
43% exchanges data with companies

Only 30% has sufficient budget to operate in a data-driven manner

To what extent does your organization have sufficient budget to make operating in a data-driven manner possible?



Response	Percentage
There is no budget for this within the organization	17%
We have a budget for this, but it is not sufficient	46%
There is sufficient budget for this	30%
Other	7%

50% experiments with operating in an data-driven manner

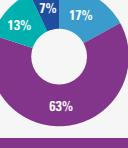
To what extent does your organization already operate in an data-driven manner?



Response	Percentage
There haven't been any initiatives yet	20%
Ad-hoc: we experiment pro-actively whenever we have a problem	20%
Lab: we experiment pro-actively	20%
Staging: we analyze our data systematically; data analysis is integrated in our systems	50%
Other	10%

Only 13% has employed sufficient data scientists to be able to operate in a data-driven manner.

To what extent does your organization have the right people (data scientist) to operate in a data-driven manner?



Response	Percentage
There are no data scientists within the organization	17%
We have a few data scientists but not sufficient	63%
We have sufficient data scientists	13%
Other	7%

References

- [APG16-1] Autoriteit Persoonsgegevens, *General Data Protection Regulation*, Autoriteitpersoonsgegevens.nl, https://autoriteitpersoonsgegevens.nl/sites/default/files/atoms/files/verordening_2016_-679_definitief.pdf, 2016.
- [APG16-2] Autoriteit Persoonsgegevens, *Directive for data protection during investigation and prosecution*, Autoriteitpersoonsgegevens.nl, https://autoriteitpersoonsgegevens.nl/sites/default/files/atoms/files/richtlijn_2016_-680_definitief.pdf, 2016.
- [APG17] Autoriteit Persoonsgegevens, *Europese privacywetgeving*, Autoriteitpersoonsgegevens.nl, <https://autoriteitpersoonsgegevens.nl/nl/onderwerpen/europese-privacywetgeving>, 2017.
- [AUSG13] Australian Government, *The Australian public service big data strategy*, Australian Government, Department of Finance and Deregulation, <http://www.finance.gov.au/sites/default/files/the-australian-public-service-big-data-strategy-archived.pdf>, 2013.
- [Bret16] Willem Brethouwer, *Big data analyseren? Vermijd deze 3 beruchte valkuilen*, Frankwatching.com, 2016.
- [Hirs16] Ernst Hirsch Ballin, Dennis Broeders & Erik Schrijvers, *Big data in a free and secure society*, Wetenschappelijke Raad voor het Regeringsbeleid, <https://www.wrr.nl/publicaties/rapporten/2016/04/28/big-data-in-een-vrije-en-veilige-samenleving>, 2016.
- [IBM16] IBM, *What is big data?*, IBM.com, <https://www-01.ibm.com/software/data/bigdata/what-is-big-data.html>, 2016.
- [Jeur17] S.E.J. Jeurissen en N.L. Martijn, *Governing the Amsterdam Innovation Arena Data Lake*, Compact 2017/1.
- [Koor15] Ronald Koorn e.a., *Big data analytics & privacy: how to resolve this paradox?*, Compact 2015/4.
- [MVJ16] Ministerie van Veiligheid en Justitie, *Kabinet ziet kansen Big Data om veiligheid te bevorderen*, Rijksoverheid.nl, <https://www.rijksoverheid.nl/actueel/nieuws/2016/11/11/kabinet-ziet-kansen-big-data-om-veiligheid-te-bevorderen>, 2016.
- [NFI17] Nederlands Forensisch Instituut, *Hansken*, Forensischinstituut.nl, https://www.forensischinstituut.nl/Forensisch_onderzoek/Digitale-en-biometrische-sporen/producten_en_diensten/hansken.aspx, 2017.
- [Over16] Overheid.nl, *Uitvoeringswet Algemene verordening gegevensbescherming*, Internetconsultatie.nl, <https://www.internetconsultatie.nl/uitvoeringswetavg>, 2016.
- [Röel17] Albert Röell, *Onafhankelijk toezicht nodig op het toepassen van data in systemen*, Financieel Dagblad, <https://fd.nl/opinie/1187276/onafhankelijk-toezicht-nodig-op-het-toepassen-van-data-in-systemen>, 13 februari 2017.
- [Will16] Dick Willems, *Het criminaliteitsanticipatiessysteem - predictive policing bij de Nederlandse politie*, Big Data Expo, <https://www.slideshare.net/BigDataExpo/politie-dick-willems>, 2016.

About the authors

- E.N. Berfelo LLM** is a manager at KPMG Management Consulting and has more than six years of experience with change management and innovation in organizations in the public sector. She recognizes the considerable added value of the use of data for public (security) organizations and society as a whole.
- J.W. Eikema LLM** is a consultant at KPMG Management Consulting and is involved in strategic issues and transformations in the public sector.
- A.P.H. de Klerk MSc** is a consultant at KPMG Management Consulting and is involved in innovative solutions for organizational issues. He has a background in technical public administration and knowledge management and has done research into the use of data and information for among others the European Commission and the United Nations.
- M.H.N.M. Hermans BSc** is writing his dissertation at KPMG Management Consulting, in which he specifically focuses on the justice and security sector. The Public Governance master gives him insight into several trends and strategic issues that have a large impact on the public sector.