

# SMALL

**BIG DATA  
CONGRESS**



**TNO** innovation  
for life

2015

› COLOFON

This publication has been created as part of TNO's Small Big Data Congress that took place on the 3rd of March 2015.

Organized & Edited by:

**Wietske Koers**

wietske.koers@tno.nl

Strategic Business Analysis

**Freek Bomhof**

freek.bomhof@tno.nl

Media & Network Services

## FOREWORD

NOWADAYS DATA IS EVERYWHERE AND IT IS RAPIDLY BECOMING BUSINESS' BIGGEST RESOURCE. THROUGH NUMEROUS SYSTEMS AND DEVICES, A HUGE AMOUNT OF DATA CAN BE COLLECTED. JUST THINK OF THE MANY SENSOR SYSTEMS, SUCH AS SURVEILLANCE CAMERAS, LOOPS IN THE ROAD, SMARTPHONES OR THE DIGITAL MARKERS WE CREATE OURSELVES BY POSTING PHOTOS ON FACEBOOK OR COMMENTS ON TWITTER.

'Big Data' has become a big thing and will change our world. In practice, however, organizations are often struggling with Big Data on multiple fronts: from the IT and governance structures to manage it efficiently, via analytics to make sense of it, to business models that create value in our everyday lives.

It is TNO's ambition to advance business and society by developing technologies for applications that collect data and unlock its valuable insights in an intelligent way. To realize this ambition, TNO investigates which parties play a role, which are the most promising business models, what technologies and standards are required, how interoperability can be achieved and how to deal with semantics. This brochure presents a small sample of research that TNO conducts on Big Data topics. It is striking

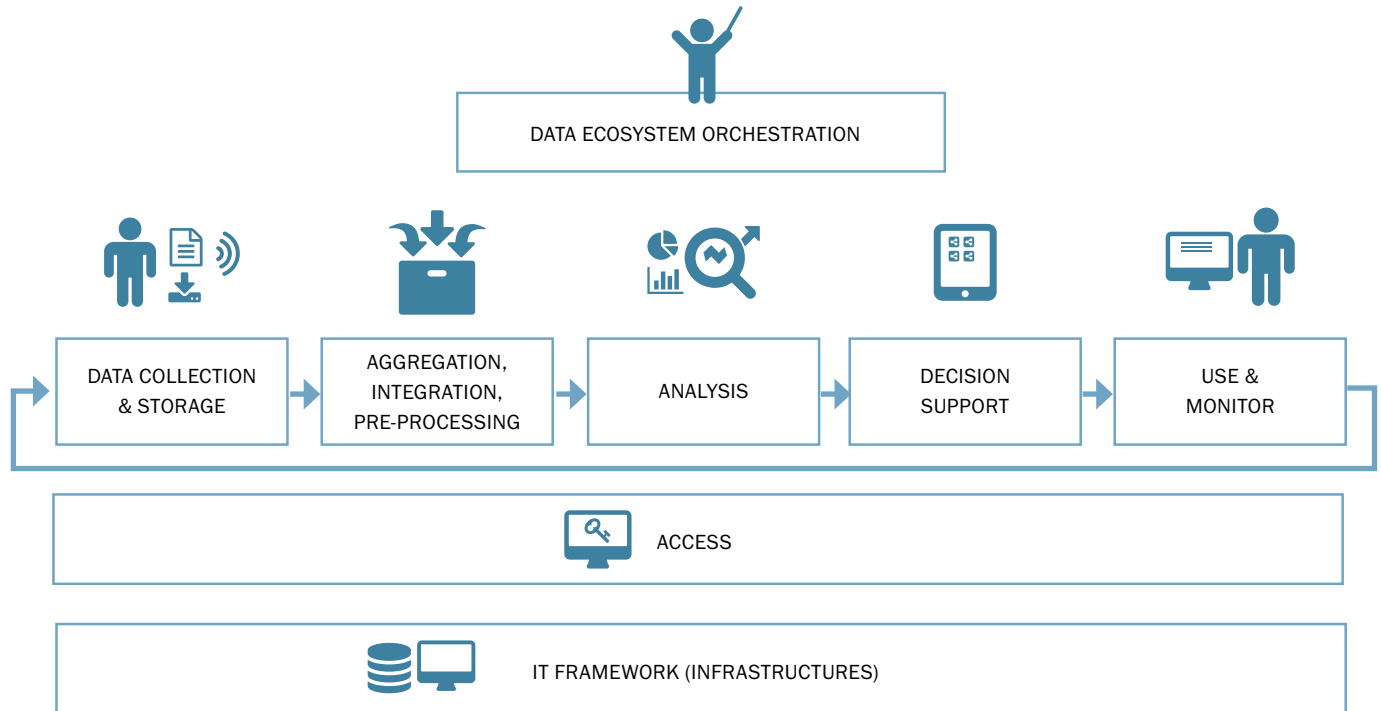
to see how Big Data creates many new linkages between the societal domains TNO is active in. Therefore, I express the expectation that data-driven innovation will have the same connecting role in society as well!



**Henk-Jan Vink**

Program Director Networked Information

# FROM BIG DATA TO VALUE



# SMALL BIG DATA CONGRESS 2015

9	BIG DATA, PRIVACY AND GOVERNANCE	31	NDW TRAFFIC DATA FUSION PILOT
11	COMMON SENSE	33	LINK(ED) DATA TO GIVE MEANING TO BIG DATA
13	STREAMING SENSOR DATA PROCESSING WITH SENSORSTORM	35	HUMAN FACTORS IN BIG DATA
15	DATA DRIVEN PREDICTIVE MODELING FOR SMART DAIRY FARMING	37	DATA SHARING IN NUTRITION (DASHIN)
17	STORMCV	39	MEASURING MOVEMENT AND INTERACTION BETWEEN CHILDREN IN SCHOOLYARDS THROUGH INTELLIGENT VIDEO AND DATA ANALYSIS
19	BIG DATA VISUALIZATION	41	DATA-DRIVEN INNOVATION AS A STRATEGIC CHALLENGE
21	VALUE CHAIN COMPLEXITY, ARCHITECTURE AND INTEGRATION CHALLENGES: DIKE DATA SERVICE CENTRE (DDSC)	43	GOOSE: SEMANTIC SEARCH IN SENSOR DATA
23	BIG DATA STORAGE CHALLENGE: HIGH RESOLUTION VIBRATION MONITORING IN 200 BUILDINGS	45	SMART GRID RESEARCH WITH BIG DATA TOOLS
25	MULTI-AGENT MODELS FOR TREND PREDICTION IN SOCIAL MEDIA	47	MONITORING WORK & EMPLOYMENT BY TNO
27	DIAMONDS: DATA-WAREHOUSE INFRASTRUCTURE FOR ALGORITHMS MODELS AND ONTOLOGIES TOWARDS NOVEL DESIGN AND SAFETY	49	NEWS GENIUS: YOUR CUSTOMIZED VIDEO NEWS BROADCAST
29	SIGNAL INTELLIGENCE FOR EMERGING FOOD SAFETY RISKS	51	MULTI-USE BIG DATA IN THE AMSTERDAM CANALS
		53	TRENDS IN VEHICLE FLEETS
		55	USING BIG DATA ANALYTICS TO IMPROVE THE USER EXPERIENCE
		57	BIG DATA WITH ROUTE RADAR
		59	LEARNING STRATEGIES IN DIGITAL LEARNING ENVIRONMENTS
		61	ADVANCED ANALYTICS ON LOGS
		63	FOUNTAIN
		65	(OPEN) DATA GOVERNANCE – DEMAND AND FIRST SOLUTIONS
		67	CANCER AWARENESS CAMPAIGNS ON TWITTER

LOCATION



# ALMERE BIG DATA VALUE CENTER

THE BIG DATA VALUE CENTER IS AN OPEN INNOVATION PLATFORM, LOCATED IN ALMERE, THAT AIMS TO SHOW HOW BIG DATA APPLICATIONS PROVIDE VALUE. WE BRING KNOWLEDGE, RESOURCES, PEOPLE AND ORGANIZATIONS TOGETHER BECAUSE WE BELIEVE THAT IN THE NEW FIELD OF DATA-DRIVEN INNOVATIONS, A MULTIDISCIPLINARY APPROACH WORKS MUCH BETTER THAN THE CREATION OF ISOLATED SOLUTIONS.

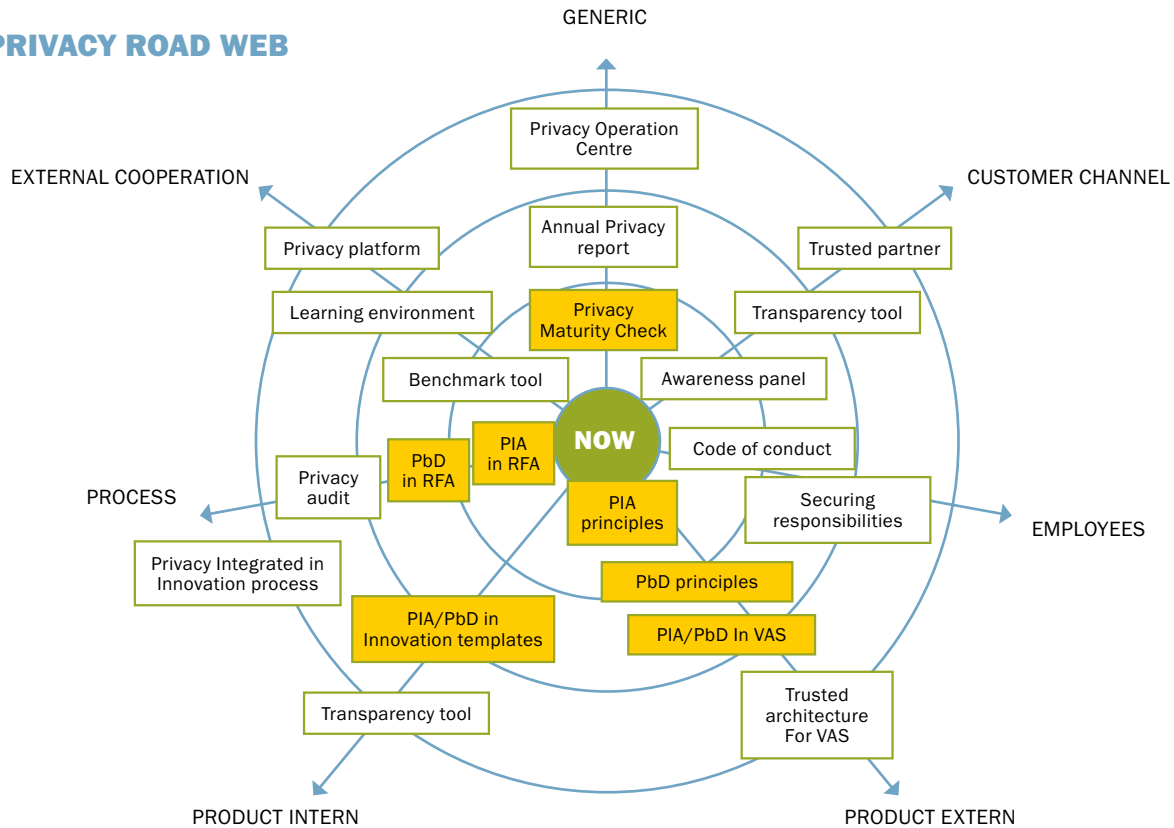
This is illustrated by the slogan 'collaboration is the new competition'. The BDVC wants to bring together science, industry, education, and government and focuses on the practical applications of Big Data.

A wide range of activities is organized by the BDVC and its partners. Examples are: meetings aimed at knowledge transfer, trainings, hands-on sessions where real datasets are analyzed, the development of a huge repository of open proprietary data, challenge assignments where tough data questions are answered in a pressure cooker. Besides these events, the BDVC hosts a number of young data startup companies, and provides working places for partners' employees.

The Big Data Value Center is an initiative of the Economic Development Board Almere, in cooperation with TNO, SURFsara, Netherlands eScience Center, Amsterdam Economic Board and Economic Board Utrecht.



## THE PRIVACY ROAD WEB





# BIG DATA, PRIVACY AND GOVERNANCE

Governance of data is a key requirement to make innovations in Big Data – which often include personal data – flourish and find their way in society. Individuals have a right to privacy and the legal framework sets strict requirements which have to be met before personal data can be processed. These requirements will become even stricter with the forthcoming revision of the EU legal framework. There will be more focus on accountability. Governance of data, in terms of access, use, editing, etc., can help facilitate the use of data, while still being compliant with the law.

An integral approach for businesses and other initiatives is recommended, making privacy part of the innovation process. From this perspective privacy can even offer an opportunity for data driven innovation, instead of a barrier. In order to facilitate such an approach, TNO has developed a strategic road web which helps identifying actions to take in the diversity of processes and communication lines of an organization. The action lines and concrete actions can be tailor-made for each organization and build up to an increasing level of privacy maturity.

## CONTACT

**Arnold Roosendaal**

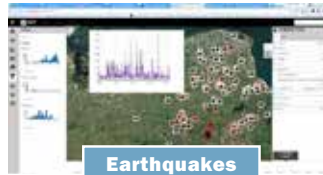
arnold.roosendaal@tno.nl

Strategy & Policy

## ACCESS



## OPEN SOURCE WEB FRAMEWORK FOR VISUALIZATION AND INTERACTION WITH (BIG) DATA



ABOUT

TECHNOLOGY



PROJECTS

ROADMAPS

📍 Geospatial

W3 Web standards

🗺️ Interactive Dashboards

📦 3D

🔍 Analysys

🐱 Open Source

📍 Sensor Data

🔗 Linked Data

# COMMON SENSE

Common Sense is a framework for visualization and interaction with (big) data. It is being developed within TNO and used for both commercial projects and research programs. Common Sense allows users to create dashboards to quickly visualize and analyze geospatial and sensor data. Using emerging web technologies, large datasets are available for exploration from all kind of devices.

Common Sense is an open source project. We are working together with IT companies and Universities to expand the capabilities and applications of this framework.

## CONTACT

**Arnoud de Jong**

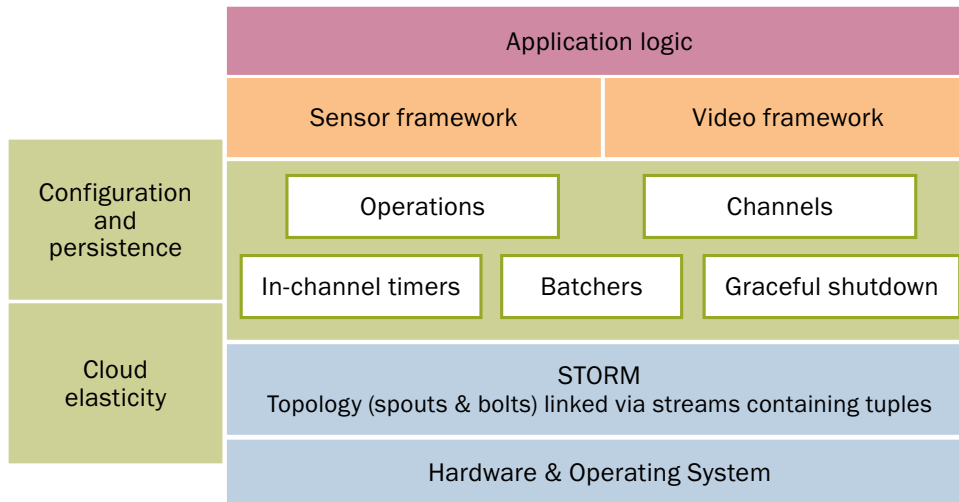
arnoud.dejong@tno.nl

Business Information Services

## DECISION SUPPORT



STREAMING SENSOR DATA PROCESSING WITH SENSORSTORM



# STREAMING SENSOR DATA PROCESSING WITH SENSORSTORM

Large scale modern sensor systems produce high volume streaming data. Monitoring infrastructures such as dikes, bridges, highways or valuable assets such as cows, is done by using a wide array of sensors from simple temperature sensors, via complex milk quality sensors, to video streams. Sometimes the volume and throughput of these data streams is so large, that storage is not a valid option. When near real-time results are required, preprocessing and/or analyzing the data in a streaming manner is a very good solution. TNO has developed a specific software application for this called SensorStorm. It offers a highly scalable and robust platform, which can process high volume streaming sensor data, from simple sensors to video. Special components for sensor data analysis can be developed by the customer. How the data must flow through a particular combination of such components can be specified in a topology. This topology can then be uploaded to the platform, running on your own laptop, on several of your own servers or on a huge number of machines hired in the cloud.

## MORE INFORMATION

SensorStorm is based on Apache Storm and SensorStorm and is released in the open source on GitHub: <http://github.com/sensorstorm>.

## CONTACT

**Bram van der Waaij**

[bram.vanderwaaij@tno.nl](mailto:bram.vanderwaaij@tno.nl)

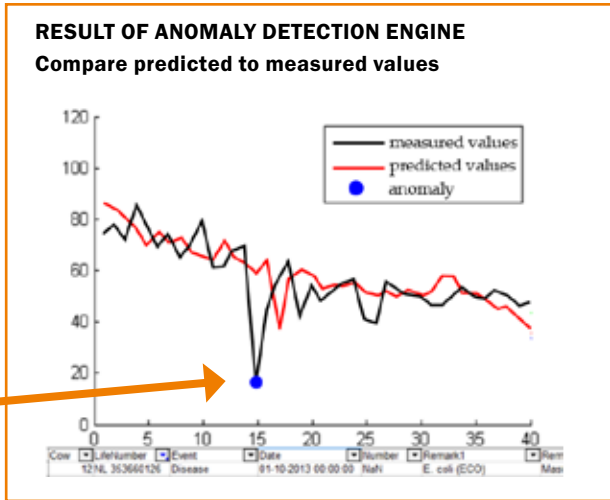
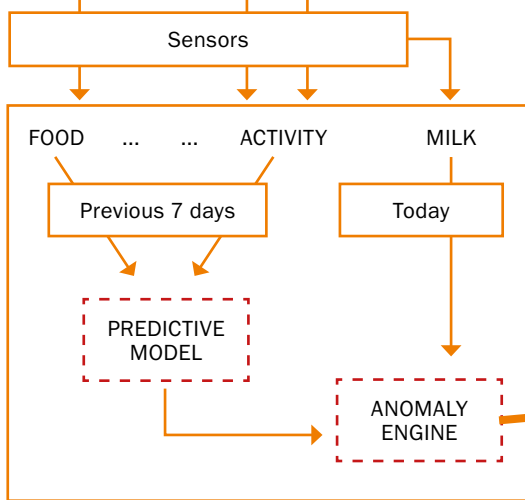
Business Information Services

PRE-PROCESSING | ANALYSIS





- No domain (expertise) knowledge required
- Automatic identification of important features and their relevance
- Reliable prediction of 2 days of milk yield using 7 days of history
- Application to anomaly detection



# DATA DRIVEN PREDICTIVE MODELING FOR SMART DAIRY FARMING

With the Smart Dairy Farming (SDF) 1.0 project we have been collecting data on a daily basis about hundreds of cows for the last two years. Such data includes the milk production, amount of food intake, the percentage of time the cow is active, and how much it ruminates. Can we learn any correlations from the data in the absence of domain knowledge?

In a recent project we showed how data driven (i.e. no domain knowledge) predictive analysis techniques can be used to make some sense out of this large volume of data. The research question was: Can we reliably predict the milk production for the next couple of days for every cow, by using various data collected in the previous days?

This is a very important topic for big farms, as it can give a lot of insight about the performance of the cows. For instance, it can indicate which cows are not conforming to their usual performance, suggesting that they might have become unhealthy. Moreover, it can give indications about which are the characteristics of the highly performing cows. In an automatic way, we identified the most effective characteristics and used them to compare multiple linear regression and support vector regression techniques to

predict the milk yield of each cow. We also demonstrated how this modeling technique can be used to identify an anomaly in the behavior of a cow. In SDF 2.0 we will investigate further for what purposes data driven predictive modeling can be used (e.g. prediction of food intake and early prediction of upcoming illnesses).

## CONTACT

**Bram van der Waaij**

bram.vanderwaaij@tno.nl

Business Information Services

## ANALYSIS







# STORMCV

## **APACHE STORM + OPENCV = LARGE SCALE DISTRIBUTED IMAGE AND VIDEO ANALYSIS.**

StormCV is an extension of Apache Storm specifically designed to build distributed computer-vision pipelines, which can be executed on a large number of servers. This platform supports reading and writing of video, manipulation of frames/images, feature extraction and much more. Most importantly it is very easy to scale up (or down) when the amount of video to be processed changes. A proof-of-concept application which detects and counts cars within 10 live streams was scaled up to 20 streams within minutes. The platform was tested up to 160 cores on Amazon EC2, analyzing over 18.000 images per minute (calculating SIFT, face detection and color histograms). Since StormCV is built on top of widely used libraries such as FFMPEG, OpenCV and Apache Storm it very future proof.

### MORE INFORMATION

Is large scale video or image processing a struggle you are familiar with? Go check out the StormCV project on Github: [https://github.com/sensorstorm/StormCV!](https://github.com/sensorstorm/StormCV)

### CONTACT

#### **Corné Versloot**

corne.versloot@tno.nl  
Media & Network Services

#### **John Schavemaker**

john.schavemaker@tno.nl  
Media & Network Services

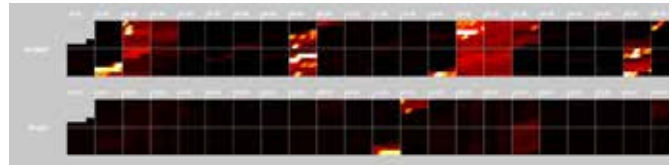
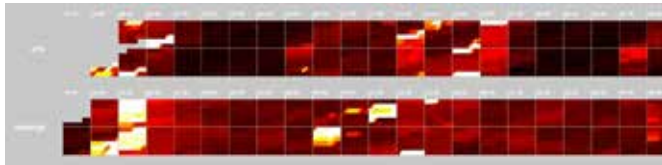
#### **Maarten Kruithof**

maarten.kruithof@tno.nl  
Intelligent Imaging

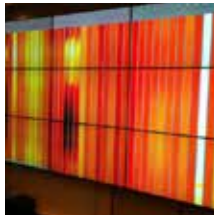
### ANALYSIS



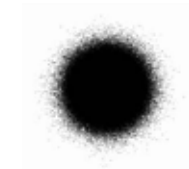
## BIG DATA VISUALIZATION



### Challenges



HIGH  
RESPONSIVENES



PREVENT CLUTTER

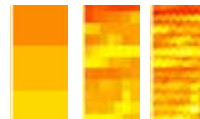
### SOLUTIONS



DATA TYPE SPECIFIC  
DATABASES



GPU DATABASES



INCREMENTAL  
VISUALIZATION



BINNING

# BIG DATA VISUALIZATION

Visualizing large amounts of data that originate from several sources of which the reliability is not always known, is a challenging task. This data, sensor data for example, comes in rapidly and needs to be processed and interpreted rapidly as well. In this project TNO has investigated the specific challenges of big data visualization through an extensive literature research and interviews with practitioners. It was found that visualization of large and heterogeneous data on relatively small screens, such as smart watches, often results in 'overplotting'. In addition, performance plays a role: on large screens it is a challenge to ensure a smooth interaction with the user when much data is displayed simultaneously. Solutions to these challenges can be found in the direction of specialized databases, use of GPUs as additional computing power and visualization techniques such as 'binning' and incremental visualization. These techniques have been presented and discussed during a well-attended workshop organized together with iMMovator. Future challenges lie in the area of intelligent visualizations: visualizations that automatically adapt to the user and her job. Also, the advent of new devices (VR goggles, 3D printers, Google Glass) offers new challenges in this domain.

## CONTACT

Soon the Technology Cluster 'Big Data Visualization' will start in which we will transfer the acquired knowledge of this study to SMEs. For more information on this topic or about the Technology Cluster, please contact us.

### **Erik Boertjes**

erik.boertjes@tno.nl

Media & Network Services

## DECISION SUPPORT



VALUE CHAIN COMPLEXITY, ARCHITECTURE AND INTEGRATION CHALLENGES: DIKE DATA SERVICE CENTRE (DDSC)



R&D projects to add advanced analysis modules



WATERBOARDS CAN ACCESS THE DDSC TO HAVE REALTIME INSIGHT IN DIKE BEHAVIOUR



LiveDikes connect the DDSC



IJKdijk experiments are recorded by the DDSC

# VALUE CHAIN COMPLEXITY, ARCHITECTURE AND INTEGRATION CHALLENGES: DIKE DATA SERVICE CENTRE (DDSC)

IT technologies, such as electronic sensors, are increasingly used for the monitoring and management of dikes. This results in a huge growth in the amount of data collected for this purpose. For this reason there is a growing need for an effective data management system to collect this data and make it practically accessible. The Dike Data Service Centre (DDSC) provides an integrated solution to store and effectively use the information for dike management. The DDSC uses the latest NoSQL database technologies (Cassandra) and hardware and is capable of processing up to 2TB per day. Examples of what can be stored are: height, movement (xyz), water pressure, temperature, infrared and radar scans.

Parallel to the development of the DDSC an R&D project provided solutions in the areas of automatic detections of anomalies in data streams, comparison of (near) real time and historical data and the incorporation of sensor data in geotechnical models for dike strength. In November 2014 the second phase of the development of the DDSC started and besides improvement in the data upload, data model, data management and Application Programming Interface, the results of the R&D project will be integrated.

## PARTNERS

Noorderzijlvest, Waternet, Wetterskip, Stichtse Rijnlanden, Rivierenland, Rijnland, Scheldestromen  
Research and development: Nelen & Schuurmans / Fugro, Deltares, HKV, Target, TNO

## MORE INFORMATION

The DDSC system has been developed in open source. You can find the project on GitHub: <https://github.com/ddsc>

## CONTACT

### **Erik Langius**

[erik.langius@tno.nl](mailto:erik.langius@tno.nl)

Business Information Services

### **Nico Pals**

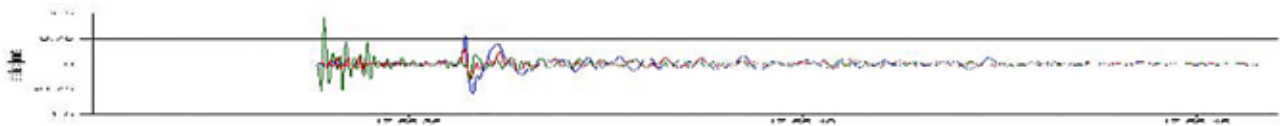
[nico.pals@tno.nl](mailto:nico.pals@tno.nl)

Strategic Business Analysis

## DATA COLLECTION & STORAGE | ECOSYSTEM

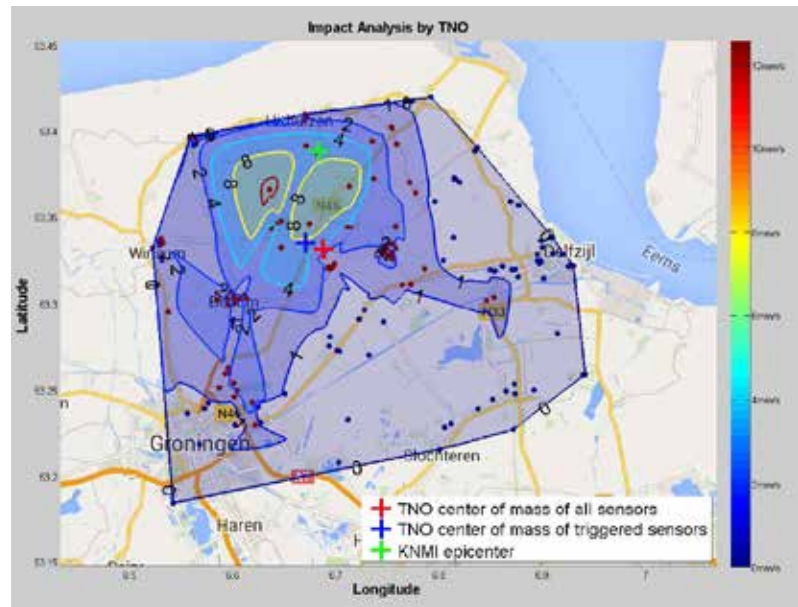


## BIG DATA STORAGE CHALLENGE: HIGH RESOLUTION VIBRATION MONITORING IN 200 BUILDINGS



Understanding the relation between measured vibration levels and occurred damage caused by gas induced earthquakes in the province of Groningen.

*Measured vibrations levels (mm/s) in the measurement area after an earthquake (November 5, 2014)*



# BIG DATA STORAGE CHALLENGE: HIGH RESOLUTION VIBRATION MONITORING IN 200 BUILDINGS

In Groningen gas induced earthquakes occur as a result of the production of natural gas. These earthquakes cause ground-borne vibrations that transfer to the foundations of buildings, causing the building itself to vibrate. These vibrations may result in damage to the building. As for the transfer of the ground-borne vibration to the buildings, both to the foundations and the building itself, there is relatively little information on the situation in Groningen. The same applies to the relationship between vibration in the buildings and occurrence of damage. This implies that currently there is relatively little insight into the effects of future earthquakes in Groningen, namely into the effects of more severe earthquakes than to date. For these reasons NAM decided to give TNO the assignment to set up a monitoring network in Groningen.

In 2014 TNO deployed the largest sensor network in the world measuring the effects of gas induced earthquakes in buildings. To safely store, process and distribute high resolution vibration data TNO build a highly robust and distributed data center based on Big Data technologies (Cassandra, OrientDB, Celery, and RabbitMQ).

The network already captured several earthquakes and recorded very valuable data on effects of gas induced earthquakes in buildings.

## CONTACT

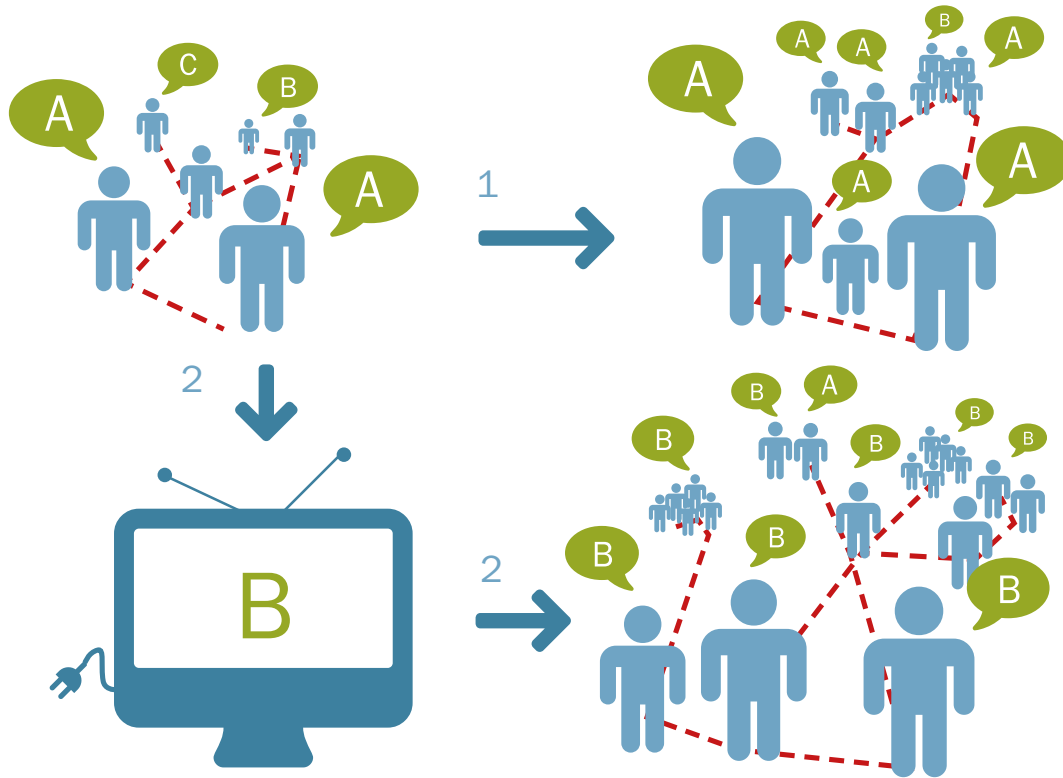
**Erik Langius**

erik.langius@tno.nl

Business Information Services

USE & MONITOR | DATA COLLECTION & STORAGE







# MULTI-AGENT MODELS FOR TREND PREDICTION IN SOCIAL MEDIA

The TNO project SON-M investigates predictive modeling in social media such as Twitter. Based on the social network structure and the observed individual behavior in the past, our models predict future message volumes on specific topics. In fact, we developed two trend prediction models which are compared against each other by using two case studies: large message volumes on banks and universities. The first model constructs the social network structure using empirical data. The second model is a multi-agent model which uses this empirical data to also estimate individual behavior parameters based on Cialdini's principles. For both case studies the results show that the combination of network structure and individual behavior parameters increases the predictive power for look-ahead-times ranging from a day to a week. As a next step, we would like to apply the generic models in specific application domains, aiming to predict trends as well as evaluate the effects of possible interventions.

## CONTACT

### **Thomas Attema**

thomas.attema@tno.nl

Performance of Networks & Systems

### **Peter-Paul van Maanen**

peter-paul.vanmaanen@tno.nl

Perceptual & Cognitive Systems

### **Erik Meeuwissen**

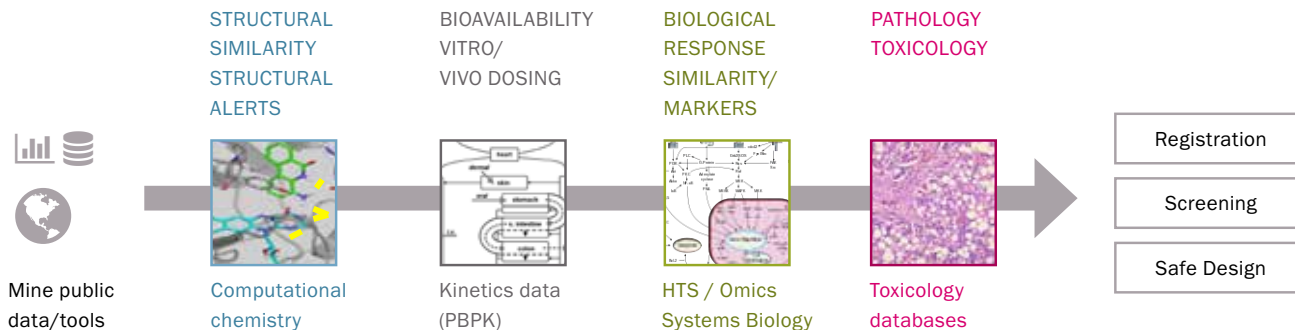
erik.meeuwissen@tno.nl

Performance of Networks & Systems

## ANALYSIS



DIAMONDS: DATA-WAREHOUSE INFRASTRUCTURE FOR ALGORITHMS MODELS AND ONTOLOGIES TOWARDS NOVEL DESIGN AND SAFETY



INTEGRATED APPROACH FOR TOXICITY PREDICTION

# DIAMONDS

DATA MANAGEMENT AND INTEGRATION

File Management	Website	Tooling Framework
Data Warehouse	Query / Browse	Workflows
Knowledge Base	Visualization	Webservices (API)
Annotation	Ontologies	

# DIAMONDS: DATA-WAREHOUSE INFRASTRUCTURE FOR ALGORITHMS MODELS AND ONTOLOGIES TOWARDS NOVEL DESIGN AND SAFETY

The current risk-assessment method for chemical substances requires a great deal of research. This can be done more quickly and effectively by combining areas of expertise and available data. For this purpose, TNO uses the DIAMONDS approach, a 'new-style' toxicological risk assessment method that makes it possible to determine the toxic properties of a chemical substance at an early stage, thereby reducing costs and preventing unnecessary testing on animals.

To this end, TNO has developed a general data management and data integration infrastructure that brings together data, knowledge and tools from diverse public as well as private sources into a single platform. DIAMONDS integrates knowledge and data relating to the chemical structure, kinetics, metabolism, system biology and toxicity of substances. We combine this with alternative experimental models for biological verification often in combination with diverse Omics technologies. By applying kinetic modelling as well as explorative and predictive algorithms to the collected heterogeneous data we are able to discover signatures that are predictive of the compound's toxicological effect. This enables faster, more

informed choices when investigating the safety of chemical substances for product development and/or registration.

## CONTACT

### **Eugene van Someren**

eugene.vansomeren@tno.nl

Risk Analysis for Products in Development

### **Dinant Kroese**

dinant.kroese@tno.nl

Risk Analysis for Products in Development

### **Peter Wolfs**

peter.wolfs@tno.nl

Sustainable Chemical Industries

## INFRASTRUCTURE



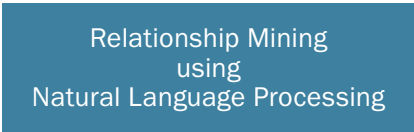
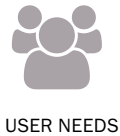
SIGNAL INTELLIGENCE FOR EMERGING FOOD SAFETY RISKS

**Melamine-contaminated powdered infant formula in China**  
18 SEPTEMBER 2008 - Over 6240 cases of kidney stones in infants with three deaths have been reported from across China as of 17 September. Kidney stones in infants are very rare.

Hindsight



Foresight



# SIGNAL INTELLIGENCE FOR EMERGING FOOD SAFETY RISKS

Can we prevent the next emerging food safety risk?

The melamine adulteration of infant formula (China, 2008) resulted in at least 54,000 sick children and four children died. This unexpected and large scale infant formula and milk powder disaster was preceded by an earlier pet food incident also involving melamine contamination. We found that often early signals precede food safety incidents. Such high impact, fortunately rare, incidents have further urged food safety authorities and the food industry to use early signal detection systems. This is a challenge and at TNO we have taken it up!

Major technical hurdles we had to deal with were very large volumes of data, variety, increasing speed of the publication process and copyright issues. During the last years we have developed Emerging Risk Identification Support (ERIS) which is based on the combination of relationship mining in scientific and technical literature and expert knowledge. Currently the ERIS system and our consultancy are used at food safety authorities and in food industry.

## MORE INFORMATION

F.J. van de Brug, N.B. Lucas Luijckx, H.J. Crossen, G.F. Houben. Early signals for emerging food safety risks: From past cases to future identification. Food Control, Volume 39, 2014, Pages 75-86.

## CONTACT

**Fred van de Brug**

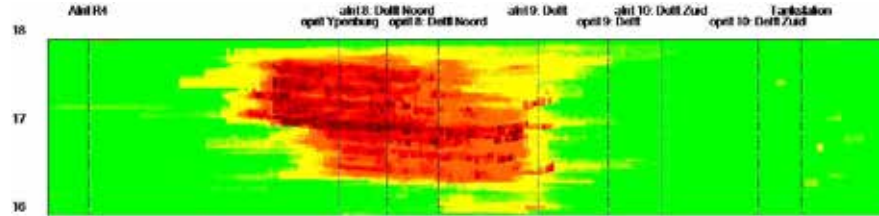
fred.vandebrug@tno.nl

Risk Analysis for Products in Development

ANALYSIS | DECISION SUPPORT



oprit 8: DELFT NOORD



- Precompetitive pilot Dutch National Data Warehouse for Traffic Information (NDW) together with DITCM Innovations
- Goal: investigate whether making more extensive use of floating car data (FCD) and data fusion is a good approach to enable a reduction of the number of fixed location sensors



# NDW TRAFFIC DATA FUSION PILOT

The Dutch National Data Warehouse for Traffic Information (NDW) has, together with DITCM Innovations, organised a precompetitive pilot to investigate whether making more extensive use of floating car data (FCD) and data fusion is a good approach to enable a reduction of the number of fixed location sensors, and to assess the added value for NDW's end users.

Three teams participated in the pilot (in-kind participation):

- Team ARS T&T and HERE;
- Team Be-Mobile, Goudappel Coffeng, MAPtm and VORtech;
- Team CGI, SAS, Grontmij, Accenture, HP, TU Delft and TomTom.

A selection of roads in and around Delft served as test site. For several scenarios it was assessed which level of quality of traffic state information (speed and volumes) could be achieved using FCD, data from fixed location loop detectors and data fusion.

Some conclusions of the project are:

- An added value of FCD and data fusion is that the visualisation of these data gives a more complete picture of traffic, and insight into the quality and plausibility of the data.

- The teams are confident that FCD and data fusion can be deployed soon.
- Joint development by public and private partners in the pilot phase lays a good foundation for future collaborations.

## CONTACT

### **Gerdien Klunder**

gerdien.klunder@tno.nl  
Smart Mobility

## USE & MONITOR



**E-government** data sources  
(CBS, Trade & Commerce, Real Estate, Public Safety),...

**UTILITIES** data:  
assets, networks,  
sensory data, CIM

**GEO-data:** cartography,  
INSPIRE, WION/WIN,  
BAG, BGT



## STEP-BY-STEP GUIDE TO PUBLISH LINKED DATA

- Step 1: Select data
- Step 2: Prepare the data
- Step 3: Model the data
- Step 4: Define a naming scheme
- Step 5: Convert the data
- Step 6: Organize governance
- Step 7: Add metadata
- Step 8: Publish the data
- Step 9: Link the data

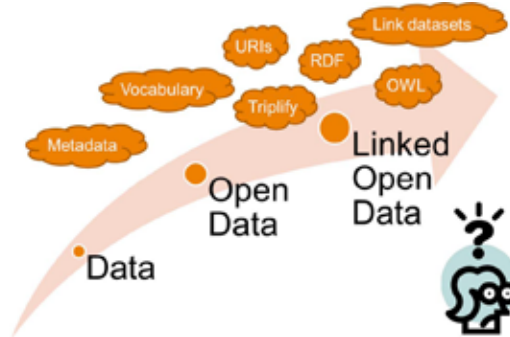
**TNO** innovation  
for life

**altlander**

**GEONOVUM**

**Geodan**

**TU Delft** Delft University of Technology





# LINK(ED) DATA TO GIVE MEANING TO BIG DATA

Within the CERISE-SG project ([www.cerise-project.nl](http://www.cerise-project.nl)) TNO, together with project partners Alliander, Geonovum, TU Delft and Geodan, explores and applies Linked Data principles to be able to give meaning to data and to be able to use data from one domain in other domains without the need to transform the data. The project is funded by the Topsector Energy and relates to the developments of Smart Grids.

By explicitly defining and modelling the meaning of data and linking the models, it becomes possible to define a question according to one specific model, get the information from a source with another model, and answer the question in the first model again. Linked Data turns (big) data into information and allows for the information to be used both inside and outside the domain in which it was originally gathered. Furthermore, Linked Data provides a way to interconnect data from different domains without the need to change the way data is acquired and stored in the different domains. It therefore allows for flexible connections between data and information.

One of the outcomes of the CERISE-SG project is a step-by-step guide to publish Linked Data in such a way that other parties can reuse the data.

CONTACT

**Jasper Roes**

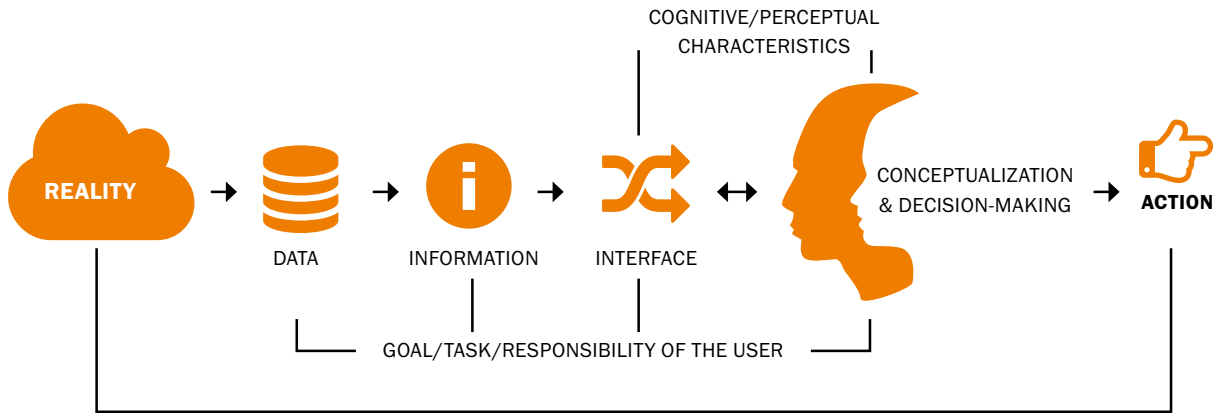
[jasper.roes@tno.nl](mailto:jasper.roes@tno.nl)

Business Information Services

ANALYSIS



HUMAN FACTORS IN BIG DATA



# HUMAN FACTORS IN BIG DATA

Transforming (Big) Data into value can look like a magical and intangible process. Only with sufficient technical expertise one can manage to make the value of the data explicit, but often this expertise is not (yet) available in organizations. By turning the process around –from value to data– organizations can be empowered and take control over the development of new information services.

The context of use as well as cognitive capacities and task of end users is the starting point for the Human Factors in Data process for the development of new information services based on (big) data. The end-users can be new or future clients or employees of an organization that need support in their work.

Human Factors in Data designs the human aspects of the information service; user interface, usability and user experience, training of end-users, work process, change management, measuring impact, etc.

## CONTACT

Organisatie-TNO-HumanFactorsinData@TNO.nl

**Jenny de Boer**

jenny.deboer@tno.nl

Human Behavior & Organizational Innovations

## USE & MONITOR



**SMALLER STUDIES WITH COMPLEX DESIGN**

Centrally available nutritional studies (currently 58) ★

Unstructured/unmapped data (e.g. local excel sheets) ★



API FAIRport compliant

(Nutritional) ontologies ELIXIR compliant

EUROPEAN NUTRITIONAL FAIRPORT

★ **DaShiN**

**FAIR =**  
**F**indable  
**A**ccessible  
**I**nteroperable  
**R**e-usable

Queries and analysis

data SHaPER  
ECRIN

API FAIRport compliant and data mapped

Structured/unmapped data (local database) ★

BioMedBridges

**LARGE STUDIES WITH SIMPLE DESIGN**

# DATA SHARING IN NUTRITION (DASHIN)

TNO is coordinator of the European project European Nutritional Phenotype Assessment and Data Sharing Initiative (ENPADASI). The main objective of this project is to deliver an open access research infrastructure (DaShiN) that will open data of most European Nutritional studies. The data will adhere to the FAIR concept (<http://datafairport.org/>) that all data should be findable, accessible, interoperable and re-usable. Many nutritional studies, including large data sets of for example organ measurements and biological molecules, have been made findable, accessible and re-usable by storing them in the Phenotype database ([www.dbxp.org](http://www.dbxp.org)), of which TNO is leading the development. This application will be used to upload data that is not yet structured – many biologists store their data in folders with excel sheets. A FAIRport compliant

Application Programming Interface (API) will be included in the application to guarantee interoperability. In the case that data of the nutritional studies is already structured, data will be made interoperable by including a FAIRport compliant API to the study specific database. Data in these systems should also be mapped to terms in the Phenotype database for re-usability purposes.

Case studies will be defined to evaluate the actual

infrastructure and to show the added value of connecting similar studies on the topic of resolving chronic diseases with lifestyle related solutions. We hypothesize that integrating research in the food-, nutritional-, social- and health sciences will increase knowledge and deliver innovative, novel and improved concepts for the prevention of diet-related diseases, thereby limiting the need for new and larger intervention studies.

## CONTACT

**Jildau Bouwman**

[jjildau.bouwman@tno.nl](mailto:jjildau.bouwman@tno.nl)

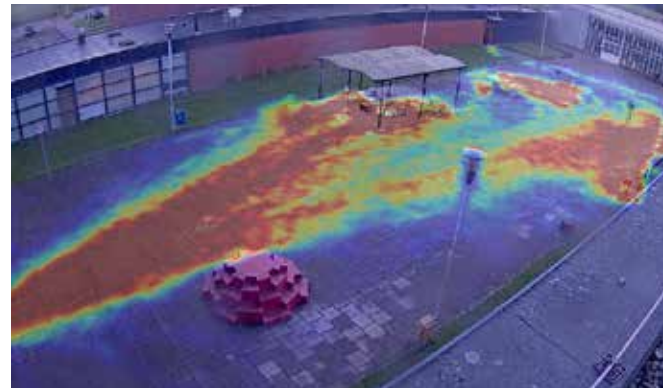
Microbiology & Systems Biology

DATA AGGREGATION | INFRASTRUCTURE

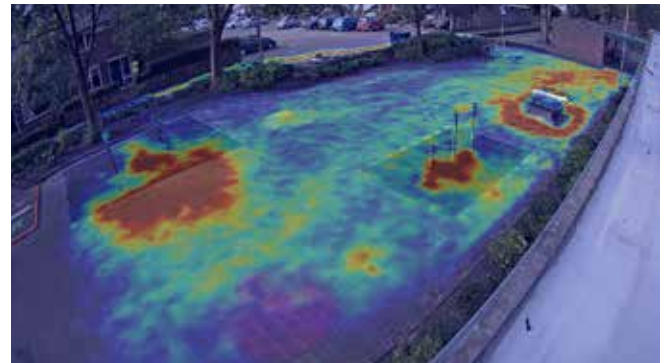


## **BIG DATA FROM SCHOOLYARDS HEALTHIER SCHOOLS, HEALTHIER CHILDREN**

The detection and measurement of movement and interaction between children in schoolyards by intelligent video and big data analysis.



*Heat map of occupancy schoolyard at secondary school*



*Heat map of occupancy schoolyard at primary school*

# MEASURING MOVEMENT AND INTERACTION BETWEEN CHILDREN IN SCHOOLYARDS THROUGH INTELLIGENT VIDEO AND DATA ANALYSIS

More and more children and young people exercise too little; only one in five gets the daily norm of 60 minutes (moderate) intensive exercise. Schools can help by, for example, designing schoolyards in such a way that they encourage children to move more. However, the effect of this type of solution can only be assessed by gathering data and analyzing it on a larger scale. Current measurement methods are not accurate enough and require every child to wear a device to measure his or her movements, which is difficult to implement in practice.

In a pilot the TNO departments Child Health, Intelligent Imaging, Media & Network Services and Life Style have developed a new technique to measure activity in children or adolescents with intelligent sensors. This method is objective, less invasive, costs less than conventional measuring methods and ensures privacy. At this time, it is possible to obtain statistics on the speed with which children in a school playground move around through automatic analysis of video images and, through heat maps, it is also possible to see at which schoolyard locations the children move most.

The added value of big data emerges as the data and statistics from multiple schools in different municipalities can be combined with each other or in combination with other (open) datasets.

## CONTACT

### **Eline Vlasblom**

eline.vlasblom@tno.nl

Child Health

### **John Schavemaker**

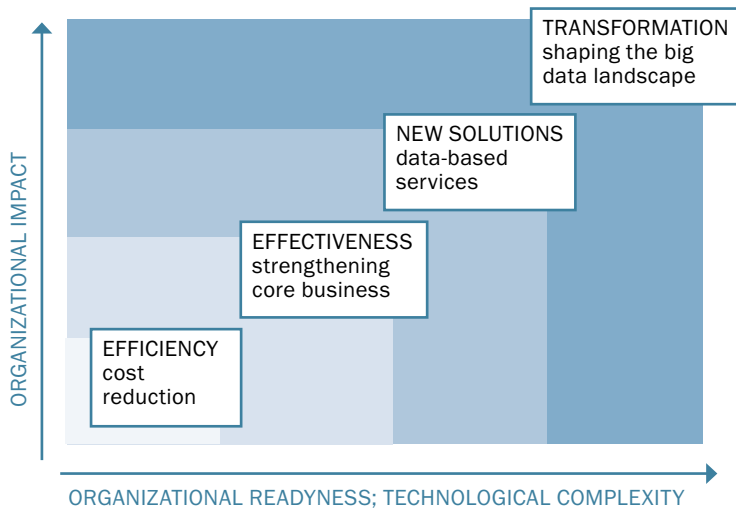
john.schavemaker@tno.nl

Media & Network Services

DATA COLLECTION & STORAGE | ANALYSIS



DATA-DRIVEN INNOVATION AS A STRATEGIC CHALLENGE





# DATA-DRIVEN INNOVATION AS A STRATEGIC CHALLENGE

Data-driven innovation (DDI) holds great potential to tackle the major challenges of our time (i.e. healthy aging, climate change, etc.). Solutions to these challenges require not merely enhancements of existing products, services and practices (which are the most common type of DDI for incumbents), but also radically new propositions, new value models, new value networks that support new types of collaboration, and thriving ecosystems that nurture these kinds of innovations.

However, it is very difficult to realize these more complex (data-driven) innovations and to foster the required innovation ecosystems. Therefore, it is important to approach DDI as a strategic challenge with data-specific mechanisms and dynamics, which demands not only technological innovations, but also social and institutional innovations that build on deep 'vertical knowledge' of the specific domain(s) of the DDIs (i.e. health and energy). The findings of this OECD project – based on empirical research – provide insights in the most common value creation mechanisms and the specific dynamics that shape new data-driven innovation ecosystems. With these insights TNO can help to instigate, orchestrate and support the development of advanced DDIs and new (or emerging and existing) ecosystems that nurture the development and adoption of these innovations, now and in the future.

## CONTACT

**Jop Esmeijer**

[jop.esmeijer@tno.nl](mailto:jop.esmeijer@tno.nl)

Strategy & Policy

## ECOSYSTEM



1. Users can pose natural language queries to GOOSE



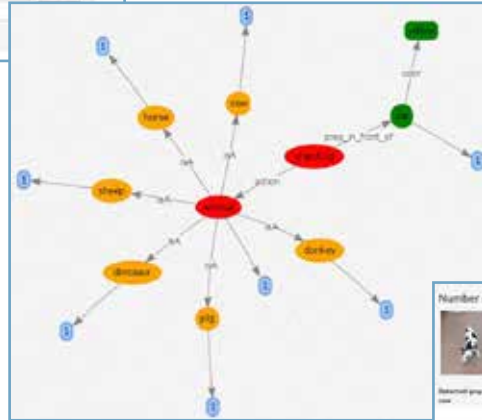
CAMERAS AS SENSORS



NATURAL LANGUAGE PROCESSING



SEMANTIC REASONING

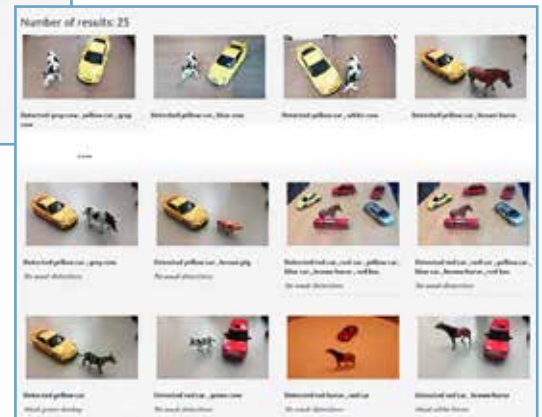


2. A semantic graph visually explains how a query has been interpreted and matched to the available image detectors



OBJECT DETECTION

3. A smart results ranking and retrieval algorithm allows for fast and effective retrieval of images



# GOOSE: SEMANTIC SEARCH IN SENSOR DATA

With the growth of open sensor networks, multiple applications in different domains make use of a large amount of sensor data, resulting in an emerging need to search semantically over heterogeneous datasets in a meaningful and, at the same time, easy and intuitive manner. In semantic search, an important challenge consists of bridging the semantic gap between the high-level natural language query posed by the users and the low-level sensor data.

In order to bridge this gap, we developed the GOOSE system, a general-purpose search engine focused on cameras as sensors that allows users to retrieve images and videos in real-time from multiple and heterogeneous sources. Users can pose queries in natural language to the GOOSE system. These queries are then interpreted using the Stanford Parser, semantic rules and the Linked Open Data source ConceptNet. Interpreted queries are presented to the user as an intuitive and insightful graph in order to collect feedback that is used for further reasoning and system learning. A smart results ranking and retrieval algorithm allows for fast and effective retrieval of images corresponding to the user query.

The GOOSE project ended in December 2014 and involved several TNO departments (BIS - Business Information

Services, II - Intelligent Imaging and MNS - Media and Network Services) with different areas of expertise (Semantic Modelling, Computer Vision and Human Media Interaction).

## CONTACT

**Laura Daniele**

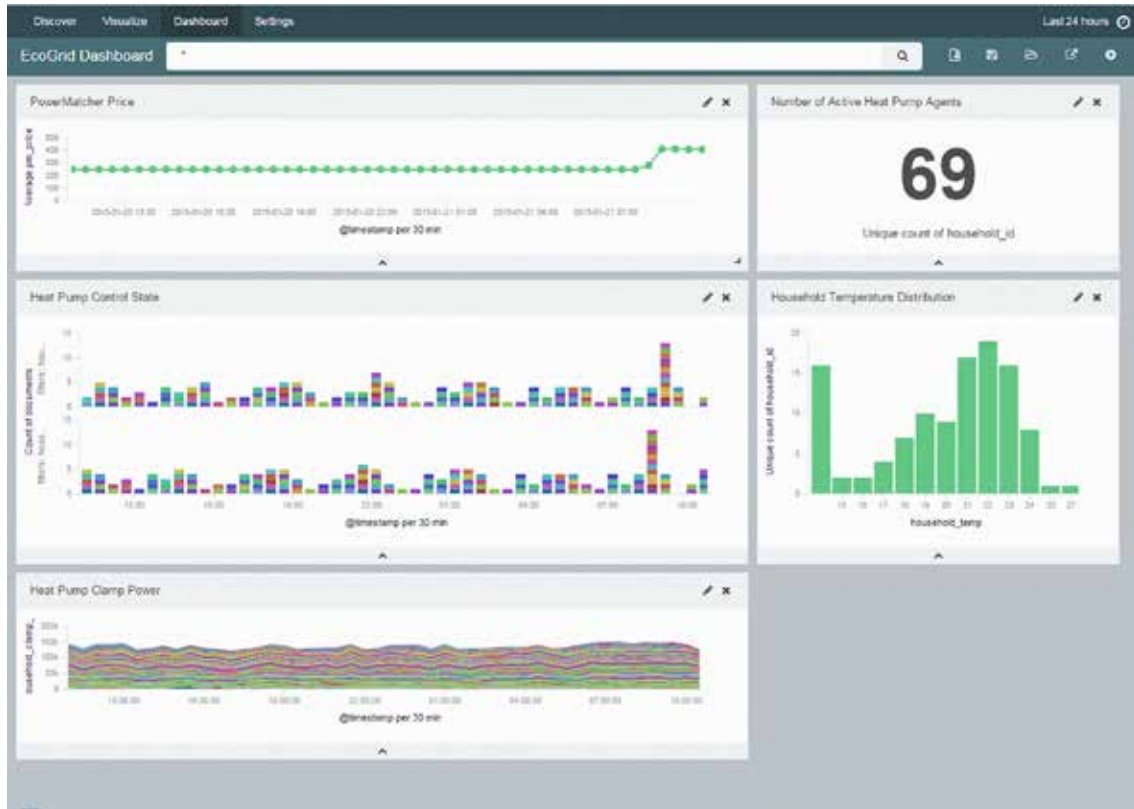
[laura.daniele@tno.nl](mailto:laura.daniele@tno.nl)

Business Information Services

## DECISION SUPPORT



SMART GRID RESEARCH WITH BIG DATA TOOLS



# SMART GRID RESEARCH WITH BIG DATA TOOLS

As TNO, we participate in several Smart Grid projects and pilots around Europe. One of them is the **EcoGrid EU** project, which uses the Danish island of **Bornholm** as a test bed. A large number of households participate in the pilot. We are controlling residential heat pumps – devices which basically use a lot of electrical energy to store and produce heat – in some of these houses. We control these pumps with **PowerMatcher**, a multi-agent, market-based control algorithm.

When you are participating in a pilot there is a certain scientific component which must be addressed. You usually do that by gathering data and doing some kind of analysis of that data. You also want to monitor if everything is working as it should on a somewhat real-time basis. It is very tempting to just keep on using your old tool set in these kinds of situations. There is just this pilot, right? We just have to use this triple SSH tunnel to connect to that off-site data gathering location one more time, right? We just have to use find-replace on all these log files one more time, right? When ‘pilotitis’ hits again and “one more time” gets repeated more than three times, it is time to think about a better solution.

That solution was using Big Data tools. We now use the **ELK stack** to monitor and analyze log data in real time,

with pretty pictures and a click-it-yourself dashboard to boot. It made our life so much easier and has been instrumental in cases where we just could not figure out why stuff was not working. And we will be using it for other Smart Grids projects as well.

## CONTACT

**Mark Bastiaans**

mark.bastiaans@tno.nl

Service Enabling & Management

ANALYSIS | DECISION SUPPORT | USE & MONITOR



## DATA

### **NETHERLANDS WORKING CONDITIONS SURVEY**

- Yearly since 2005
- 23.000 employees (40.000 in 2014)
- Labor market, working conditions, health
- In collaboration with Statistics NL & Ministry of Social Affairs and Employment

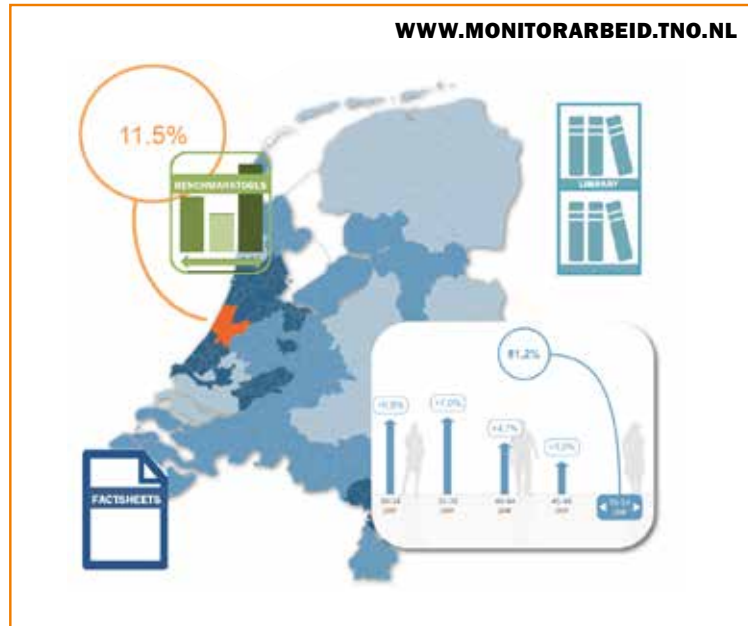
### **NETHERLANDS EMPLOYER SURVEY ON WORK**

- 2008, 2010, 2012, 2014
- 5.000 employers
- Working conditions, company polity, business outcomes

### **SELF-EMPLOYED SURVEY ON WORK**

- 2012, 2014
- 4.000 self-employed
- Working conditions, company polity, business outcomes

## VISUALS AND TOOLS



# MONITORING WORK & EMPLOYMENT BY TNO

The employment situation of the Dutch workforce is changing constantly. TNO investigates these changes and their impact on productivity, innovation, health and sustainable employment within the Monitoring Work & Employment Program. The program is financially supported by the Dutch Ministry of Social Affairs and Employment.

The program relies on several data sources:

**The Netherlands Working Conditions Survey** (Nationale Enquête Arbeidsomstandigheden) is a large scale periodical investigation into the working conditions of Dutch employees. Eleven surveys have been performed to date, in 2003 and 2005-2014. Some 23,000 employees per year have responded to the surveys

**The Netherlands Employers Work Survey** (Werkgevers Enquête Arbeid) systematically collects data on work and employment in establishments of profit as well as nonprofit organizations in the Netherlands. Three surveys have been performed (20010, 2012, and 2014). Some 5,000 establishments per year have responded.

**The Self-Employed Survey on Work** (Zelfstandige Enquête Arbeid) collects data on work and employment of self-employed workers. The survey was performed in 2012 and

2014. Some 4,000 self-employed responded.

Close collaboration with Statistics Netherlands (CBS) enables the program to match data from the surveys on data from registrations in order to facilitate policy-evaluation of ministries, municipalities, policy makers, industry associations etc.

## MORE INFORMATION

Visit [www.monitorarbeid.nl](http://www.monitorarbeid.nl) to find out more about the program, the data and the data-tools that are developed.

### **Seth van den Bossche**

[seth.vandenbossche@tno.nl](mailto:seth.vandenbossche@tno.nl)

Work & Employment

### **Marloes van der Klauw**

[marloes.vanderklauw@tno.nl](mailto:marloes.vanderklauw@tno.nl)

Work, Health & Care

## DATA COLLECTION & STORAGE | USE & MONITOR



# NEWS GENIUS: YOUR CUSTOMIZED VIDEO NEWS BROADCAST





# NEWS GENIUS: YOUR CUSTOMIZED VIDEO NEWS BROADCAST

What is the future of news? Smartphones, tablets, social networks and other new media allow people to follow the news around the clock. But at the same time, a day still only counts 24 hours, in which people have to work, travel, raise children, to eat and also get some sleep. How can we provide you with the news you want to see?

To help people to keep up to speed on the most important news in an efficient way, RTL, TNO and User Intelligence jointly developed an app that automatically creates an RTL News Bulletin based on individual preferences. This app offers a video news bulletin, a personal combination of video clips from the RTL News media repository. Users can give an indication of how much time they want to spend, and the system picks up on their tastes through implicit user feedback such as video skips.

The project provides good insights of the role of Big Data in consumer central media services, such as:

- What constitutes a 'proper' video news broadcast?
- The importance of high quality metadata.
- What user feedback / app usage statistics to use?
- Big Data is only part of the problem/solution.
- And the importance of user studies.

## CONTACT

**Martin Prins**

[martin.prins@tno.nl](mailto:martin.prins@tno.nl)

Media and Network Services

DECISION SUPPORT | USE & MONITOR



MULTI-USE BIG DATA IN THE AMSTERDAM CANALS

AIS positions of boats



The image shows a black AIS display unit with a screen showing a map and a compass. To its right is the RWS logo, which consists of a stylized blue wave and the letters 'RWS'.

LIGHT BARRIERS for passage detection of vessels



The image shows a concrete light barrier with a black sensor unit mounted on top, used for detecting vessels.

WEATHER DATA



The image shows the logo for 'weeronline' with the tagline 'Altijd jouw weer'.

OBSTRUCTIONS



The image shows a red and white triangular warning sign with a black silhouette of a person carrying a load, and several orange and white buoys.


AnySense – data management  
Traffic model – traffic forecasts  
Signaleringsysteem – licenses



  
European  
Excellence  
Award 2013


  
Now also for  
bridges in  
Zuid-Holland!

VAARKAART AND VAARWATER-APP  
For tourists



The image shows a computer monitor displaying a website with a map and a weather forecast, and a smartphone displaying a similar map application.

OFFLINE TRAFFICMODEL  
For Traffic Management



The image shows a map of a canal network with yellow and blue lines indicating traffic flow and management.

INCIDENT REPORTS  
enforcement



The image shows a blue boat on a canal and a computer screen displaying an incident report form with a map.

# MULTI-USE BIG DATA IN THE AMSTERDAM CANALS

To combine various real-time data sources to achieve multiple goals is the promise, if not the holy grail, of big data. However, the amount of discussions and ideas on this topic far outnumbers the practical real-life implementations.

For Waternet Amsterdam, we have realized a real-time, multi-use big data solution for the Amsterdam Canals delivering up-to-date and forecasted traffic information for tourists (via apps) and for traffic management (via a traffic model), and tracking information for enforcement of licenses for tour boats via a position tracking license enforcement system.

For this purpose, TNO combines real-time AIS position information of tour boats, real-time information of passing vessels from light-barrier sensors, current and forecasted weather information and static information about obstructions in the Canals. Data from these sources are collected, combined and enriched in the proven TNO AnySense platform, analyzed and modeled in traffic model software and made available for application in apps, websites, traffic models, and reporting.

In 2013 the passage detection system and apps (the first application of the abovementioned) won the European

Excellence Award for mobile apps and communications. This concept is now also piloted in the province of Zuid-Holland for monitoring boat passages and heights.

## CONTACT

**Michel Emde**

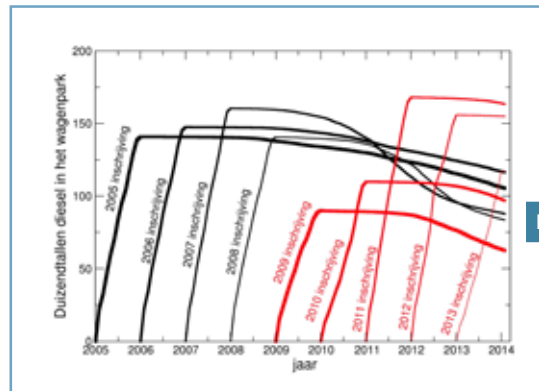
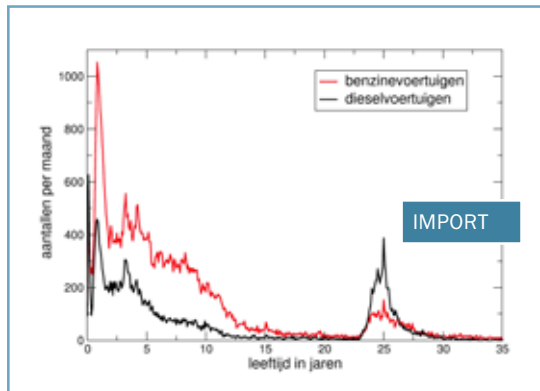
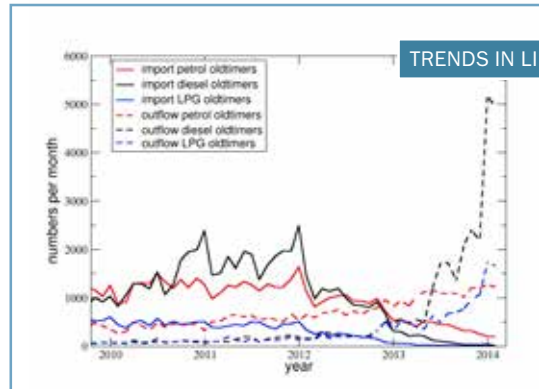
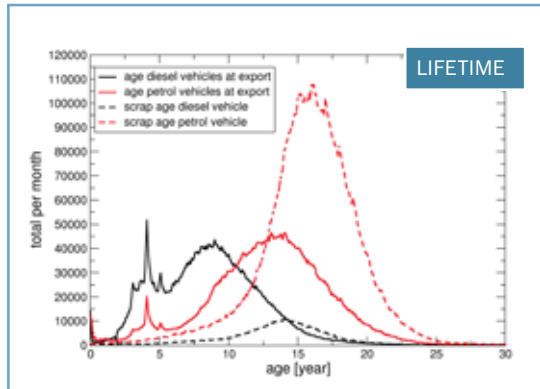
michel.emde@tno.nl

Business Information Services

DECISION SUPPORT | USE & MONITOR



TRENDS IN VEHICLE FLEETS



# TRENDS IN VEHICLE FLEETS

The development of the Dutch vehicle fleet of passenger cars is driven by economics, taxes, and the vehicle lifetime. For the Ministry of Infrastructure and Environment TNO analyzes the trends. For this purpose we acquire the complete mutation database of the RDW (Road Authority) over a period of nine years, which consists of several hundreds of million records. In this data we recover import and export, changing fleet compositions, trading characteristics, etc. There are about 12 million cars in this time period, which each are bought and sold many times in their lifetime. Preferably we want to trace the whole fleet development.

The data file is about 6 GB. Generally we approach this as a data stream from which relevant properties are extracted via scripts in AWK, C, and MATLAB. Finding niche vehicle groups such as plug-in hybrid and electric vehicles are done by matching this data with RDW sales database, which have more complete records on vehicle characteristics. The appropriate angle to present the data is important to highlight main trends, relevant for policy makers. We are therefore in the process of automating the procedure to select views, without preselecting certain features and groups among the different axes: technology, lifetime, ownership, etc.

## CONTACT

**Norbert Ligterink**

[norbert.ligterink@tno.nl](mailto:norbert.ligterink@tno.nl)

Sustainable Transport & Logistics

ANALYSIS | DECISION SUPPORT



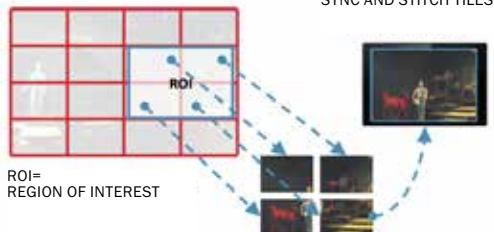
### COMMONWEALTH GAMES 2014



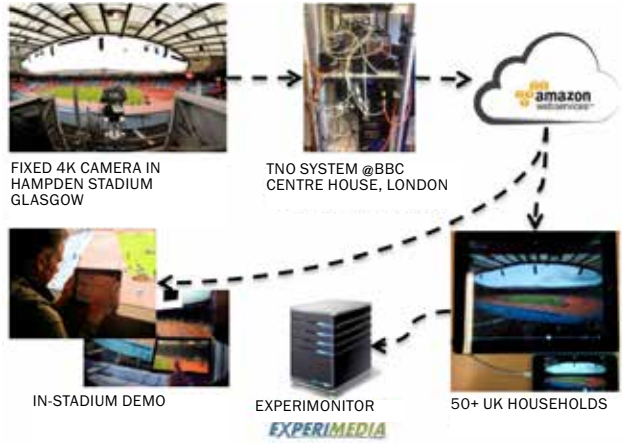
### CONCEPT



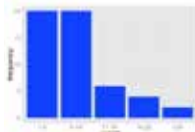
### TECHNOLOGY



### IMPLEMENTATION



### RESULTS



Participant	Age	Gender	View	Time	Score	Rating	Engagement	Attention	Retention	Completion	Dropout	Abandon
Participant 1	25	M	1080p	1:30	85	4.5	90	85	95	100	0	0
Participant 2	30	F	720p	1:45	75	3.8	80	75	85	95	5	10
Participant 3	35	M	1080p	1:20	90	4.8	95	90	100	100	0	0
Participant 4	40	F	720p	1:55	65	3.2	75	70	80	90	10	15
Participant 5	45	M	1080p	1:35	80	4.2	85	80	90	95	5	10
Participant 6	50	F	720p	2:00	55	2.8	70	65	75	85	15	20
Participant 7	55	M	1080p	1:40	70	3.5	78	73	82	90	10	15
Participant 8	60	F	720p	2:10	45	2.5	65	60	70	80	20	25
Participant 9	65	M	1080p	1:50	60	3.0	72	67	77	87	12	17
Participant 10	70	F	720p	2:20	35	2.2	60	55	65	75	25	30

- 52 participants
- objective and subjective metrics
- 1M+ measurements in 5 days

# USING BIG DATA ANALYTICS TO IMPROVE THE USER EXPERIENCE

Multimedia are an important and valuable source for insight into end user experiences. Big Data analytics allows us to investigate and improve the Quality of end user Experience for consuming multimedia. TNO, in collaboration with BBC R&D, performed an interactive video application trial during the 2014 Commonwealth Games. Over 50 UK households enjoyed athletic games, captured in ultra-high resolution video (known as UHDTV or 4k), on their smartphone or tablet. For this new way of watching TV, TNO employed their tiled streaming technology and iXperience application. That is, viewers could watch a live UHDTV broadcast interactively, by navigating through the video. TNO measured end user experience via the EXPERIMonitor Big Data analytics platform, by capturing objective data (e.g. application and data usage) as well as subjective data by asking viewers to rate application and service aspects. The trial provided the following insights: 1) data collected in a real-world setting need to be properly cleaned before analysis 2) QoE changes over time, from being limited initially when users are trying out service functionalities, and increasing afterwards once users get acquainted to the service 3) for long term usage, QoE is correlated to specific network parameters (buffer/switch times).

## CONTACT

### **Omar Niamut**

omar.niamut@tno.nl

Media & Network Services

### **Lucia D'Acunto**

lucia.dacunto@tno.nl

Media & Network Services

## USE & MONITOR



## RTL ROUTE RADAR APP 200.000+ DOWNLOADS





# BIG DATA WITH ROUTE RADAR

At the first TNO Small Big Data Congress in 2014, the RTL Route Radar project was pitched with the promise to operationalize the traffic models of TNO Smart Mobility in the cloud and to make the outcomes available to RTL Nieuws and the Route Radar app of Buienradar via Application Programming Interfaces (APIs) in order to obtain location data with the consent of the end users. These location data can then be used to calculate current traffic conditions.

This year, I want to tell you that this has been successful to a great extent! The Route Radar app has been downloaded more than 200,000 times, and the result of the TNO models is used in the daily RTL news broadcast to discuss the traffic situation.

This is interesting, because we are dealing with large amounts of data that must be processed quickly and must be combined with a variety of other sources (fusion), while the reliability of the data plays a major role. To tackle this, TNO has set up a pragmatic but effective platform that efficiently processes the obtained data in the traffic models, resulting in 'bite-size' pieces that are offered to the market in standard formats.

## CONTACT

**Paul van den Haak**

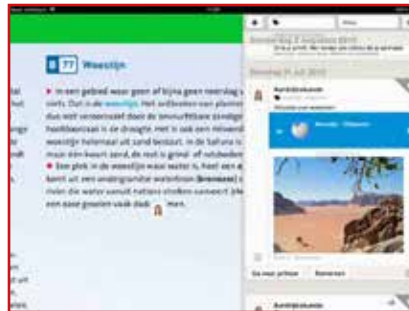
paul.vandenhaak@tno.nl  
Smart Mobility

ANALYSIS | DECISION SUPPORT | USE & MONITOR



## LEARNING STRATEGIES ANALYTICS

- Combining multiple sources of information &
- expertise : learning psychology & data mining
- Self efficacy and motivation have more influence on learning performance than the use of the digital learning environment



Raw Log Data

Digital Learning Environment

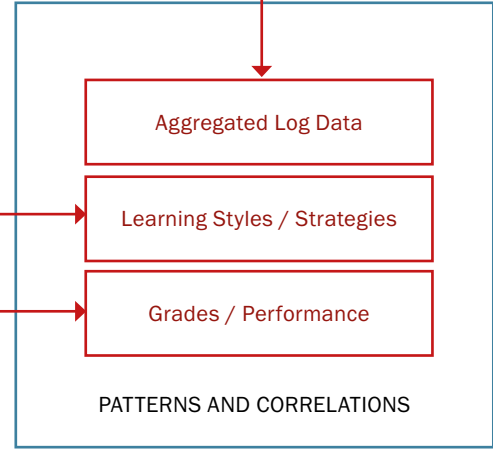
Questionnaires

School Tests

Aggregated Log Data

Learning Styles / Strategies

Grades / Performance



# LEARNING STRATEGIES IN DIGITAL LEARNING ENVIRONMENTS

The log data of digital learning devices provides information on the learning processes of students. By analyzing patterns of learning behavior and performance, adaptive applications can be developed to provide effective individual personalized support. Learning strategies include the order of learning, use of social media, use of different information types, application of meta-cognitive skills etc.

In this project we have followed groups of students from different schools during 2-3 months. The students fill out questionnaires to determine learning styles and strategies, grades on a test are collected, and the activities in the digital learning environment are logged. A control group that does not use the digital learning environment is also included in the study.

From our study we can conclude that self-efficacy and motivation have most influence on learning performance, more than the use of the digital learning environment. Effectivity of learning depends on factors like the didactics of the teachers and individual differences. Correlations are found between log data and questionnaire data validating the usefulness of the log data and giving insights into which data should be logged.

## CONTACT

### **Rianne Kaptein**

rienne.kaptein@tno.nl

Media & Network Services

### **Gerard Veldhuis**

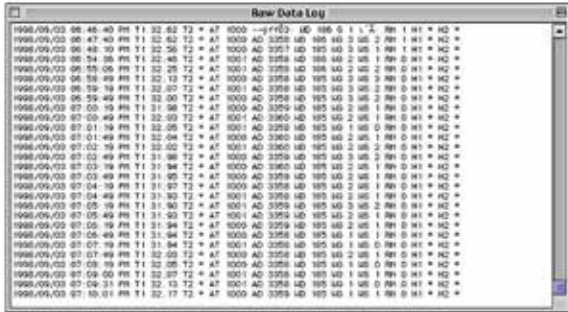
gerard.veldhuis@tno.nl

Training & Performance Innovations

## ANALYSIS



### LOG DATA



### RELATIONS



TIME, EVENTS, ATTRIBUTES



# ADVANCED ANALYTICS ON LOGS

Data is exploding; not only due to the internet, social media, sensors, and the internet-of-things, but also because each and every machine or network is now monitoring and logging what is happening. Obviously there is much value in such a log, especially when things go wrong. However, getting that value (the cause of the malfunction) out is getting complicated due to the enormous amount of information.

In this project we have developed a novel method that can mine such timed activity logs, and is able to find patterns (through time) that can explain under which conditions a malfunction is likely to occur. For example, if the pattern A-B-A-G, possibly even with specific attributes (or values), always leads to failure.

We have applied this method on logs of tactical data messages during a military exercise. Aim was to identify which chain of events leads to a missile being, or not being, intercepted by a task group. Within seconds the method was able to find the reason behind this, while the analyst that was working concurrently had hardly started, and was only finished after an hour. Obviously the patterns that are found still need interpretation, but at least it saves the analyst the effort of digging through all the data.

CONTACT

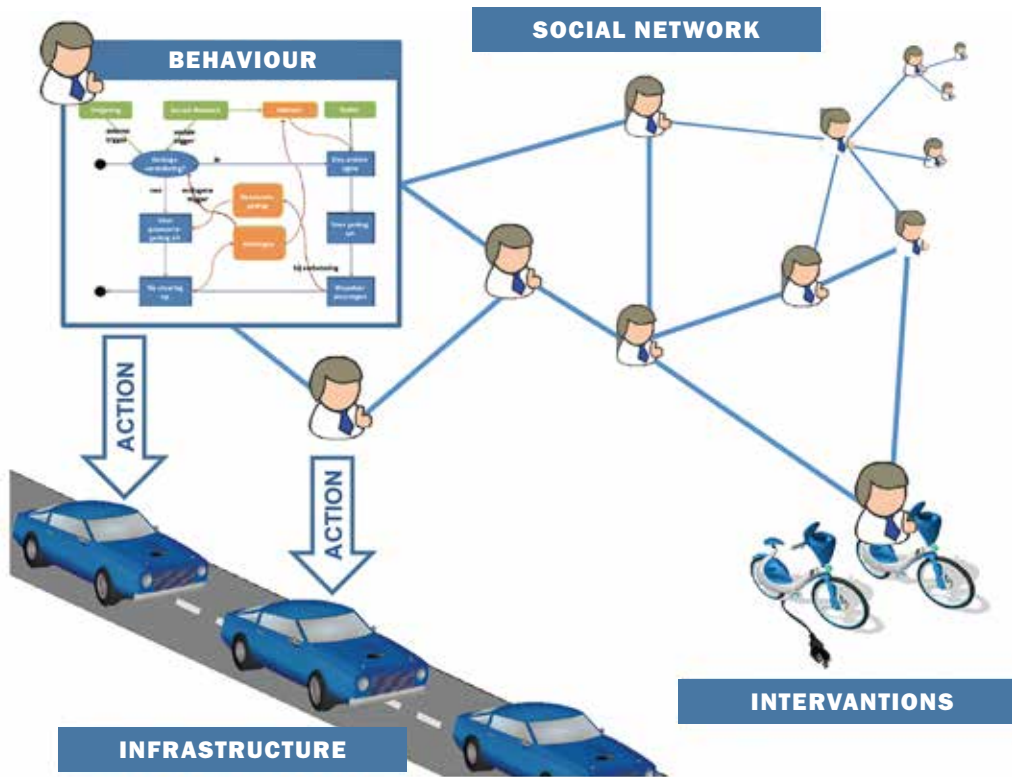
**Selmar Smit**

[selmar.smit@tno.nl](mailto:selmar.smit@tno.nl)

Modelling, Simulation & Gaming

ANALYSIS





# FOUNTAIN

FOUNTAIN is a framework developed within the ETP behavior and innovation 'project Mobiliteit', and allows for prescriptive analytics on habitual behavior. This means, it can not only be used to predict what will happen but also why behavioral change will happen. The main use case that has been tested is that of commuter behavior, however we have also implemented snacking and citizen participation.

The main objective of FOUNTAIN is to serve as a what-if-generator to policymakers. For example, what would happen if we give commuters - who normally travel by car during rush hour - 2 euro's if they don't? Who would change their behavior? Why do they change their behavior? And would that new behavior last, even if we stop stimulating? In order to answer these types of questions we developed a model of habitual behavior, together with social scientists, behavioral scientist and specialists on transportation.

Obviously, in order to act as a what-if-generator, it needs to be thoroughly calibrated. For this, we did a first step by adjusting the model parameters to fit the outcomes of an actual intervention in Brabant. After this, we tested the model by replacing the infrastructure (roads, trains, demography) of Brabant, with that of Utrecht. From this we also had the evaluation data of an intervention and, somewhat surprisingly with so little calibration, we got very

similar results to the real ones. The province of Utrecht learned a lot using this model; it brings structure to defining interventions and allows them to study the possible effects of interventions, including side effects, on the long run.

## CONTACT

**Selmar Smit**

[selmar.smit@tno.nl](mailto:selmar.smit@tno.nl)

Modelling, Simulation & Gaming

## ANALYSIS





**FACT 1:**  
MANY OPEN DATASETS CAN NOT  
BE ACCESSED AND USED



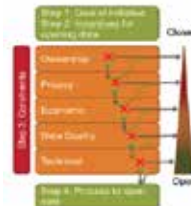
**FACT 2:**  
(OPEN) DATA MANAGEMENT IS OFTEN NOT WELL  
DEFINED AND IMPLEMENTED

**THERE IS A NEED FOR  
DATA GOVERNANCE**



**SOLUTION: DATA GOVERNANCE METHODS:**

- OPEN DATA DECISION TREE
- BOMOD





# (OPEN) DATA GOVERNANCE – DEMAND AND FIRST SOLUTIONS

Having access and actually being able to open and use data is essential when talking about creating value with data. Although this sounds trivial, availability and technical compatibility are still an issue when working with many open datasets. Data governance can help data owners to better manage their data throughout the process of publishing and re-using data. The TNO department Business Information Services has developed several methods and tools that support organizations to keep control over their data in order to be able to realize value from it.

One of the tools that have been developed is an open data decision tree. It helps organizations to structurally analyze which data could possibly be published as open data. During the analysis several aspects (ownership, privacy, economic value, data quality and technical format) are checked to decide if they form barriers for publishing the data and how interventions can help to overcome these barriers.

BOMOD (Beheer- en OntwikkelModel Open Data) is the outcome of another project. BOMOD defines 22 activities that help organizations with managing their data from a strategic, tactical, operational, implementation support and communication perspective. These activities are prioritized

using a maturity model. Each activity is described in the form of step-by-step guidelines and practical tools.

## CONTACT

**Silja Eckartz**

[silja.eckartz@tno.nl](mailto:silja.eckartz@tno.nl)

Business Information Services

ECOSYSTEM



## HOW AND UNDER WHICH CONDITIONS DO ONLINE CANCER AWARENESS CAMPAIGNS HAVE POSITIVE OFFLINE BEHAVIORAL EFFECTS?



## #DATAGRANT:

**TWEETS FROM 9 CAMPAIGNS, ABOUT 6 CANCER TYPES IN 4 LANGUAGES SINCE 2008**

# CANCER AWARENESS CAMPAIGNS ON TWITTER

Cancer is arguably the most pressing contemporary health issue worldwide: about every minute 15 people in the world die of cancer. One third of these cancers deaths could be prevented by research, screening, lifestyle change and vaccinations. According to the WHO the number of cancer deaths is still on the rise. In the battle against cancer, many health organizations use online campaigns to raise awareness, acquire funding and stimulate prevention and early detection behavior among citizens. Yet, according to leading scholars, the research into the effect of online campaigns on offline behavior is still in a very early stage. Therefore, the research question this project aims to answer is: “How and under which conditions do *online* cancer awareness campaigns have positive *offline* behavioral effects?”

Our research team received a Twitter data grant to study the diffusion process and effectiveness of cancer awareness campaigns. We analyze archival Twitter data from 9 campaigns (e.g. Movember and Pink Ribbon), covering 6 cancer types from 20 countries over a 6-year time period. First, we map the diffusion process in detail: what key events and actors accelerate or decelerate the spreading of these campaigns? Second, we assess the effect of the campaigns on the frequency and sentiment of tweets about a particular type of cancer. Last, we

determine the campaigns’ offline effects: when do online cancer awareness campaigns result in offline behavior, e.g. more donations to cancer research? Academically, our research adds to previous research about healthcare awareness campaigns. Practically, our research helps healthcare organizations, patients, and (social) enterprises to improve their online cancer awareness campaigns.

## CONTACT

**Tijs van den Broek**

tijs.vandenbroek@tno.nl

Strategy & Policy

## ANALYSIS





**TNO** innovation  
for life

**TNO.NL**