Strategic cooperation

Creating education programs

The data scientist of the future

Data science = capturing reality

Technische Universiteit Eindhoven
University of Technology
Dear reader,

These are exciting times! Not only scientific research, but also business processes, healthcare, mobility, communication and everything else in our daily lives are influenced by today’s abundance of data. Data are the fuel driving innovation. Without recent breakthroughs in data science, we would not be able to use this fuel in an effective and scalable manner.

The Data Science Center Eindhoven (DSC/e) is the epicenter of data science research in the Brainport region with international visibility and some of the best data science researchers in the world. This first issue of the DSC/e magazine highlights some of the ongoing activities related to research, education, and collaboration with industry. Enjoy reading!

Wil van der Aalst,
Scientific Director DSC/e

Table of Content
04. Calendar
05. Data science is here to stay!
06. Data science = capturing reality
Wil van der Aalst
09. Data science in quotes
10. Creating education programs
An adventure
12. TU/e’s response to the growing volume and importance of data
14. The data scientist of the future
16. Strategic cooperation Philips and TU/e: New insights on how to improve people’s lives
19. Picture this
20. Data science and psychology go hand in hand
22. ‘The nerd is going to shape our world’
23. DSC/e in short
Data science is here to stay!

The Data Science Center Eindhoven (DSC/e) was founded in 2013 by the Eindhoven University of Technology (TU/e). Its goal is to set up a world leading research program in data science. The center is run as a doctoral school with a scientific research program consisting of seven major research programs:

- Process analytics
- Customer journey
- Smart maintenance & diagnostics
- Quantified self
- Data value and privacy
- Smart cities
- Smart grids

DSC/e research contributes to the challenges of the TU/e Thematic Research Areas: Health, Energy, and Smart Mobility. Each of these areas witnesses a rapidly growing volume of data triggering a variety of scientific and societal challenges.

Currently four university departments work together in close cooperation on the research programs of the DSC/e:

- Department of Mathematics & Computer Science
- Department of Industrial Engineering & Innovation Sciences
- Department of Industrial Design
- Department of Electrical Engineering

Professor Wil van der Aalst is the scientific director of the DSC/e. Given the empirical nature of data science, DSC/e collaborates with a wide range of organizations. Collaborations include larger joint research projects, PhD projects, master projects and contract research.

For more information, visit our website www.tue.nl/dsce. Please contact us via email dsce@tue.nl or telephone the DSC/e program manager Maurice Groten at +31 (0)40 247 8639 with specific questions or remarks.
Prof.dr.ir. Wil van der Aalst
Wil van der Aalst is a full professor at the Department of Mathematics and Computer Science of Eindhoven University of Technology (TU/e). He is also the scientific director of the Data Science Center Eindhoven. His personal research interests include business process management, process mining, Petri nets, workflow management, process modeling, and process analysis. Wil van der Aalst has published more than 380 journal papers, 18 books (as author or editor), 400 refereed conference/workshop publications, and 60 book chapters. He is one of the most cited computer scientists in the world.

Society, organizations, and people are ‘always on’

The data scientist is the engineer of the future. The ultimate goal is not to collect more data, but to turn it into real value. Organizations that are unable to use (big) data in a smart way will not survive. These are some strong statements made by Wil van der Aalst, scientific director of the Data Science Center Eindhoven (DSC/e). Statements that he can make, for he is one of the most cited computer scientists in the world. His ideas have influenced researchers, software developers, and standardization committees. DSC/e Magazine introduces Wil van der Aalst by means of six statements on data science.

Data is collected about anything, at any time, at any place. The ultimate goal is not to collect more data, but to turn it into real value.

“Society shifted from being predominantly ‘analog’ to ‘digital’ in just a few years. This has an incredible impact on the way we do business and communicate. Society, organizations, and people are ‘always on’. The term Big Data is often used to refer to the incredible growth of data in recent years. We can turn that into real value. Data should be used to improve existing products, processes and services, or enable new ones.

Data Science is a new engineering discipline and the main driver for innovation in the years to come. "Just like computer science emerged as a new discipline from mathematics when computers became abundantly available, we now see the birth of data science as a new discipline driven by the torrents of data available today. We believe that the data scientist will be the engineer of the future. Innovations will be data driven. Also, scientific research is becoming more data driven. Therefore, TU/e established the Data Science Center.”

Organizations that are unable to use (big) data in a smart way will not survive.

“It is not sufficient to focus on data storage and data analysis. Data scientists in organizations also need to relate data to process analysis. For that, process mining is very important. It bridges the gap between traditional model-based process analysis (simulation and other business process management techniques) and data-centric analysis techniques such as the mining of event data.”

Event data are the most important source of information. Events may take place inside a machine (such as an X-ray machine or baggage handling system), inside an enterprise information system (an order placed by a customer), inside a hospital (the analysis of a blood sample), inside a social network (exchanging e-mails or Twitter messages), inside a transportation system (checking in, buying a ticket, or passing through a toll booth), etcetera. Events may be life events, machine events, or both.

Automatically discovered process model based on hospital data. The model reveals the real treatment process of a particular group of patients.
Everyone knows data is the new black
Jonathan Goldman, former principal data scientist at LinkedIn
I keep saying that the sexiest job in the next 10 years will be statisticians, and I’m not kidding Hal Varian, chief economist at Google

Data! Data! Data!
He cried impatiently. I can’t make bricks without clay!
Sherlock Holmes, The Adventure in the Copper Beaches

Data science is the profession of the future.
Organizations that are unable to use (big) data in a smart way will not survive
Wil van der Aalst, scientific director Data Science Center Eindhoven

In Philips, data science is part and parcel in business but as well as how we run the organization and in fact, we would be losing out on our competitive advantage if we would not be using data science
Bart Luijten, CIO Philips

It took a long time for people to recognize that data science is something that we need to improve on
Gert Streutker, ASML

I am looking for the people which data mine the data we already have
Roland Schneiders, Program Manager SaaS at Cofely GDF Suez

We recognized the opportunity to develop new products and services that use big data
Jeroen Tas, CIO Philips

as machine learning and data mining. Process mining seeks the confrontation between event data (observed behavior) and process models (hand-made or discovered automatically). This technology has become available only recently, but it can be applied to any type of operational processes (organizations and systems). Example applications include: analyzing treatment processes in hospitals, improving customer service processes in a multinational, understanding the browsing behavior of customers using a booking site, analyzing failures of a baggage handling system, and improving the user interface of an X-ray machine. All of these applications have in common that dynamic behavior needs to be related to process models. Hence, it is no longer acceptable to just look at processes and data in isolation.

Expertise in data- and process mining, probability, statistics, algorithms, privacy, stochastics, and visualization are vital.

“Data science aims to answer questions such as ‘what happened (reporting)’, ‘why did it happen (diagnosis)’, ‘what will happen (prediction)’ and ‘what is the best that can happen (recommendation)’. To adequately answer such questions there is not just a need for raw data and more computing power. Expertise from different disciplines is needed to transform data into actionable information and provide real value to people and organizations.”

When analyzing data, we cannot stop time.

“Data seems static but is often about people, machines or organizations that change constantly. The moment you start to analyze data, the data can already be outdated. Mining techniques can be used to describe the behavior of people, machines and organizations, but processes and data populations change over time. The challenge is to cope with these dynamics. In essence, what we do with process mining is trying to capture reality. We’re not interested in behavior of people in the past, we want to know the behavior of people in the present or rather: in the future. If we want to stimulate people to change their lifestyle or make machines reliable under changing circumstances, we have to focus our data analysis on the future, not the past.”

“Data can answer questions that have never been asked before. When we visualize data in a smart way, we can generate questions that have never been asked before. It is very hard, but it can be done. Normally, we ask questions and come up with a hypothesis and only then analyze data that supports (or rejects) the hypothesis. Visualization of data can help. It is amazing how we are capable of discovering remarkable patterns. Within the DSC/e we have people doing groundbreaking research on this topic.”

By replaying event data on an automatically discovered process model, it is possible to see the most frequent paths, deviations, and bottlenecks.
Eindhoven University of Technology (TU/e) creating education programs in data science

An adventure: be a whole new Department of Data Science.

Several master, PDEng and PhD programs. Eventually, there might even be a bachelor program in data science in September 2016, together with corporations, public authorities and more. The goal is to start a broad education program that is broad and fine-tuned to allow for specialization. It will have a lot to offer to many different students.

Bachelor and masters Starting September 2016, a broad bachelor program will be available at both Eindhoven and Tilburg University. In both cities, the program will be roughly the same. In addition to the basic components of any TU/e bachelor (calculus, design, modeling, etc.), the program will offer students a solid basis in the technical side of data science (statistics, programming, visualization, data mining), but will also include relevant courses in areas such as business analytics, cognitive science, and law. In addition, there will be plenty of room for practical application and integration of the course content, both within the individual courses, and through exciting Data Challenges and, of course, students’ final projects.

After students have graduated the bachelor program, they can continue studying data science in one of three available master programs:

- Data science & engineering (Eindhoven)
- Data science & society (Tilburg)
- Data science & entrepreneurship (Den Bosch)

In the master programs, an additional city is involved: Den Bosch. There, students study at former convent Marienberg, where they can also live on campus. The education program there will be set up in close collaboration with industrial partners. They will provide masterclasses, internships and cases students can work on. IJsselsteijn: “Creating these new master programs is very exciting. I hope the master program in Den Bosch, for example, will result in the next Google, Amazon, Netflix, or Blendle. After all, those are all data driven companies as well.”

PDEng: creating innovative solutions TU/e now offers several technology design postmaster PDEng programs (Professional Doctorate in Engineering) for candidates with a university engineering degree at master level. Their academic, research oriented education, together with a two-year training program focusing on the application of recent research results to practical, complex industrial problems, allows them to create innovative solutions for these problems. Seeing its many application aspects, TU/e also works on a new PDEng program in data science.

In the PDEng program, different elements will be integrated. The two-year interdisciplinary program combines mathematics, statistics, and computer science theory with the business acumen to explore data sets, gather insights, visualize results, and communicate meaningful findings. Graduates make sense of data and have the ability to articulate their discoveries and recommendations to those not schooled in the world of data in the frame of industrial and business design and decision processes. “A successful data science professional has technical and soft skills at a high level and has integrated them optimally”, says Stef van Eijndhoven, associate professor at TU/e. “In the creation of the new PDEng program in data science, that should be a main objective.”

Cooperating with Tilburg University “In developing the new education programs, we are on an adventure with Tilburg University”, says Wijnand IJsselsteijn, professor at TU/e and one of the founding fathers of the DSC/e. “Together, we have a broad perspective on data science, bringing together a unique combination of alpha, beta and gamma sciences. In Eindhoven the focus is more on the technical side, Tilburg brings in valuable knowledge on topics like law, governance, and larger societal perspectives, while both parties closely collaborate in areas such as psychology, ethics, and entrepreneurship. Together we can build an education program that is broad and complete, with additional electives and assignments to allow for specialization.

TU/e creates data science & society Eindhoven the focus is on data science, bringing together a unique combination of alpha, beta and gamma sciences. In Eindhoven the focus is more on the technical side, Tilburg brings in valuable knowledge on topics like law, governance, and larger societal perspectives, while both parties closely collaborate in areas such as psychology, ethics, and entrepreneurship. Together we can build an education program that is broad and complete, with additional electives and assignments to allow for specialization.

For more information on the new education programs, please contact Edwin van den Heuvel (e.r.v.d.heuvel@tue.nl).
The Data Science Center Eindhoven (DSC/e) is TU/e’s response to the growing volume and importance of data. Its goals: to set up a world leading research program in data science and to provide a platform for data science knowledge dissemination amongst partners in industry, public authorities and academia.

Globally, data is growing Ninety percent of the data in the world today has been created in the last two years alone. The total volume of data will grow by 50 times in the next 10 years. Moreover, human and organizational activities are intertwined with the digital universe. Therefore, data science is growing in importance and becoming an integral part of different types of engineering and scientific research. Many of our grand societal challenges can be addressed by exploiting the data at hand. Data analytics provide meaningful solutions to problems in a large variety of domains, such as manufacturing, banking, logistics, mobility, sustainability and health and wellbeing. Therefore, opportunities to turn data into value are everywhere.

The birth of DSC/e TU/e has gained a lot of experience on data science over the years. We have some of the best scientists working and educating in the broad field of data science. As a logical result of that, in 2013 the TU/e established the DSC/e.

Research areas
In line with TU/e policy, DSC/e research contributes to the challenges of the TU/e Thematic Research Areas: Health, Energy, and Smart Mobility. The research is conducted within the TU/e, and the Data Science Center also closely collaborates bilaterally with companies, such as Philips and ASML, with corporations and public authorities and in national or European research projects. The center has its doors wide open for new alliances and joint programs on scientific research with companies and organizations within the brainport region and beyond.

Buzzing with new possibilities
From its birth in 2013 until now, the DSC/e has been buzzing with work, research, new possibilities and new alliances. At the annual Data Science Summit, leaders of industry and leading data scientists share their latest insights. Key speakers share their thoughts and knowledge at the Data Science Lecture Series, that are freely accessible. And of course, interesting research results have been presented in reports and articles. There are many examples ranging from large-scale industry collaborations to individual Master projects. Consider for example the Master project by Marijn of M. Grootjans: he made a tool to visualize a World Cup soccer game, based on data collected during that soccer game.

A promising future
Looking at the (near) future, the DSC/e promises interesting new developments. The center will continue to develop the seven research programs. As data will grow, we are ready for the future. The Data Science Center Eindhoven is ‘always on’. Welcoming new research challenges, partnerships, and opportunities. You will hear from us.

Departments
Four TU/e departments are involved in the DSC/e:
- Department of Mathematics & Computer Science
- Department of Industrial Engineering & Innovation Sciences
- Department of Industrial Design
- Department of Electrical Engineering

Research programs
The DSC/e is run as a doctoral school with a scientific research program consisting of seven major research programs:
- Process analytics
- Customer journey
- Smart maintenance & diagnostics
- Quantified self
- Data value and privacy
- Smart cities
- Smart grids
How do you prepare for that as a student? Isn’t it so that what you study now, will already be ‘old news’ next year?

‘I don't think so. New things will come, but there are also many existing techniques that we’ll be using to analyze data in the future. What really is changing, are the sources of data. Almost every day we create new sources of data. The internet, social media, but also organizations that keep track of what their clients do, provide us with so much data that can tell us so many things.’

How do you see the future?

‘I believe that innovation will be much more driven by data. Companies already use data to improve their products, and they will continue to do so. For me, I don’t know where I will be working three years from now but I would like to do something that has to do with the behavior of people. You can use data to predict behavior. For example, when data analysis shows that there are a lot of people on a certain highway at a certain time, you can improve the infrastructure to accommodate that. And that can be life changing.’

What qualities does the data scientist of the future need to have?

‘A good data scientist has to be a ‘people person’ as well as a ‘technical person’. You have to know what data is usable, and what isn’t. And because the field of data science is still so new, you have to look in all possible directions. It’s like we’re standing at a crossroads. We have to look ahead of us, behind us, left and right. But it’s not only that. What’s important, is that we also look at the roads that are ‘hidden’. Maybe you can even walk through a building that’s standing on a road, and leave that building again through the backdoor, having found new information and insights on your way.

So, I will keep my eyes wide open and discover as much as I can.’

The data scientist of the future

‘We are standing at a crossroads’

He is the data scientist of the future: Jeroen van de Ven (24). Open to anything, very driven and curious of what’s going to happen in the wide field of data science in the years to come. Although he is studying a data science track at the TU/e now, no one really knows what Jeroen’s future professional career will look like exactly. Data science is relatively new and young, but moving at an unbelievable high speed. How does a student prepare himself for that? A little chat about what is waiting around the corner.

Why do you study data science?

‘I’ve always had a great interest in computers and programming. So I did a bachelor in Computer Science and I’m now doing a master in Computer Science and Engineering at TU/e. The stream ‘Data Science’ is part of the master. It’s very interesting to see all the developments in the field of data science. At this moment in time, we don’t really know what will be the standard techniques everyone uses ten years from now. That makes data science such an interesting topic.’

How do you prepare for that as a student? Isn’t it so that what you study now, will already be ‘old news’ next year?

‘I don’t think so. New things will come, but there are also many existing techniques that we’ll be using to analyze data in the future. What really is changing, are the sources of data. Almost every day we create new sources of data. The internet, social media, but also organizations that keep track of what their clients do, provide us with so much data that can tell us so many things.’

How do you see the future?

‘I believe that innovation will be much more driven by data. Companies already use data to improve their products, and they will continue to do so. For me, I don’t know where I will be working three years from now but I would like to do something that has to do with the behavior of people. You can use data to predict behavior. For example, when data analysis shows that there are a lot of people on a certain highway at a certain time, you can improve the infrastructure to accommodate that. And that can be life changing.’

What qualities does the data scientist of the future need to have?

‘A good data scientist has to be a ‘people person’ as well as a ‘technical person’. You have to know what data is usable, and what isn’t. And because the field of data science is still so new, you have to look in all possible directions. It’s like we’re standing at a crossroads. We have to look ahead of us, behind us, left and right. But it’s not only that. What’s important, is that we also look at the roads that are ‘hidden’. Maybe you can even walk through a building that’s standing on a road, and leave that building again through the backdoor, having found new information and insights on your way.

So, I will keep my eyes wide open and discover as much as I can.’

Jeroen van de Ven

• Student at TU/e
• Completed the bachelor program in Computer Science at TU/e
• Now studying master program in Computer Science and Engineering
• Chose the stream Data Science
• Fascinated by data science and the behavior of people
In 2014, Philips and the TU/e started a long-term strategic cooperation. Researchers of Philips and PhD students of TU/e work together on scientific and industrial challenges in the fields of data science, health and lighting. Currently 16 PhD students associated with TU/e’s Data Science Center Eindhoven work on data science topics. The so-called ‘Flagship project’ is a success and already starts to show the first results.

The Data Science Center Eindhoven provides PhD students with the opportunity to carry out their four-year research project at Philips. The projects are divided into four streams in the fields of data science and health: Data Driven Value Proposition, Health Care Smart Maintenance, Optimizing Healthcare Workflows, and Continuous Personal Health.

Digital innovations

With the scientific research that is conducted in the projects, Philips and TU/e aim to come up with digital innovations in healthcare and data science. The innovations will focus on bringing solutions to grand societal challenges. Such challenges include affordable access to high-quality healthcare, the prevention of cardiovascular diseases and the treatment of various forms of cancer.

‘We are discovering things we have never seen before’

Ronald Begeer, program manager at Philips, is very happy with the alliance and the first results that are already showing. “Our researchers are working closely with the PhD students of TU/e. It’s very nice to see that all those involved form a successful, multi-disciplinary team. There’s a lot of knowledge in the team. We don’t have to wait four years to see results: already we get new insights, things we can really use to improve our products and support our clients. Now, we really have the people and the time to discover things we have never seen before. It’s very exciting.”

New insights

Ronald Begeer, program manager at Philips, is very happy with the alliance and the first results that are already showing.

Long term relationship

In the years to come, Philips and TU/e will continue these four streams and possibly add more. “It’s a long-term relationship. The topics we are working on now are very important to us, now and in the future. We’ll hire more PhD’s and keep the scientific research going. Data science will only become more and more important in the way we do our business.”

PhD projects

PhD projects are research projects carried out by young researchers with a master degree. They work dedicatedly on a specific, long-term research problem. The research is performed under the joint supervision of the senior researchers from both the university and Philips Research. PhD projects take four years.

The stream Data Driven Value Proposition is about the development of new business models for smart internet-based services, such as a personalized device to measure your blood pressure.

With Health Care Smart Maintenance Philips wants to improve the maintenance processes of medical machines (scanners and X-ray machines, for example) with the help of user data analysis.

Continuous Personal Health has to result in improvements in personal care in different stages of healthcare: prevention, diagnosis, treatment, aftercare.

Optimizing healthcare workflows looks at how workflows in hospitals can be improved. Begeer explains: “Machines used in hospitals, health care protocols and work flow processes all provide us with an enormous amount of data. With the analysis of that data, the researchers try to find answers to questions like ‘how can we improve and accelerate the treatment of a patient?’ and ‘how can you use the data to predict the steps in the treatment of a patient?’”

Strategic cooperation Philips and TU/e:
New insights on how to improve people’s lives

In 2014, Philips and the TU/e started a long-term strategic cooperation. Researchers of Philips and PhD students of TU/e work together on scientific and industrial challenges in the fields of data science, health and lighting. Currently 16 PhD students associated with TU/e’s Data Science Center Eindhoven work on data science topics. The so-called ‘Flagship project’ is a success and already starts to show the first results.

The Data Science Center Eindhoven provides PhD students with the opportunity to carry out their four-year research project at Philips. The projects are divided into four streams in the fields of data science and health: Data Driven Value Proposition, Health Care Smart Maintenance, Optimizing Healthcare Workflows, and Continuous Personal Health.

Digital innovations

With the scientific research that is conducted in the projects, Philips and TU/e aim to come up with digital innovations in healthcare and data science. The innovations will focus on bringing solutions to grand societal challenges. Such challenges include affordable access to high-quality healthcare, the prevention of cardiovascular diseases and the treatment of various forms of cancer.

New insights

Ronald Begeer, program manager at Philips, is very happy with the alliance and the first results that are already showing. “Our researchers are working closely with the PhD students of TU/e. It’s very nice to see that all those involved form a successful, multi-disciplinary team. There’s a lot of knowledge in the team. We don’t have to wait four years to see results: already we get new insights, things we can really use to improve our products and support our clients. Now, we really have the people and the time to discover things we have never seen before. It’s very exciting.”

Digital innovations

With the scientific research that is conducted in the projects, Philips and TU/e aim to come up with digital innovations in healthcare and data science. The innovations will focus on bringing solutions to grand societal challenges. Such challenges include affordable access to high-quality healthcare, the prevention of cardiovascular diseases and the treatment of various forms of cancer.

‘We are discovering things we have never seen before’

Ronald Begeer, program manager at Philips, is very happy with the alliance and the first results that are already showing. “Our researchers are working closely with the PhD students of TU/e. It’s very nice to see that all those involved form a successful, multi-disciplinary team. There’s a lot of knowledge in the team. We don’t have to wait four years to see results: already we get new insights, things we can really use to improve our products and support our clients. Now, we really have the people and the time to discover things we have never seen before. It’s very exciting.”

Long term relationship

In the years to come, Philips and TU/e will continue these four streams and possibly add more. “It’s a long-term relationship. The topics we are working on now are very important to us, now and in the future. We’ll hire more PhD’s and keep the scientific research going. Data science will only become more and more important in the way we do our business.”

PhD projects

PhD projects are research projects carried out by young researchers with a master degree. They work dedicatedly on a specific, long-term research problem. The research is performed under the joint supervision of the senior researchers from both the university and Philips Research. PhD projects take four years.

The stream Data Driven Value Proposition is about the development of new business models for smart internet-based services, such as a personalized device to measure your blood pressure.

With Health Care Smart Maintenance Philips wants to improve the maintenance processes of medical machines (scanners and X-ray machines, for example) with the help of user data analysis.

Continuous Personal Health has to result in improvements in personal care in different stages of healthcare: prevention, diagnosis, treatment, aftercare.

Optimizing healthcare workflows looks at how workflows in hospitals can be improved. Begeer explains: “Machines used in hospitals, health care protocols and work flow processes all provide us with an enormous amount of data. With the analysis of that data, the researchers try to find answers to questions like ‘how can we improve and accelerate the treatment of a patient?’ and ‘how can you use the data to predict the steps in the treatment of a patient?’”
Governments of third world countries are coping with a great shortage of data they can use in building their policy. The analysis of data provided by mobile phones can be a solution, according to Stef van den Elzen, software engineer at SynerScope and PhD student at TU/e.

In the context of the Data for Development (D4D) challenge Van den Elzen analyzed the data of 2.5 billion phone calls and text messages in Ivory Coast, Africa. With a smart combination of analysis and visualization he connected social events to changes in telephone communications. By doing that, he won the prize for ‘Best visualization’. Who knows, in the near future, we can use such visual analysis to predict big social changes.

Bart Hompes: “At the moment, we are looking at workflows involved in radiology and interventional X-ray in hospitals. The data we use in our process mining, is recorded by machines and support systems. We have a lot of data, but the challenge is that the data is quite sparsely distributed, as each patient follows a unique process. We have to link several sources and identify related cases to get to the end-to-end process instead of just generating the data for one machine or one sub-process or procedure.”

Real data, real processes, real value

What Bart finds most exciting, is the fact that his group of researchers can do evidence-based recommendations to improve the workflows in healthcare. “We are working with real data and real processes, which is provided to us by machines in hospitals, logs and people. Our recommendations won’t be based just on conceptual ideas, but on real data and people. This way, we can add value.”

The goal of the project Bart is involved in is to make evidence-based recommendations to improve the end-to-end workflows in terms of cost, efficiency, speed, reliability, and conformance. “We want to learn from the past and compare different process variants and instances in order to obtain deeper insights into the processes at hand.”

New techniques needed

For Bart, it is interesting to see that the existing techniques used in process mining are not always suitable for process mining in the field of healthcare. “The existing techniques work well in structured processes, like applying for a credit card for example. In that process, the steps each applicant takes are the same. But in healthcare, that’s not the case. Every step a patient takes, depends on the step that was taken before that, or on which doctor is involved. The ‘structure’ in the data is unique for every patient. That makes the process mining and the analysis of the data very complex. We have to create new techniques to get good results we can actually use.”

Real data, real processes, real value

What Bart finds most exciting, is the fact that his group of researchers can do evidence-based recommendations to improve the workflows in healthcare. “We are working with real data and real processes, which is provided to us by machines in hospitals, logs and people. Our recommendations won’t be based just on conceptual ideas, but on real data and people. This way, we can add value.”

The goal of the project Bart is involved in is to make evidence-based recommendations to improve the end-to-end workflows in terms of cost, efficiency, speed, reliability, and conformance. “We want to learn from the past and compare different process variants and instances in order to obtain deeper insights into the processes at hand.”

New techniques needed

For Bart, it is interesting to see that the existing techniques used in process mining are not always suitable for process mining in the field of healthcare. “The existing techniques work well in structured processes, like applying for a credit card for example. In that process, the steps each applicant takes are the same. But in healthcare, that’s not the case. Every step a patient takes, depends on the step that was taken before that, or on which doctor is involved. The ‘structure’ in the data is unique for every patient. That makes the process mining and the analysis of the data very complex. We have to create new techniques to get good results we can actually use.”

Make an impact

“I love that with my work, I can really make an impact. The new techniques we come up with, the things we find out, can change people’s lives. For Philips, the work we do already provides good insights on how their processes work, what happens around them and how they can be improved.”

Bart Hompes is one of the 16 PhD students that work on the data science research projects in cooperation with Philips. He started his research in November 2014 and focuses on ‘Process Mining for Healthcare Workflows’, part of the stream ‘Optimizing Healthcare Workflows’. 

Bart Hompes: ‘Our work can change people’s lives’
When we think of data science, subjects that come to mind are mathematics, computer science, and statistics. However, as the sciences of understanding human behavior, psychology and behavioral science also make a number of fundamental contributions to data science. Wijnand IJsselsteijn, neuro- and media psychologist, professor at TU/e, and one of the founding fathers of the Data Science Center Eindhoven (DSC/e), explains.

**Data science and psychology go hand in hand**

*In essence, data science is a lot about tracking human behavior in (near) real-time and at a large scale, such that it improves our understanding of the internal and external forces that shape human behavior, and enables meaningful action*, says Wijnand IJsselsteijn. "As such, data science is not so much about the sheer quantity of data that can be captured, processed, and accessed. It is rather about the capacity to search, aggregate, and cross-reference data, such that relational inferences can be drawn."

**Insight, guidance and feedback**

“Psychology makes a number of fundamental contributions to data science. First, psychology helps to create insight into the kinds of behavior that are most fruitful to track and analyze. Secondly, it can provide theoretical and methodological guidance on behavioral inferences that can be drawn, distinguishing spurious correlations that frequently emerge in large datasets from meaningful behavioral patterns. Thirdly, psychology can help to provide users with psychologically meaningful and actionable feedback, and to elicit sustainable behavior change: taking action on the basis of data.”

**The other way around**

“In turn, data science has important contributions to psychological research. The possibility of large-scale online and mobile data acquisition is redefining the traditional notion of a psychology lab. Tools and techniques from information sciences (network analysis) can complement existing methods in the measurement and modelling of psychological processes. Analysis of online behavior of large communities (Tweeting, picture uploads, Facebook shares, behavior in multiplayer online games) allows research into, for example, language, personality and mood, at an unprecedented scale.”

**Quantified self**

IJsselsteijn is involved in many research projects in the fields of data science and behavioral science, also associated with the DSC/e. One thing he, his students and colleagues are intrigued by is, ‘quantified self’: individuals collecting data about themselves by using devices. "It’s a big trend that started in Silicon Valley and that will only get bigger in the years to come", IJsselsteijn points out. "Smartphones, smart watches and personal trackers provide us with the possibility to collect data on literally every step we take. Think of a tracker that counts how healthy your diet is. That is all about collecting data and using that data to change your behavior. For me as a social scientist, it is very interesting to see the effect it has on people. There are people to whom personal tracking is just a very useful way of quantifying their performance and health habits. But it appears that some people will also change their behavior just to get more positive feedback from the tracker they use. We have to be careful that those trackers don’t take over our life. For example: a friend of mine uses a tracker that tracks his physical behavior. The device rewards him for walking and running, but not so much for riding a bike. To get better results on his tracker, my friend stopped riding his bike. That is the kind of behavioral change we don’t want to see. But in turn, for me it’s a very interesting research topic. Personal tracking is a fascinating phenomenon, as is data science related to psychology and behavioral science in general. There’s a whole new world to discover."
Professor statistics:

‘The nerd is going to shape our world’

The Data Science Center Eindhoven (DSC/e) can’t exist without having a group of people involved, with different backgrounds and a great understanding of data science. People who know the world will never be the same again since data, the way we look at it, and the way we use it, became so important. Edwin van den Heuvel, professor of Statistics at the TU/e department of Mathematics and Computer Science, is one of them. Since the start of his job at TU/e last October, he is closely involved in the development of DSC/e.

Before Van den Heuvel came to TU/e, he was a consultant in industrial statistics, and head of the statistics department at the pharmaceutical company MSD. But, to ‘satisfy his hunger for research’, as he explains it, he took the opportunity to become a professor. First at the University Medical Center Groningen and now full time in Eindhoven.

New techniques, new people

With the new large data sets we have at our disposal, we also need new statistic techniques, and new, well trained people who can develop them. “That is why I am also involved in the development of new education programs in data science. TU/e is working on a bachelor program and master programs in data science. The data scientist of the future has to be someone with a knowledge of many different things: statistics, computer science, but also psychology and sociology. He or she has to have good communication skills. You can’t be a data scientist living in an ivory tower. A data scientist has to be a bit of a ‘nerd’ as well. What we call a nerd now, is the success candidate of the future. The nerd is going to be very influential, because data scientists are the ones that are going to shape our world.”

Data science and statistics go hand in hand

“Without statistics, data science isn’t complete”, says Van den Heuvel. “In this day and age, we have enormous amounts of data that we can analyze and learn from. But I think we still don’t use all that data the way we should. If we really want to use data to improve certain products, or to make our lives easier, you simply have to use statistical models and methods. It is crucial to understand and model data.”

‘You can’t be a data scientist living in an ivory tower’

New alliance, new possibilities

Alliances with other businesses, organizations and educational institutes bring new possibilities. That is one of the reasons that the Data Science Center Eindhoven joined the international Big Data Alliance. The alliance is born out of the idea that Big Data & Business Analytics parties in academia and in the industry could benefit more from each other. The BDA focuses on four groups: students, industry, entrepreneurs and researchers. The members include Erasmus University Rotterdam, Tilburg University and Amsterdam Data Science. Together, the members want to help students to find the best fitting education and training programs, and to help the industry to find academic expertise to do joint projects. They also help entrepreneurs to start or advance their own enterprise and they help researchers in finding complementary expertise.

European Data Forum 2016 hosted by TU/e

TU/e is proud to be chosen as the university that will host and organize the European Data Forum (EDF) in 2016. Partners in the organization amongst others are of Amsterdam and VU University Amsterdam. This year, the event is held in Luxembourg. The 2016 version of the EDF will welcome 600 participants from countries all over Europe. The EDF is a meeting for industry professionals, researchers, policymakers and members of community initiatives to discuss the challenges of Big Data and the emerging Data Economy and to develop suitable action plans for addressing them. The range of topics discussed at the EDF ranges from novel data-driven business models, and technological innovations to societal aspects.

DSC/e member of Big Data Value Association

The Data Science Center Eindhoven recently became a member of the international Big Data Value Association. Members of the association comprise small and large European organizations from public and private sectors of business, industry, academia and research. The objectives of the association are to boost European big data value research, development and innovation and to foster a positive perception of big data value. It aims at strengthening competitiveness and ensuring industrial leadership of providers and end-users of big data value technology-based systems and services. It also promotes the widest and best uptake of big data value technologies and services for professional and private use. On top of that, the association establishes the excellence of the science base of creation of value from big data.