**Institute of Data Science**
**Maastricht University**

# 2018 Annual Report

**Table of Contents**


**Preface** **3**

**Ambition** **4**

**Roadmap** **6**

**Research** **8**

IDS Research Themes 8

Research programs & projects, publications and funding 9

Responsible automation of data-driven research 9

Clinical Data Science 10

Ethics, Law, and Social Implications of the Data-Driven Society 11

LawTech 12

NeuroTech 12

AgriTech 13

Recommendation Systems 13

**Education** **14**

**Research IT** **17**

Towards a research IT infrastructure at Maastricht University 17

**Community** **18**

University network 18

FAIR University Initiative 23

Community of Data Driven Insights 24

Regional network: 26

National network 27

VSNU Digital society 27

International network 27

**Governance** **28**

**Appendix A Data Science Research Seminars 2018** **33**

# Preface

Vast amounts of data are now being generated across all segments of society. These data offer an unprecedented opportunity to drive dramatic advances across academic disciplines, strengthen the very fabric of our society and improve the health and well being of citizens. Realizing such a vision requires a complete rethink of the role of data and its analysis in society. We must not only ensure that the data and the resources to innovate with the data are broadly and equitably accessible, but also that the development of new methods are done in a responsible manner.

Data Science is an emerging interdisciplinary discipline that is focused on obtaining insights from data. Data Science builds on theories and techniques from many fields including mathematics, statistics, information science, and computer science - particularly in data management, data mining, machine learning, and visualization. Data scientists leverage their expertise to find, retrieve, manage, clean, integrate, prepare, visualize, interpret, and build models from data while consciously navigating hardware, software, and bandwidth constraints along with social, legal, and ethical issues. Indeed, while Data Scientists exhibit interdisciplinary skills, they must also work with others to maximize their productivity and the impact of their work.

We established the Institute of Data Science in 2017 to foster a collaborative and interfaculty environment in scalable and responsible data science through multi-disciplinary team research, interdisciplinary training, and data-driven innovation. Our vision is that the Institute will act as an academic home for data scientists across campus and to foster cooperation across disciplinary boundaries. Over the past two years the Institute has a great number of successes to claim: a vibrant Data Science research seminar series, a multitude of community events held in collaboration with units across the university, acquisition of national and international research funding, first steps in creating interdisciplinary education, and the appointment of joint positions across career ranks. Looking forward into 2019, we have an ambitious program to increase our research excellence, strengthen interdisciplinary data science education, and continue to use data science to build bridges across and within faculties and with our community.

Yours sincerely,

Michel Dumontier
Distinguished Professor of Data Science
Director, Institute of Data Science
Maastricht University

# Ambition

A well known African proverb states that ‘it takes a village to raise a child’ – the proverb suggests that it takes an entire community of different people to interact with a child for it to grow up in a safe environment and to be in a position to reach their potential. We believe that Data Science at Maastricht University, which has been over 25 years in the making, can only fulfill its ambition of being a globally recognized force in Data Science when all stakeholders come together and meaningfully contribute to a shared vision. Fulfillment of the vision will require long term strategic investment for interdisciplinary activities, the synergistic alignment of teaching programs, the reorganization of research group structures, the coordination of marketing campaigns, and the cooperation of researchers and staff alike. The creation of the Institute of Data Science coupled with the establishment of the Distinguished Professor of Data Science, was only one step in this process.

Comprised of a growing number of staff and researchers, the Institute believes that a strong, vibrant, and rewarding academic environment is required to foster excellence in Data Science research and education, and to enable members of the Data Science community to work together to fulfill this mission and their own personal ambitions. As such, we have two strategic aims:

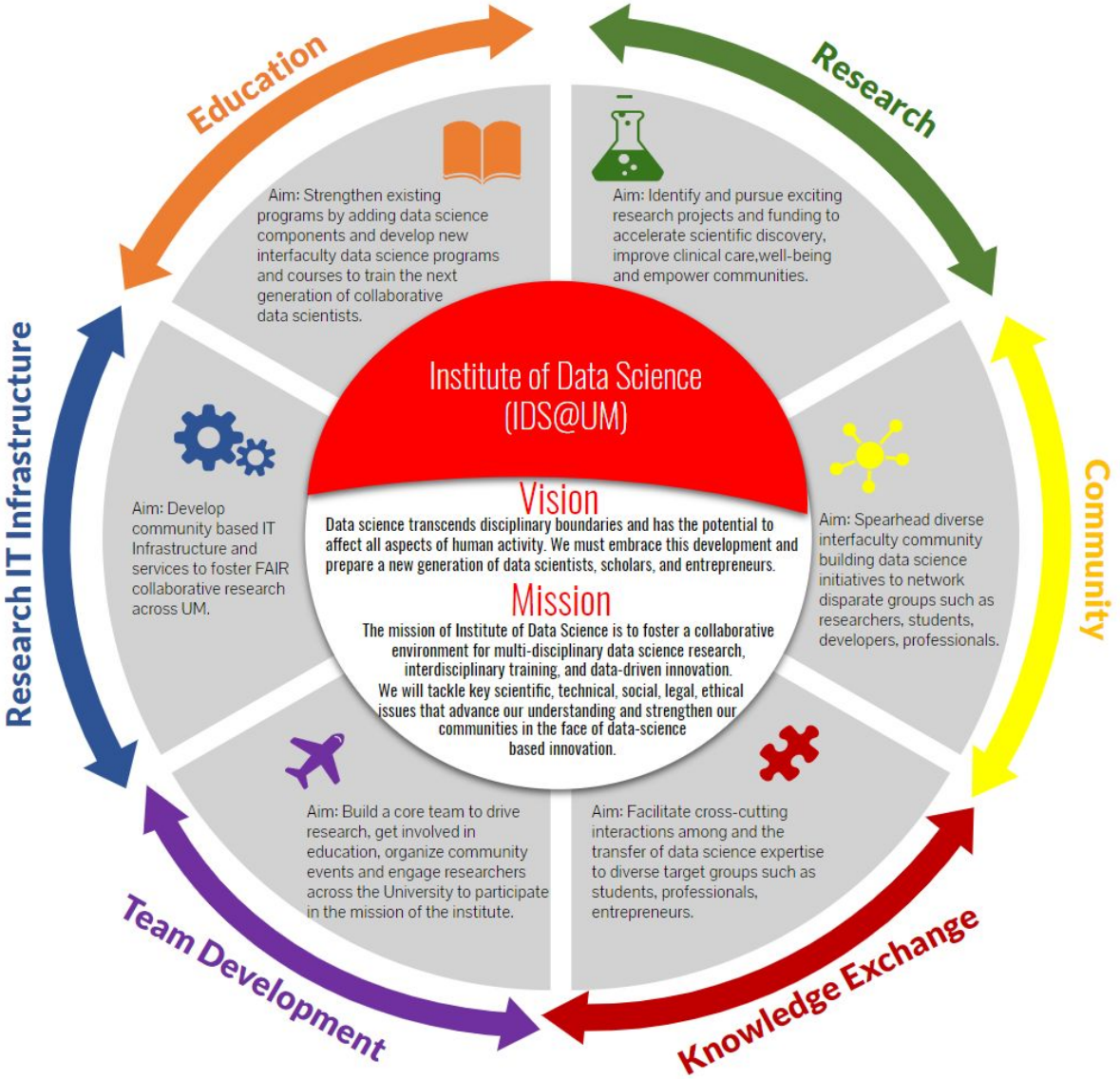
*(1) To coordinate, strengthen, and promote Data Science activities at Maastricht University.*

*(2) To spearhead new interfaculty initiatives targeting Data Science research, education, infrastructure, and innovation.*

Data Science is a subject of research and education in many faculties, departments, and research groups across Maastricht University. However, the lack of coordination and communication of their activities diminishes their potential prominence on an increasingly competitive world stage. Preliminary meetings with faculty and staff revealed that most self-identified data scientists are entirely unaware of each other’s contributions to research and education, whether in the same department or across departments. They recognize that their hard work is not getting the attention it deserves, which results in a missed opportunity to promote the Institution as a premier place for Data Science research and education. Concurrently, the lack of coordination of computing infrastructure to manage research data and workflows has resulted in a patchwork of non-interoperable systems that each cannot scale with anticipated growth. These represent only some of the threats to the data science activities at UM.

Our ambition is to spearhead groundbreaking interfaculty initiatives in Data Science research, education, infrastructure, and innovation. New incentives and collaborative structures are desperately needed because many faculty and staff want to demonstrate true societal impact – they want to go much further than just publishing a thesis that few will read, or publish papers that only academics can understand, or spend their valuable time only educating handfuls of specialists in a particular domain. Going beyond academia demands that contextualize our initiatives in a bigger picture – that we bring to the table individuals that can make contributions across the spectrum of data science – from infrastructure, to

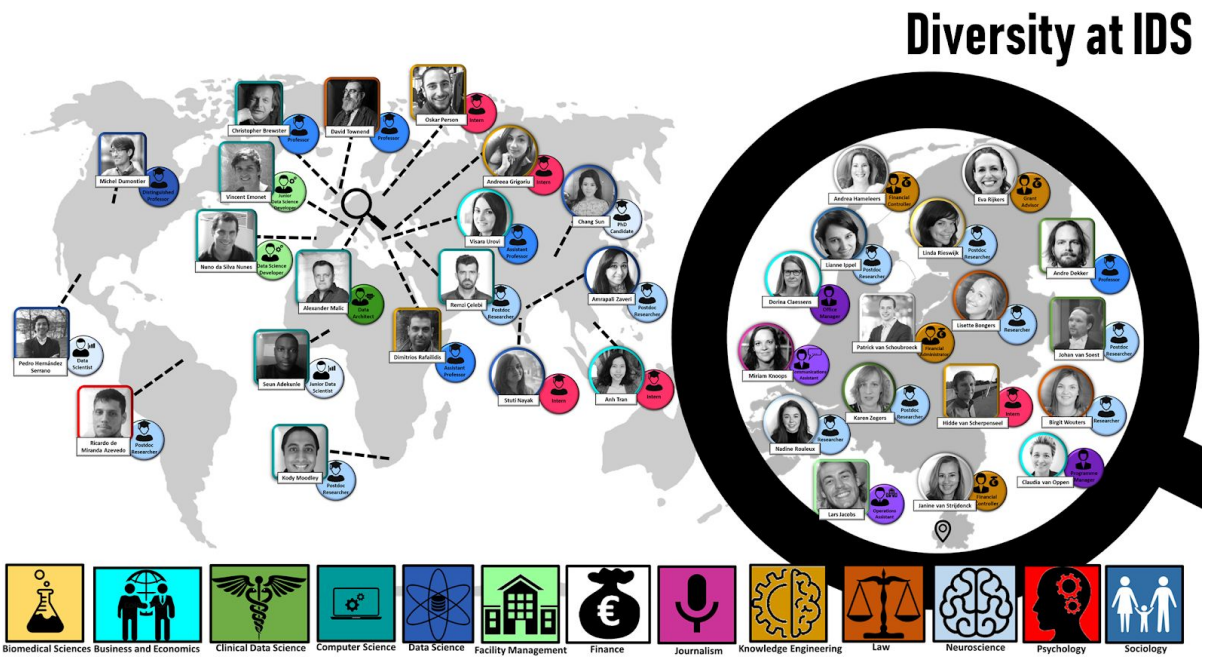
methods, to application areas, to social, legal, ethical, economic and political aspects. Thus, we aim to take this to heart and develop new frameworks to foster multi-disciplinary teams that can contribute to innovation to all of those aspects, and address, in a realistic manner, their application in society. To this end, we will work to strengthen multi-disciplinary team science, further develop interdisciplinary training opportunities, create new avenues for innovation by youth and professionals, and establish scalable infrastructure to effectively share and compute with data across the disciplines. An overview of our vision, mission, and aims is illustrated in the figure below.



# Roadmap

We have developed a roadmap to guide our work.  It comprises of 6 key components:

**1) Build a core team to drive research, deliver education, develop infrastructure, and organize the Institute's activities.** The Institute is led by Dr. Michel Dumontier and Institute Manager dr. Claudia van Oppen. As of January 1 2019, the team has grown to 26 individuals. The team is highly diverse with members coming from the Netherlands and far and wide across the globe. The team was comprised of 1 chair of emerging technologies, 2 assistant professors, 7 postdocs, 1 PhD student, 3 researchers, 2 Master's students, 2 data scientists, 2 developers, 1 data science architect, 1 office manager and 2 participation employees. In 2019, we plan to recruit 2 additional faculty members as core staff with rank up to Full Professor. We will additionally recruit researchers and staff into the Institute that identify with IDS@UM mission, can provide complementary elements, and have a proven track record of data-driven research, grant funding, educational programs, and community-building.




**2) Pursue exciting research** that tackle society-relevant problems and exemplify the vision and mission of the Institute. Platform research projects will demonstrate cross-cutting multi-disciplinary interactions on the main research themes of i) accelerating scientific discovery, ii) improving health care and well-being, and iii) empowering communities. These projects will involve core members, students, postdocs, researchers, partners, and stakeholders inside and out of the University. Projects will be funded internally to develop initial proofs of concept while pursuing external funding opportunities to sustain and growth these research directions. An elaboration of research directions and achievements in provided in the Research section.

**3) Establish a state of the art research IT infrastructure** to strengthen Data Science community at Maastricht University and to support the activities of the Institute. This will be done in close conjunction and interaction with research groups and the University's Information Technology groups. This includes, but is not limited to, state of the art hybrid-cloud infrastructure, data storage, data science applications and software and licenses. In 2018, we secured funding through the IT Board to create a first version of this infrastructure, and will develop models of governance and sustainability through the pilot project.

(4) **Train the next generation of collaborative data scientists**. Several units at the University have already developed well subscribed educational courses and programs to educate students in the field of Data Science. Many more opportunities exist to make Data Science more widely relevant and available. We are creating formal courses and informal workshops that bring the expertise of our researchers to students and staff. Our work illustrates the principle of research-based learning as part of the Collaborative Open Research Education (CORE) initiative at Maastricht University. We work with our partners to create new *interfaculty* programming at the high-school, undergraduate, postgraduate, and professional levels that drive curiosity and offer new and exciting career directions. We plan to develop a data intelligence and innovation lab with mentorship from IDS researchers and other partners (e.g. Brightlands, MECC) to create entrepreneurial opportunities for Data Scientists.

(5) **Engage researchers across the University to participate in the mission of the Institute.** Our goal is to be the academic home for data scientists at the University and in the region. We have developed infrastructure to support this idea (e.g. website) along with a series of activities to network disparate groups, whether they be researchers, students, developers, alumni, professionals, and entrepreneurs. We have done so through a series of open forums, working groups, seminar series, journal clubs, and other community-building events, and as we onboard new members to the organization. We create additional interactions by hosting data scientists from different organizations, faculties, departments, and research groups.

(6) **Transfer knowledge to industry and society**
With its excellent research, IDS strives to provide a significant contribution to the Scientific Data Research community @ Maastricht University. Importantly, we seek significant additional recognition arising from high quality research that contributes to various societal issues such as economy, public policy or quality of life. Dissemination of research findings of direct interest to specific stakeholder groups will take place through journals, seminars, conferences, engaging in professionals networks etc. Prof. Dumontier is also one of the Principle Investigators in the BISS institute, a UM valorization platform at the Brightlands Smart Services Campus.

# Research

## IDS Research Themes

We have identified *three* key research themes to focus our efforts on. These are:
(1) **Accelerating scientific discovery** through the development of powerful Artificial Intelligence (AI) platforms coupled with FAIR (Findable, Accessible, Interoperable, Reusable) data and services to systematically unlock knowledge about the world we live in. Most aspects of scientific inquiry remain entirely dependent on human expertise and effort. However, no one person can keep track of the vast amounts of data, tools, and knowledge being generated each and every day. Moreover, with recent reports claiming rates of non-reproducibility of 64% in psychological studies and up to 89% in pharmacological studies, there is an urgent need to find effective approaches to conduct scientific investigations that takes into account and makes sense of the vast amount of data in a more reliable manner. Recent developments in data-driven systems show that machines can effectively ingest complex information and outperform humans in many tasks. Examples of research topics in this area include:

- Creating and sharing FAIR data and other digital research objects (e.g. data, software, publications) in a responsible manner.
- Uncovering evidence that supports or potentially disputes a scientific assertion
- Designing, prioritizing, and orchestrating experiments to fill in knowledge gaps
- Tracking scientific progress, evolution of scientific disciplines, and scientific impact.
- Adaptive data-driven learning to maximize the accuracy and completeness of any human endeavour

(2) **Improving clinical care and well-being** through the creation of intelligent systems that bring the science of medicine back into the practice of medicine. Despite ritualistic and time-consuming chronicling of patient encounters in electronic health care systems worldwide, these on their own cannot translate into improved patient outcomes. Meanwhile, new technology is filling the gap to monitor and advise patient health on a per second basis, with unclear health benefits. Once again, vast amounts of data are becoming readily available, but we lack the infrastructure, methodology, and understanding of the social, legal, and ethical aspects of health information systems to maximally advantage of these developments. Our goal in this research theme is to exploit emerging sources of health-relevant data to improve clinical care and well being. Topics include, but are not limited to:

- Using distributed data in a privacy-preserving manner to identify environmental determinants of health at the level of an individual.
- Platforms to undertake systematic comparative effectiveness research using millions of patients worldwide
- Methods for continuous and personalized diagnosis, prognosis, and treatment.
- Methods to improve the quality of the physician-patient interaction. e.g. constructing a brief, but accurate patient summary from patient's social networking data.

- Methods to aid care workers find and work with people who need help at the time they need it.

(3) **Empowering communities** to characterize, implement, and monitor data-driven solutions that optimize their investments to maximize their quality of life. It wasn't that long ago that the tools for multivariate analysis, data mining and machine learning had limited utility and could only be used by specialists. Today, data science is more accessible and easier to use than ever – open source frameworks are freely available, compute-hungry analyses can be readily executed on cloud infrastructure, and enormous online communities are sharing their recipes and their accomplishments. Indeed, people from all walks of life are applying data science out of curiosity and to create altogether new applications to meet a societal need. We believe that data science can be the bridge that brings youth and community leaders together to better understand the problems that they face and craft solutions to make life better for them and their neighbours. In this research theme, we will focus on bringing data science to a region nestled in the heart of Western Europe, which is making investments to overcome serious economic challenges, and must strive to include youth as part of a strategy for economic and social prosperity. Topics of interest include, but are not limited to:

- Digital infrastructure to enable collaboration on policy making
- Infrastructure for data sharing while strengthening privacy controls and cybersecurity for citizens and industry
- Data driven methods to foster job creation and reduce social inequality

Identification and removal of barriers to digital innovation and commercialization for young data scientists and entrepreneurs

# Research programs & projects, publications and funding

Aligned with these strategic research directions, several research programs are now in development. These include Responsible automation of data-driven research, clinical data science, Ethics, Law, and Social Implications of the Data-Driven Society, LawTech, NeuroTech, AgriTech

## Responsible automation of data-driven research

Research remains a time consuming, highly manual, and error prone endeavor. Researchers must find and make sense of thousands of relevant research articles in order to synthesize the state of the art, identify gaps in current understanding, synthesize new hypotheses, and design and execute increasingly complex workflows. Consequently, the reproducibility of hundreds of landmark studies in cancer, pharmacology, and psychology have been brought into question, suggesting that new approaches to the design and execution of experimental research are urgently required.

Targeting IDS' Accelerating Scientific Discovery research theme, this line of groundbreaking research led by Professor Michel Dumontier investigates the manner and extent to which

empirical research can be automated, while keeping humans in the loop in order to achieve maximal benefit to the field, and to society. Our bold vision is grounded in four main research aims: 1) development of architectures, frameworks, softwares, and specifications for the establishment of an Internet of FAIR data and services that will enable automated discovery and reuse of relevant content, 2) the development of systems capable of optimizing the interplay between humans and machines to create and use high quality and FAIR data and services pertinent to the realization of a specific research objective, and 3) the establishment of a social, legal, ethical frameworks that affords benefits to all stakeholders (e.g. researchers, funders, industry, society), and 4) the application of automated research to biology, medicine, agriculture, and law.

In collaboration with the MAASTRO clinic and BISS institute, we are developing a lung cancer detection solution that harnesses the power of non-experts to annotate clinical images that are then used to train a machine learning algorithm to detect tumors. We received funding from The Limburg University Fund program that will enable us to train researchers in this approach. In January 2019, we submitted an Innovative Training Networks grant to recruit up to 15 Early Stage Researchers (ESR) to develop the theme of responsible automation of data-driven research and apply it to problems in medicine and agriculture.

## Clinical Data Science

The clinical data science group, led by Professor Andre Dekker, is one of the initiators of the Personal Health Train (PHT) infrastructure. A major goal of the PHT is to accelerate scientific discovery by making health data available to scientists. The train metaphor is used to explain the PHT infrastructure: stations with FAIR health and health-related data are connected by secure and monitored tracks via which care professionals, researchers or citizens can run AI trains that ask questions to the stations and return answers. Bringing questions to data rather than moving data is a key differentiator of the PHT and avoids legal, ethical, societal and technical barriers associated with physical data sharing. The PHT is the health implementation of the IDS goal of accelerating scientific discovery that will create powerful Artificial Intelligence (AI) platforms coupled with FAIR data and services to systematically unlock knowledge about the world we live in.

The clinical data science group has pioneered the application of objective, outcome prediction models in the form of decision support systems in medicine, and as such contributes directly to the IDS research theme of improving clinical care and well-being. Over the years, clinical, imaging and outcome data of thousands of patients has been used to learn dozens of prediction models. Some of these have reached routine clinical care and are applied to make more individual and more shared decisions in medicine. Besides improving clinical care by making better decisions, the group has developed AI based automation that has found its way into clinical practice, increasing efficiency and quality of routine tasks. Current work includes the inclusion of non-medical data into prediction models, implementation studies into the sage and effective introduction of clinical decision support as well as extensive ethical-legal-societal investigation into the proper role of big data and AI in medicine.

The clinical data science group's goal is first and foremost to empower citizens and patients to take meaningful control of their health data, and this aligns to the IDS theme of empowering the health community – policy makers, care professionals, academics, industry, citizens and patients. Specific, informed consent on the use of personal health data by others and using one's own health data for citizen science are the cornerstones of trust in the health data community which currently spans 5 continents and more than 25 institutes. By emphasizing security, trust, control and transparency and taking its academic leadership seriously by publishing open access, open source and, as much as possible, open data the clinical data science group has provided guidance to the World Economic Forum, the Dutch Health Ministry and partners globally on how to build a health data community. Similarly, tools have been and continue to be developed to support health care professionals and now scientists starting their own AI program on the health data of the community.

## Ethics, Law, and Social Implications of the Data-Driven Society

Ethics, Law, and Social Implications (ELSI) research is vitally important to the success of modern data science. The ELSI group at IDS, led by Dr. David Townend, is leading the field by embedding this collaboration strongly at the core of its work and structure. The legal issues involved are complex, covering, for example, human rights, particularly privacy and data protection, intellectual property, administrative legal issues, liability and contract issues. There are questions of colloquial and practical ethics - how the developments are seen as 'fair' or 'just' in society, and how ethics are applied in research ethics approvals - and there are broader questions of how modern data science (in different settings) conform to and challenge different concepts of justice in society (and between societies). The broader 'justice' question is a philosophical and sociological question of who we are and whom we want to be?

This translates into specific questions that the ELSI group asks within the Institute: Does the data science research conform to the relevant international, European and domestic laws and ethics principles binding on that work (i.e. according to the various jurisdictions within which the work is undertaken)? Is the governance structure (i.e. law and ethics) that exists for the area created by the work or where the work is seated 'fit for purpose', and if not, how can it be modified to produce a better fit? What law and ethics problems do the new technological (data science) situations bring? How far do existing interpretations of laws and ethics meet the governance needs of particular situations? How far can existing laws and ethics be interpreted to accommodate the governance challenges of new situations and technologies? How can new laws and ethics be developed to fill in gaps?

The work will be informed by two further considerations: How is the work and the governance structure understood by citizens? How do individuals perceive the harms and benefits from the new technologies and data science? How do individuals respond to the governance structures? Are the governance solutions appropriate in the eyes of European and national regulators (e.g. the EU European Data Protection Board, and the Dutch Autoriteit Persoonsgegevens and Inspectie voor Gezondheidszorg en Jeugd)?

These questions are addressed by academics (law, ethics, sociology, philosophy, anthropology, etc.), and practitioners (lawyers and data protection officers, etc.). These ELSI actors also interact in an interdisciplinary way with the rest of other members of the IDS

team. Examples of current projects where the four questions above are addressed include governance of the Personal Health Train, governance work for Artificial Intelligence (AI) and Machine Learning (ML), and how far and in what ways can AI and ML produce technologies for the operation of governance in society.

## LawTech

The Institute has initiated several research collaborations with the Faculty of Law. Our research theme is to use cutting-edge data science techniques to **interpret the complex structure of the legal system.** Together with our collaborators from Law, we explore this agenda in **three research directions**:

1. Artificially intelligent conversational agents (chatbots) to educate consumers about their rights and to resolve legal disputes,
2. Data-driven analysis of court decisions over time to verify if the law is being consistently applied and enforced, and
3. Generating quantitative evidence to support qualitative legal analyses that inform policy-making

All three directions align with IDS' core research themes of: **accelerating scientific discovery, strengthening communities and improving the wellbeing of citizens**. This is achieved through the development of powerful AI systems built on FAIR data to educate consumers about the law, and through generating insight from legal data which drives policy development that improves societal wellbeing. In the near future we envision these research lines to be strengthened with coordination through the recently launched **Maastricht Law and Tech Lab** which aims to investigate interactions between law, technology and data science.

## NeuroTech

The NeuroInformatics research program aims to use big data, data science and AI technologies to 1) improve our understanding of brain function and dysfunction, 2) power more accurate diagnosis, prognosis and treatment of individuals, and 3) develop new therapeutic interventions for brain disorders. Our plan is to foster and drive collaborative research, education and community activities with neuroscience departments embedded within FHML and FPN. For instance, we co-hosted a very successful joint research seminar on Machine Learning and Personalized Medicine in Psychiatry with Massimiliano Grassi and Koen Schruers in 2018, with a follow up joint session in our Data Science Research Seminar Series in February 2019. We have partnered with Daniel vd Hove, Ehsan Pishva & Jos Prickaerts from MHeNS to organise a session on 'Big data & AI in neuroscience' during the Dutch Neuroscience Meeting 2019. We are pursuing a collaborative research project with PhD student Massimiliano Grassi & Professor Koen Schruers on developing a clinically translatable machine learning model to predict Alzheimer's disease in MCI patients. This work is being continued in collaboration with Inez Ramaekers, Sebastian Kohler & Frans Verhey from the Alzheimer Center Limburg. We have agreed to have a joint PhD position with MHeNS.

The collaborations with these research groups & faculties will enable the NeuroInformatics research program within IDS to enrich its expertise in cognitive neuropsychiatry, clinical neuroscience, mental health, human behavior, brain imaging & translational neuroscience, and enable the neuroscience research groups and institutes at Maastricht university to expand their research into the areas of big data, data science and advanced computation. Finally, to become a world-class center in data science, and more specifically in neuro/biomedical/health informatics, we are intending to establish sustainable collaborations with research institutes, industry partners and other research organisations beyond Maastricht University.

## AgriTech

The establishment of Dr. Brewster as the Chair of Emerging Technologies at IDS, sponsored by TNO, will focus on the application of semantics and data science to the food and agriculture system, a major contributor to climate breakdown and one facet of the most import social, environmental and economic challenge of our times. Data science can make a major contribution in providing tools for measurement, analysis, data integration, optimisation and mitigation of the effects of the agrifood system on our natural environment.

Our initial priority has been to develop research funding proposals at the intersection of agrifood and data science. Two H2020 proposals were submitted in January, and a further one is in preparation for the end of March. February will see a training focussed Erasmus + proposal submitted in the application data science for the agritech sector. We will continue this line of activities in the foreseeable future.

Concurrently, we are examining collaborative opportunities between TNO and UM. This is initially appearing in the context of the Brightlands campus, with TNO responsible for the Techruption project which we hope will work closely with the UM BISS team to provide various types of consultancy and contract research support to companies and entrepreneurs in the Limburg region. Finally we are planning to get researchers from both institutions to visit each other sharing research interests and exchanging ideas.

## Recommendation Systems

Assistant Professor Dimitrios Rafailidis, jointly appointed in August 2018 between IDS and the Department of Data Science and Knowledge Engineering, leads the Recommendation Systems (RecSys). The RecSys group will incorporate machine learning strategies and data analytics to study different aspects of RecSys. Aligned with the IDS' research core theme of accelerating scientific discovery, the RecSys research program targets to the following goals:

- perform adaptive data-driven learning to support people with personalized services
- develop AI agents to elicit the interests and preferences of individuals via conversations
- use AI to improve human-computer interactions in daily life

Recommendation strategies are among the most popular Decision Making algorithms in Data Science, and are becoming integral parts of daily life by generating personalized recommendations. In 2018, we submitted a bi-collateral proposal to the Swiss National

Science Foundation to learn people's contextual preferences and produce recommendations when using mobile applications on a daily basis in collaboration with researchers at Amazon/Alexa and the University of Lugano. We also submitted a Comenius scholarship to generate personalized recommendations for students at UM to improve their competencies and better guide them in their studies. This work will personalize recommendations for professional, academic and personal development by making effective use of existing knowledge e.g., grades, extracurricular activities, self-reflections and information about our student population e.g., employability rankings, graduate profiles, course and programme evaluations. This will result in students in becoming more employable, in that our students are capable to continuously improve their knowledge and skills during their studies.

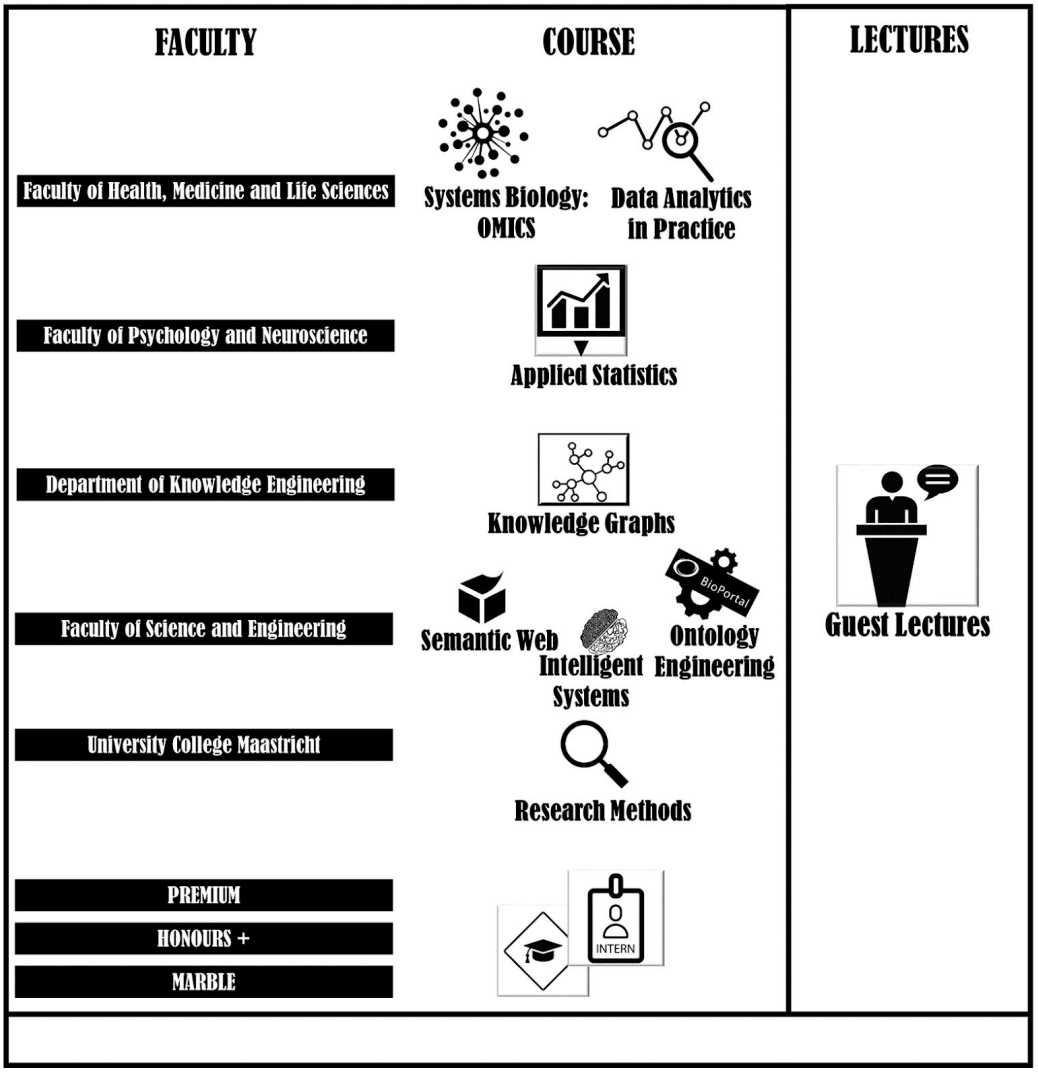
# Education

Report after report indicates that we are simply not training enough data scientists to meet world demand. Even KE@Work, an internship program out of the Department of Knowledge Engineering and Data Science, shows that the supply of qualified students is less than half of the demand, and that demand is growing much faster than the supply. We must do much more to entice people to consider data science as a career.

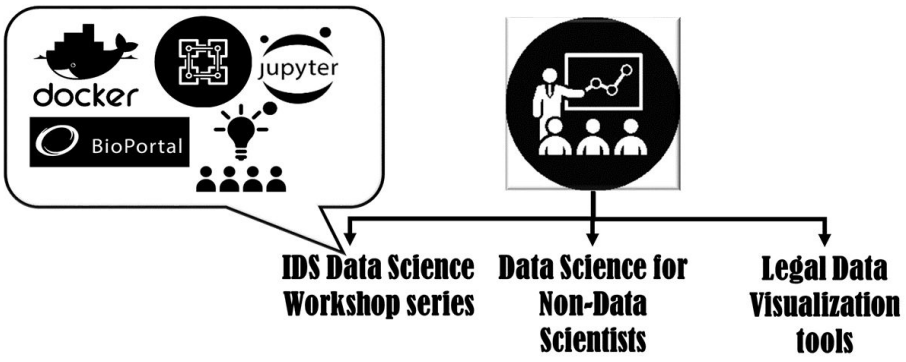
IDS holds three objectives in educating the next generation of collaborative data scientist @ UM:

1) Enhance existing data science programs and courses
2) Create learning opportunities for youth and professionals
3) Develop new (interfaculty) programmes featuring data science and artificial intelligence in society

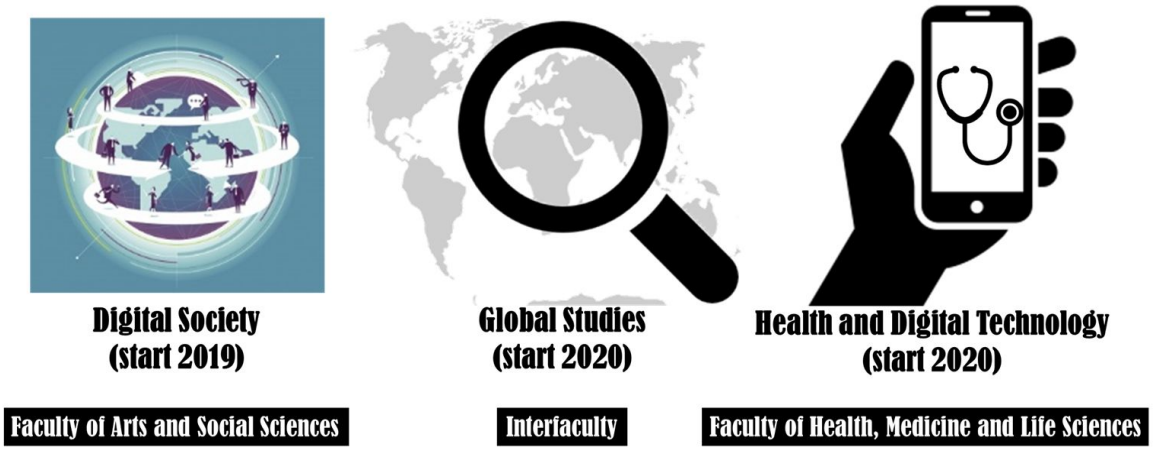
Our first objective is to work with our partners at the University to **catalogue, promote, and strengthen existing data science programs and courses.** We have conducted a thorough survey of existing courses and Data science related curricula to identify the options to enhance the content and explored possibilities for collaboration. This has helped us understand what is being offered, where we could help out, and what gaps still remain. We have contributed to existing programs and courses across faculties through a variety of guest lectures (e.g. Semantic Web, Ontology Engineering (FSE/DKE), System Biology (FHML), Data analytics in practice (SBE)), workshops (e.g. Research Methods (UCM)) and supervising internships, master theses and project based programmes like Honours +, Premium and MARBLE related to diverse data science topics. An entirely new Masters course "Building and Mining Knowledge Graphs" will be taught in Q1 2019 and made available to Masters students in the Masters in Artificial Intelligence and Masters in Data-Driven Decision Making.




Second, we will **create learning opportunities for youth and professionals**. Here, we intend to leverage the strength of our researchers to provide cutting edge content. We will explore both formal and informal ways of delivering data science education. For instance, we offered Data Science Workshops as a series of hands-on workshops on topics including Docker, Machine Learning, Crowdsourcing, Ontologies, and Data Science for Non-Data scientists. These workshops were attended by UM-lecturers, researchers, IT professionals and students. We will explore further how these workshops can be incorporated in a sustainable teaching infrastructure e.g. via the executive teaching program @ BISS or in collaboration with Edlab.




One specific group that IDS intends to focus its activities on are young researchers including kids, high school students, (under)graduates, PhD students and Postdocs. We delivered an interactive lecture and a challenging assignment for the UM KidzUniversity on the topic of “Can a robot be your friend?”. These activities to provide young minds with the opportunity to learn not only of the technical nature of data science, but also elicit their social and ethical implications.




Our third objective is to develop **new interfaculty data science programs and courses** that feature data science and artificial intelligence in society. A key criteria for our involvement is whether the educational offering include interest and support from more than one faculty or department. The inclusion of diverse stakeholder representation that includes multiple faculties working professionals will be critical to the development and delivery of new data science courses and programmes. We anticipate that our educational activities will grow gradually, but strategically. As a first step IDS staff are involved in the development and educational planning of new educational programs such as Bachelor Digital Society (FASOS), interfaculty Bachelor Global studies and the track Health and Digital Technology (FHML). Additionally, we will be involved in the development and teaching of new FSE Bachelor programs: Business Engineering, Circular Engineering, and Medical Engineering. IDS and the University Library are exploring the development of research data management courses and an educational program (BK-FAIR). We anticipate that within a few years the institute of data science will also host new BSc/MSc programmes. These programmes will focus on combining fields from alpha, gamma, beta and diverse domains such as Health, Psychology, Law & Ethics. In developing interfaculty education we strive to make effectively use of existing governance, administrative and support structures.

# Research IT

Data science research requires agility and flexibility in the procurement and use of state of the art IT infrastructure and software. Unfortunately, the absence of a unifying research IT vision or policy at Maastricht University has resulted in the development of multiple isolated, incompatible, and independently managed infrastructures that have differing policies and patchy compliance to organizational, national and international regulations. All of this makes it difficult for our own researchers to undertake cutting edge research. It also marks a missed opportunity to tout a shared infrastructure towards attracting the best students and researchers to the University.

**Towards a research IT infrastructure at Maastricht University**
To provide cutting edge technologies for data science research at Maastricht University will required a change in policy, organization, and investments in key technological infrastructure. We must establish policies that make it easier for researchers to procure advanced information technologies, while ensuring that their digital assets can also be discovered and used by others inside and outside the university, in a secure and reliable manner (in line with the FAIR principles). We must make sense of the myriad of incompatible faculty and department policies so as to make sure that researchers are able to do their work in a straightforward, but compliant manner.

It is our goal to develop a Research IT to support the growing Data Science community at Maastricht University. In 2017, using Professor Dumontier's research funding, we procured 2 servers to host databases, applications and web services that were deployed at Stanford University. As part of the transition, we developed the expertise to make components Docker compatible, so they could be deployed on any number of platforms including the servers, development laptops as well as with other cloud providers.

In 2018, we have shared our expertise and made our compute platform available for **others to reliably deploy their software and applications**. This includes: a Neuron docker container for Menica DiBenedetto (MAcsBio) that reduced compute time for Neuron analysis by several weeks; a Matlab Docker container for Michele Maurel (MAcsBio) to enable a large scale Matlab computation for a Master project; a XNAT Docker prototype for Elia Formissano (FPN) and Federico DeMatino (FPN) to make a storage for sharable MRI images possible;
an RStudio Docker for international workshop (16 users) for Carlos Collares (EDU) that to perform a large scale Adaptive Learning Test evaluation. As part of our strategy, we have also provided formal Docker training as part of the Data Science Workshop Series and have provided informal training and knowledge transfer to IT focused individuals such as Dries Boers (MAcsBio).

Towards creating a shared researcht IT infrastructure at Maastricht University, we are running a pilot project called the **Data Science Research Infrastructure** (DSRI) with the University's IT group (ICTS) with initial contributions via the IT Board. The pilot will procure state of the art computer, network, and storage infrastructure for data science research. The pilot will explore a variety of issues including: shared policies and operational procedures, clarify governance and participant roles and responsibilities, and inform maintenance and support costs beyond the pilot phase. It will be used to support data science research groups across the university, with an initial focus on the pilot projects that are part of the **Community of Data-Driven Insights (CDDI)**, which aims to incentivize high quality research data management with new insights from the computational analysis of institutionally deposited data.

This state of the art hybrid cloud infrastructure enables IDS and its partnering organization to meet their current and future storage needs. All IDS applications will run within the cloud infrastructure (data less client). This will enable high performance access to data and easier sharing of results. The reuse of existing services and code will enable IDS to reuse smaller applications and frameworks into bigger applications and frameworks. Successful methodologies, tools and software from the Open Source community will be applied to all public IDS applications. As services and applications move from existing data centers into the IDS cloud those freed up computational resources can either be used for new services and applications. Those resources can also be used to extend the cloud capacity, resulting in a scalable and sustainable shared Research IT infrastructure @UM.

# Community

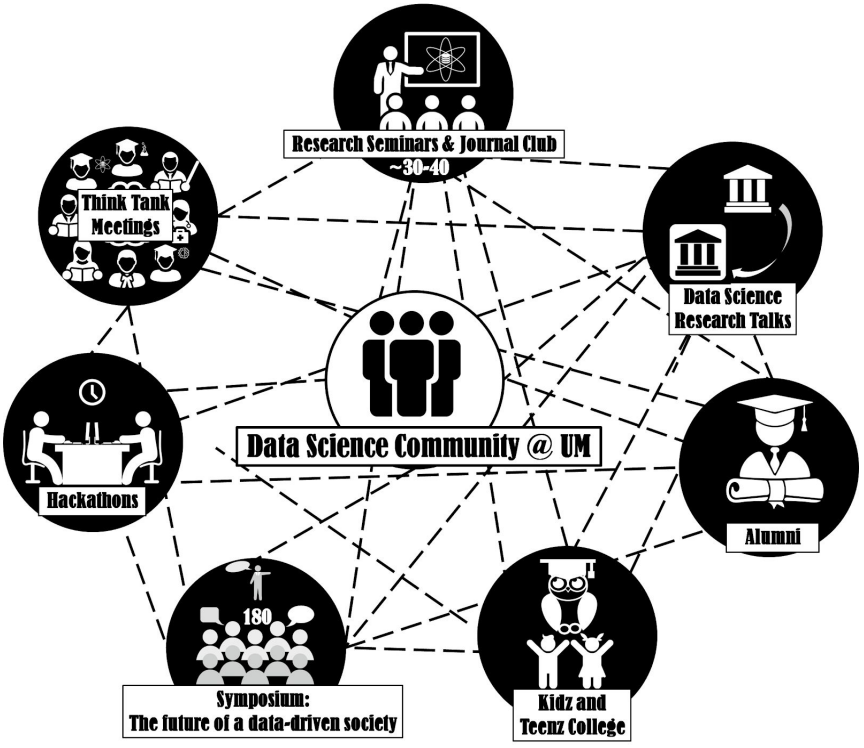
## University network

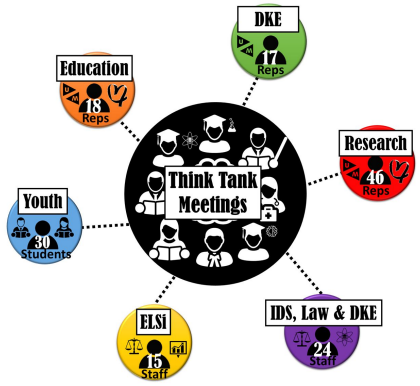
Cross-cutting interactions are at the core of IDS@UM and building a strong, vibrant, and diverse Data Science community feeds into our strategic aims. We have invested in variety of activities to foster growth of this community. One tangible outcome is working with UM Marketing and Communications with our partners at UM to create an overall Data Science value proposition for UM:

**Value Proposition Data Science@UM**
*"We are a pioneering, university-wide network of data-driven researchers at an institution with more than 25 years' experience in Data Science education and research. Our community consists of multidisciplinary teams that address significant scientific problems and major societal issues through and related to data science including social, legal and ethical concerns in a wide range of academic disciplines".*

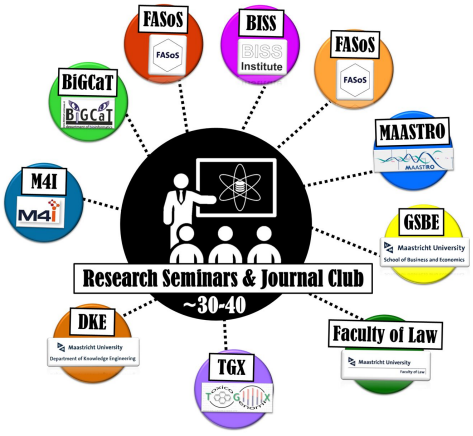
**Data Science Community Activities**
The institute has initiated a series of activities to foster the Data Science Community at Maastricht University (Figure X). These include Data Science Think Tanks, Data Science Seminar Series, Data Science Research Talks, the Data Science Symposium, the 2018 Theme for the Dies Natalis,

# Community Building Activities at IDS

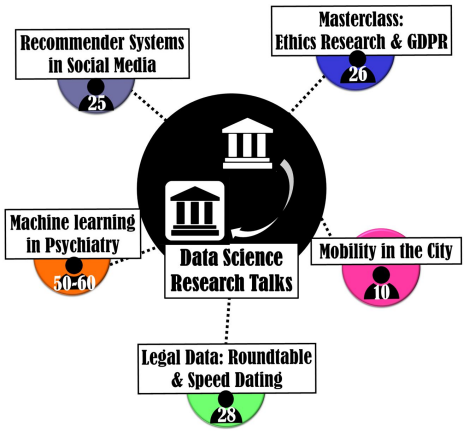



**The Data Science Think Tanks** are stakeholder consultations to gain insights into their ideas, needs, and dreams to foster a valuable Data Science Community. The Think Tank meeting series kicked off with three thematic meetings targeting Research, Education and Youth, followed by targeted think tanks with DKE, FHML, LAW and ELSI. In 2019, a new series of Think Tanks will be organized for Neuroscience, ELSI and Innovation in Education.




**The Data Science Research Seminar series** brings together researchers from across the University of Maastricht to present and discuss their research in Data Science. In 2018, ten Data Science Research Seminars were co-organized with departments across the university with an average of 35 participants per session.




**Data Science Research Talks** were organized by IDS to share the knowledge of our visiting experts with members of the data science community @ UM. The main goal was to acquire more knowledge of the application of data science in other fields or research areas and to discuss about this.

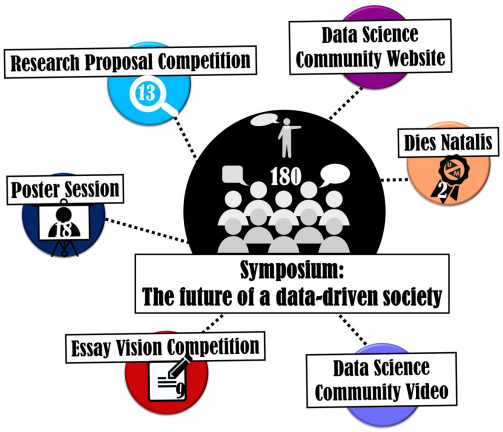
One very successful example was entitled 'Data Science Research talk: Machine learning in Psychiatry' co-organized with MHeNS Koen Schreurs as part of Massimiliano Grassi's research visit on the topic of our joint research work: Machine Learning and Personalized Medicine in Psychiatry. This research talk was very well attended by approximately 60 staff members and (Phd) students from MHeNS and other faculties.

In partnership with the European Centre on Privacy and Cybersecurity, (ECPC), The LeD (Legal Data) community, IDS organized a Summer Masterclass on "Ethics, Research and the GDPR", with the experts Cosimo Monda, Paolo Balboni and Christopher Kuner. The target group consisted of UM staff IDS & faculty of Law, Master students, PhD students and Post docs.

At The research talk "Mobility in the City" dr. Loreto Bravo Director of the Data Science Institute of the Universidad del Desarrollo (Chile) presented the results of several research projects that made use of Telcom data complemented with other data sources to study the city through several lenses e.g. How do people move?

Dr. Dimitrios Rafailidis presented diverse aspects of "Recommender Systems in Social Media" such as the fundamental aspects of recommender systems, the collaborative filtering strategy and he explored how recommendation algorithms can exploit the selections of social friends to improve the quality of the recommendations.

As a follow up of the LAW-IDS Think Tank meeting a Legal Data (LeD) Speed Date & Roundtable was organized", to explore common grounds for joint academic output in the areas of "Big data, Trustworthy Computation & Legal Analytics" which resulted in some joint research projects of staff of IDS and LAW.



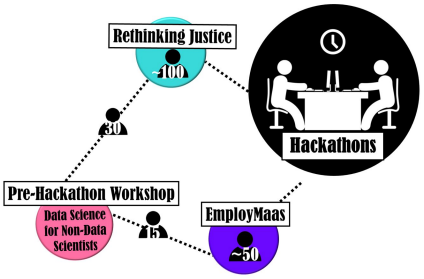

**The Dies Natalis** is an annual celebration of the establishment of the University. In 2018, the celebration was themed "The Future of a Data-Driven Society", in which Professor Michel Dumontier discussed 'A social and technological infrastructure for data science', while Professor Sally Wyatt from the Faculty of Arts and Social Science (FASoS) reflected on the question of

‘Where is the knowledge we have lost in data?’. Professors Carole Goble (Manchester University) and Lucy Suchman (Lancaster University) were awarded Honorary Doctorates for their contributions to Data Science and the Digital Society by Professors Michel Dumontier and Sally Wyatt, respectively. Just prior to the Dies Natalis event, we organized the **The Future of a Data-Driven Society Symposium** to bring the Data Science community together with keynote lectures from Professor Goble and Professor Suchman, coupled with a research proposal and visionary essay competitions and poster session:

**The Data Science research proposal competition** aimed to stimulate the development of innovative research proposals that use data science to tackle key problems of societal relevance within multi-disciplinary teams. In total, 13 research proposals were submitted and 3 proposals were selected by the interdisciplinary jury and pitched by the research teams. The winner of the research competition was the research proposal entitled: “Intelligent games for assessing cognitive, social and physical capabilities of elderly and children”. The winning team received a cheque from €5.000 sponsored by University Fund Limburg and mentoring from Prof. Michel Dumontier and Prof. Sally Wyatt to follow up. The other nominated teams received spontaneously €500,- from the Executive Board.

**The visionary essay competition** concerning the future of a data-driven society received 9 submissions of UM staff and students. The essays were submitted in any form of web published digital media, including text, visuals, software, an application, a website and so on. The best essay was selected by an interdisciplinary jury and awarded a price. The winner was Claudia Egher (PhD student FASOS) with her remarkable essay titled: *Digital morning or manic by design.*

**The Poster Session** aimed to highlight innovative and impactful on-going data science research. 18 posters were presented, highlighting research from diverse backgrounds & disciplines in data science and various departments of Maastricht University.




**Hackathons** are gatherings of students, professionals, developers, designers and business strategists to celebrate free thinking, enthusiastic doing and aims to create a nurturing environment and mindset for knowledge sharing and collaborating in challenges within a limited time frame. The Institute co-organized two Hackathons: 1) the **Rethinking Justice Hackathon** in collaboration with Technolawgeeks was a 24 hour event in which approximately 100 participants worked together in multidisciplinary teams to find solutions to one of four challenges on Social Justice, E-Commerce Conflicts, Data-Driven Justice and Courts of the Future proposed by Hill, Ebay, IDS and DFC court; 2) the 1 day **EmployMaas Hackathon** organized via UM student employability initiative UM,

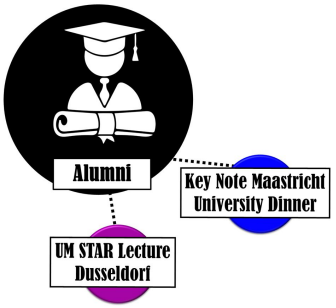
IDS, the Municipality of Maastricht and the Province of Limburg Prior brought together 35 students, staff and alumni from Maastricht University and 15 professional mentors to find solutions to employability in South Limburg. IDS also organized a pre-Hackathon workshop **Data Science for Non-Data Scientist** to introduce key concepts to the hackathon participants.

IDS contributed lectures to the **Kidz, Teenz and Alumni programs** @UM. Within the **KidzCollege** 2018 series, Dr. Rico Mockel (DKE), Dr. Claudia van Oppen (IDS), Prof. Michel Dumontier and Prof. Sally Wyatt (FASOS) gave an interesting and interactive lecture titled “Can robots be friends?” to 65 kids aged 11-13. In this Kidzcollege some robots were show, explained how they work, and discussed with the kids what still has to be done to make them more intelligent and what makes them a friend?

**TeenzCollege** puts major current themes in historical perspective, drawing parallels between primarily the Second World War and today’s day and age. This program is targeted towards ambitious and talented high school kids aged 13-16. IDS proposed a research project focussing on analysis of Second World War data, to map map the bombing patterns and strategies used during many Second World War missions and to develop a predictive model, based on these bombing patterns, which will allow us to estimate the likely bombing locations. Unfortunately only 3 students were interested in this project, so it was postponed.

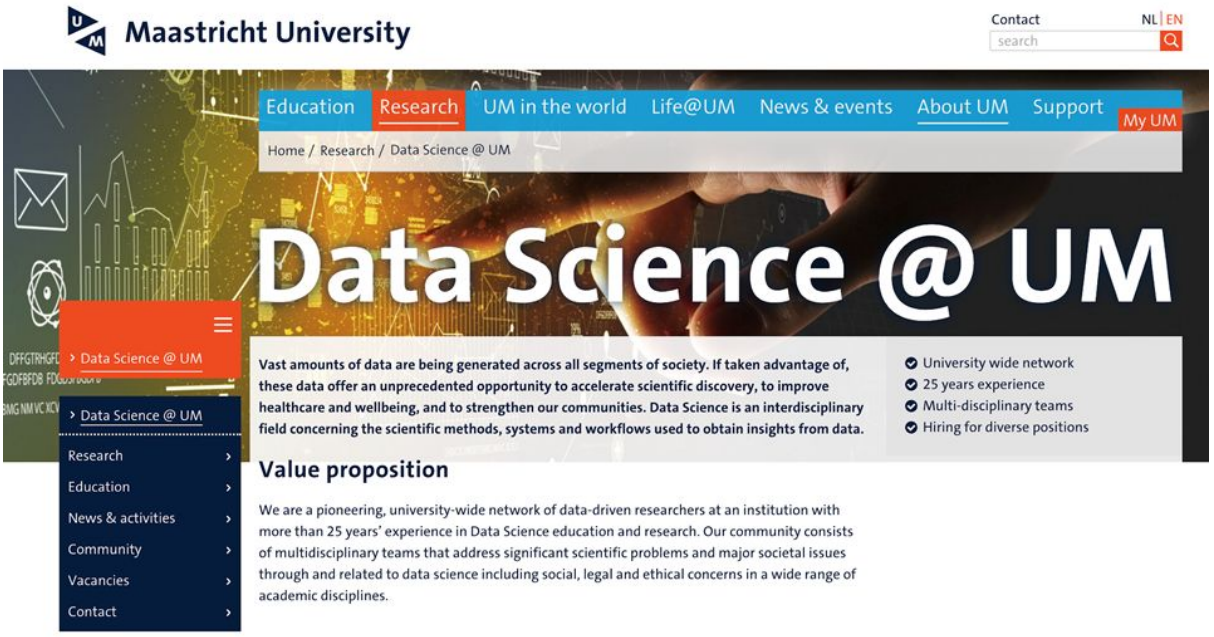
**Alumni**

At the **UM Start Lecture series** for UM Alumni in Düsseldorf professor Michel Dumontier presented the “The Future of Science and Business”. During this lecture prof. Dumontier discussed how data science is affecting our way of life and how we at Maastricht University are preparing the next generation of leaders to address opportunities and challenges in responsible manner.




During the **UM University Dinner 2018** Professor Michel Dumontier gave a keynote speech titled: “Why Data Science changes everything?” This dinner is organized by the University Fund Limburg, in cooperation with Maastricht University. The dinner is a meeting between university staff and scientists with representatives of regional and national companies, partners of Maastricht University, and alumni. The dinner primarily aims at establishing ties in education or research, between all the mentioned partners (friendraising).

The **Data Science @ UM community website** was launched during the Symposium in January 2018. The Data Science @ UM website provides information about research, education, news & activities of the diverse Data Driven research and education groups at UM. As part of the Data Science Marketing and Communication strategy this community website will the upcoming period be further developed towards a truly interactive collaboration platform.




The Data science community is also connected via diverse **social media channels (facebook, linkedin, twitter)** to keep them appraised of news, events, and vacancies.

## FAIR University Initiative

Maastricht University firmly believes that expedient access to all forms of research and research materials is an indispensable aspect of a public university. Maastricht University therefore commits itself to implementing the FAIR Principles across all of its disciplines, with the intention to becoming the first 'FAIR University'. Maastricht University will foster an environment to open and reproducible research, generate new opportunities for data-driven research, and ensure that the general public and other researchers benefit from investments in public universities.

In order to become a truly "FAIR University" the UM2025 FAIR university vision has to be further developed, aligned with the MUMC+ vision on FAIR research Data management and implemented. Next to that a research data management infrastructure and services (support, teaching etc) needs to be in place. Moreover, to enhance the value of this FAIR research data platform for the individual researchers besides added value of the infrastructure and support for data management, scientific value needs to be created by means of automated data driven discovery as described below.

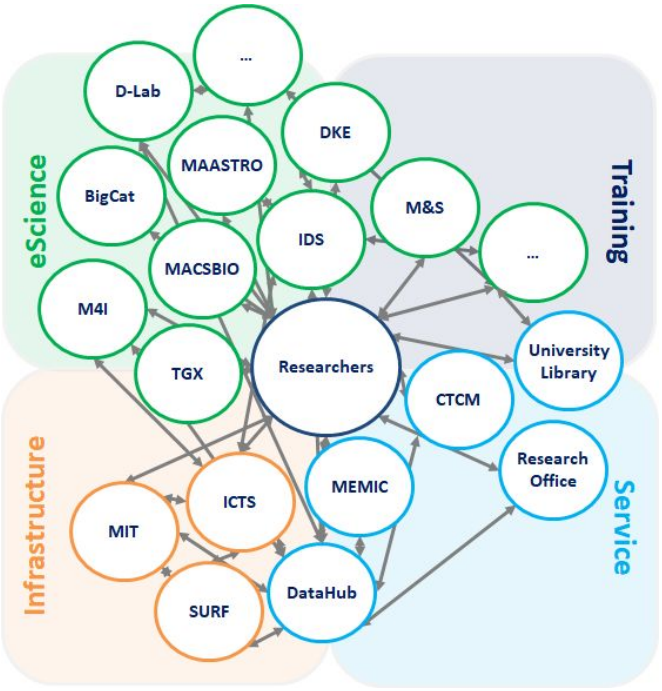
## Community of Data Driven Insights

Researchers are increasingly being asked by journals and funding agencies to make their their research products (data, software, workflows, etc) FAIR - Findable, Accessible, Interoperable, and Reusable. The key insight here is that the availability of research products (and their metadata) could be useful in i) reproducing published research results, ii) validating findings, and even iii) generating new hypotheses. However, it is rare for the people that do all the hard to prepare research products to be FAIR to also benefit from its reuse. As such, data producers are either reluctant to share their data, want to impose limitations, or simply don't care about the reusability of the data, all of which results in limited participation or poor quality data and metadata, that ultimately compromise public investments in research. If we are ever to realize an institutional goal to have most (or all) research data FAIR by 2025 (the UM2025 vision), then we need to address this critical barrier that prevents people from fully participating.

We have proposed **a new strategy** to motivate data sharing and reuse across the university. At the core of this strategy is the establishment of a Community for Data-Driven Insights (CDDI). CDDI aims to ensure that the people who make the effort in making their data FAIR will also gain **immediate professional benefit in the form of new insights produced by automated and actionable data-driven analysis**. We envision users will be able to use, compare, or combine their data with those produced by others in order to promote their work, build confidence in their results, and generate new findings. For instance, when a biomedical researcher adds their gene expression data arising from exposure of a chemical substance to a strain of mice, the envisioned platform will identify and be able to execute relevant public workflows that use submitted data in combination with other public data to, for instance, identify a gene signature that is robust across a variety of attributes such as species, tissues, diseases, chemicals classes, etc. Similarly, the social scientist will be able to more readily ask and answer questions by creating and reusing reproducible workflows involving their data and other data to create compelling and periodically updated visualizations that communicate novel findings.

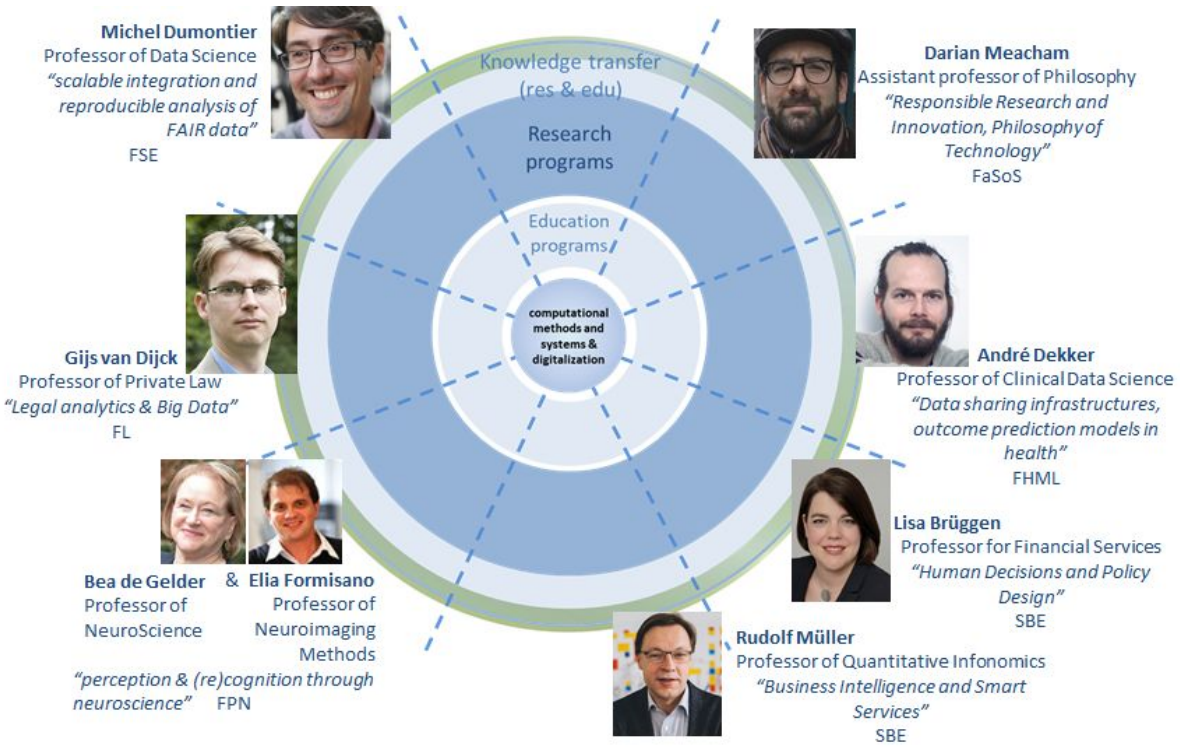
Realizing the vision of in which data-driven insights are automatically generated for researchers at UM will require a tighter form of collaboration and coordination with various units at Maastricht University and the Maastricht University Medical Center. It also necessitates further development of four key components: methods, infrastructure, training, and other services. We need to develop and experimentally validate new methods and approaches. The state of the art needs to be further developed and customized for the ultimate user experience, and embedded into a **shared infrastructure** along with the support to make the most of it. **New training** in the form of workshops, seminars, guides and best practices need to be developed to train students and professionals in how to make use of the system and to share knowledge within the ecosystem. Other services such as marketing and communication are needed to promote the work we are doing at Maastricht University and to build new collaborations worldwide.

The Institute of Data Science is one of the founders of the Community of Data Driven Insights. It will work with a variety of data-intensive research groups (BigCAT, MACSBIO, M4I, MAASTRO, etc). It has an active research program targeting the discovery and reuse of digital resources (FAIR) and the establishment of methods for responsible data science (FACT), both of which are supported by the NWO and the NIH. Together with its service-oriented partners (The CIO-office, the DataHub, the UM Library and ICTS), we believe that we can transform the face of research data management and knowledge discovery at the University, and set a precedent for data reuse and analysis worldwide.



# Regional network:

**Brightlands Institute for digital Smart Society (BISS)** is an Interfaculty platform for translational data-driven research & education in the Brightlands ecosystem. Prof. Michel Dumontier is one of the Principal Investigators of the Institute.




The Brightlands Institute for digital Smart Services (BISS) aims to empower people and organizations through the socially responsible application of digital technologies. BISS offers a transformative set of products and services including co-creation of digital solutions, professional education, applied research and consultancy. The BISS team comprises of a highly collaborative and interdisciplinary team of internationally recognized experts in data science, artificial intelligence,human computer interaction, responsible research and innovation, with diverse backgrounds spanning health, biotechnology, neuroscience, psychology, economics, social science, and law. Professor Dumontier will represent the Faculty of Science and Engineering as one of the Principal Investigators for BISS. His work at BISS will be geared towards products and services that enable people and organizations to create and reuse FAIR data, particularly with sensitive human data and in competitive data sharing environments.

**Maastricht, Working on Europe**

*Maastricht, Working on Europe* is a joint ambition of the City of Maastricht, the Province of Limburg and Maastricht University to create an expertise center of knowledge and vision on Europe and European integration. Research theme 4 focuses on Knowledge, Technology and Digitalisation. The use of social media may help boost citizen engagement and

participation in politics. It can produce knowledge in areas such as medicine, science and politics. However, the use of these tools is not without critics. More research is needed into how knowledge is created and shared by digital platforms. IDS will contribute by examining how youth opinion can be learned from the integration of traditional surveys and social media tools. Surveys are routinely performed to better understand political thought and opinion. However, surveys routinely underestimate the opinion of youth, particularly those that are conducted through telephones or by mail. In contrast, youth are highly active (>90% in Netherlands) in expressing their thoughts and opinions on social media platforms such as facebook, youtube, twitter, and instagram. The analysis of social media for sentiment and opinion is in its infancy, and new research is needed to process these media (e.g. estimate demographics) and promote the voice of youth across europe on topical european issues such as Brexit. The proposed research will explore questions relating to youth aspirations and concerns, topics relating to immigration, employment, and lifestyle. The project will be positioned in collaboration with the Maastricht Europe researchers Christine Arnold (FASoS), Giulia Piccillo (SBE), and Mathieu Segers (UCM), and coordinated with York-Maastricht Europe theme leads Thomas Christiansen (Maastricht) and Kathryn Wright (York).

# National network

## VSNU Digital society

The Digital Society Research programme is a collaboration between the fourteen universities within the Netherlands. This unique programme, led by the Association of Universities in the Netherlands (VSNU), has the intention to promote the collaboration between universities and across disciplines in order address the many pressing questions raised by the emergence of a digital society. The Digital Society programme aims to secure the Netherlands' leading international position in the field of human-centered information technology, and to find solutions to global challenges. The agenda focuses on seven themes: Citizenship & Democracy, Responsible Data Science, Health & Well-Being, Learning & Education, Work & Organisations, Digital Cities & Communities and Safety & Security.

*Digital Society programme @ UM and IDS*
Sally Wyatt, professor of Digital Cultures in the Faculty of Arts and Social Sciences at UM, is one of the coordinators of the programme. Within IDS, David Townend (professor of Law and Legal Philosophy in Health, Medicine and Life Sciences) and Michel Dumontier are two of the lead professors in the Responsible Data Science line while Andre Dekker (professor of Clinical Data Science) is one of the principal investigators in the Health & Well-being trajectory.
From the first quarter of 2019 on, two postdoctoral researchers will be supporting the efforts of prof. Michel Dumontier and Andre Dekker in their respective research lines.

# International network

Both at **RWTH Aachen (RWTH) and at Maastricht University (UM)**, scientific research and education is conducted in the area of data science, in accordance with the FAIR/FACT principles. Both universities have given a boost to data science research and education very recently. At RWTH, this has happened a.o. by awarding a Humboldt chair to professor Wil van der Aalst and at UM by creating IDS@UM led by distinguished professor Michel Dumontier. Data science and digital transformation do not stop at the border of countries and regions, nor at the boundaries implicated by different disciplines. On the contrary, we all can benefit from a cross-border approach. That is why UM, RWTH and BSSC are investigating a cross-border, joint approach in the area of data science. We take a bottom-up, pragmatic approach to achieve this, in order to quickly get first results. An agile approach, fitting the area we are working in. The first step consists of the set-up of a jointly funded team of jointly mentored staff, in connection with the ecosystems of Brightlands Smart Services Campus but also obviously existing networks of RWTH and UM.

We envision **the York-Maastricht Data Science Initiative** as an internationally recognized transnational research program focused on responsible data science by design. It will create innovative software architectures, frameworks, methods, and tools for the representation, integration, transport, analysis, and persistence of heterogeneous, distributed, streaming, and sensitive small and big data, in a manner that is, by design, secure, responsible and sustainable and can be applied to a wide range of problems. This collaboration will attract world class talent and attract new funding to drive further research and innovation, and create training programs to produce highly skilled and in demand data scientists.

# Governance

The Institute of Data Science is directed by Michel Dumontier, the Distinguished Professor of Data Science at Maastricht University. Distinguished Professors are appointed by and are accountable to the Executive Board. The Institute is part of the Faculty of Science and Engineering (FSE) and coordinates its activities with the office of the Dean of FSE.

As of January 2019, an **International Advisory Board** has been appointed to advise the Institute in terms of strategic positioning, to inform it of the latest developments regarding academic synergy, government policy, and commercial opportunities, offer an outsider's perspective and to bring visibility to the Institute. The Advisory Board, in which each member is appointed for a duration of 3 years, consists of:

Professor **Carole Goble** is a Fellow at the Royal Academy of Engineering and the British Computing Society. She has a Honorary Doctorate at Maastricht University and established and co-directed myGrid, which focuses on data intensive e-Science. The team collaborates with scientists worldwide, from many disciplines like Life Sciences, Biodiversity, Astronomy, Chemistry, Health informatics, Social Science and Digital Libraries. In 2010, Carol co-founded the Software Sustainability Institute. In 2014, she became the deputy for the EU ESFRI ELIXIR UK Node. In 2008, Carol won the Microsoft Research Jim Gray e-Science

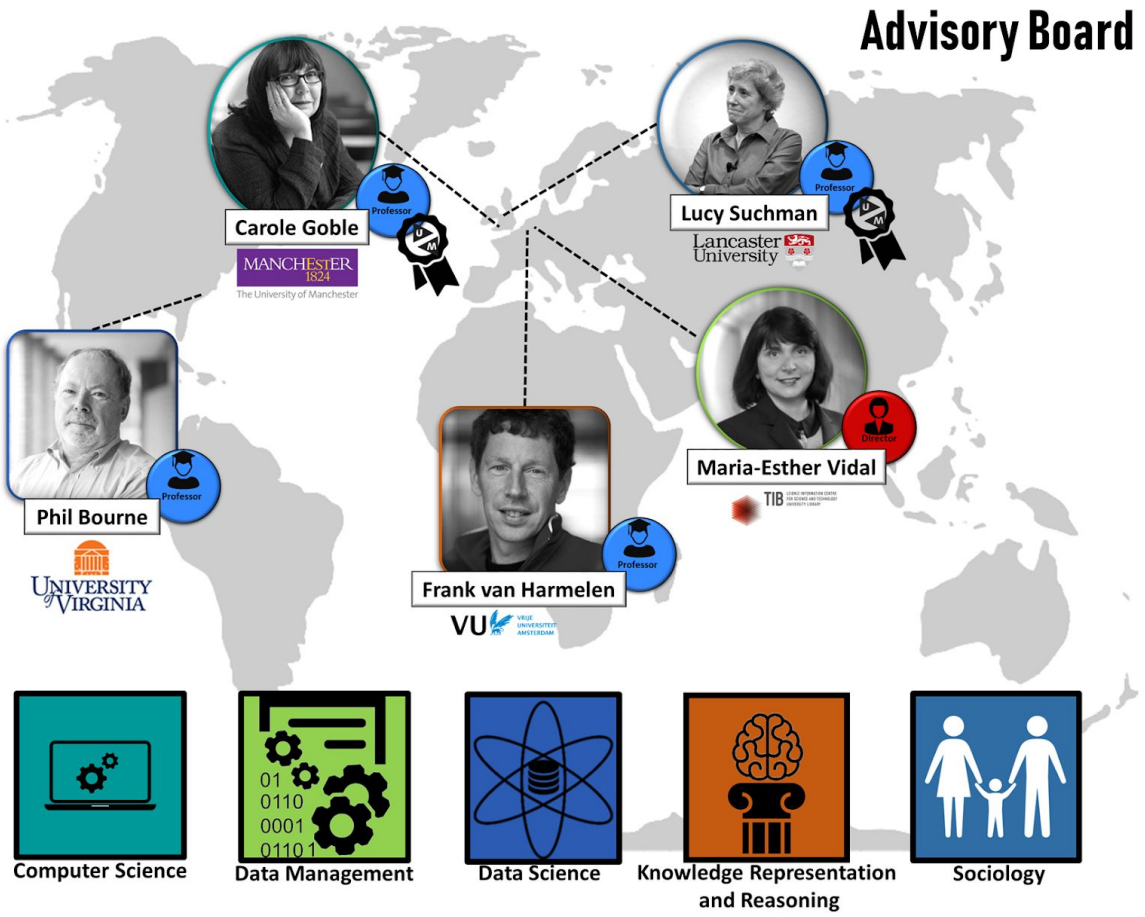
Award. In addition, in 2002 she received the Sun Microsystems Award for Significant Achievements in Advancing Life Science Computing.

**Lucy Suchman** is a Professor at the Sociology Department and the Centre for Science Studies at Lancaster University and member of the International Committee for Robot Arms Control. She is engaged in the field of human-computer interaction to contemporary warfighting. In 2010, she received the Lifetime Research Award from the Association of Computing Machinery (ACM) Special Interest Group on Computer-Human Interaction and an Honorary Doctorate from the Faculty of Culture and Society, Malmö University, Sweden. In 2014, Lucy received the Society for Social Studies of Science (4S) John Desmond Bernal Prize for Distinguished Contribution to the Field.

Prof. Dr. **Maria-Esther Vidal** is computer engineer and both Head of the Scientific Data Management Research Group at the Simo'n Bolivar University and dean assistant for research and development in applied science and engineering. Here she leads the Semantic Web Group, whose research focusses on solving problems from databases, distributed systems and artificial intelligence. She also obtained her master in Computer Science here and became the director of direction for faculty development.
She is staff member of the Leibniz Information Centre for Science and Technology University Library
(Hannover, DE) and was a faculty research assistant of the Institute of Advanced Computer Studies (UMIACS) in the University of Maryland.

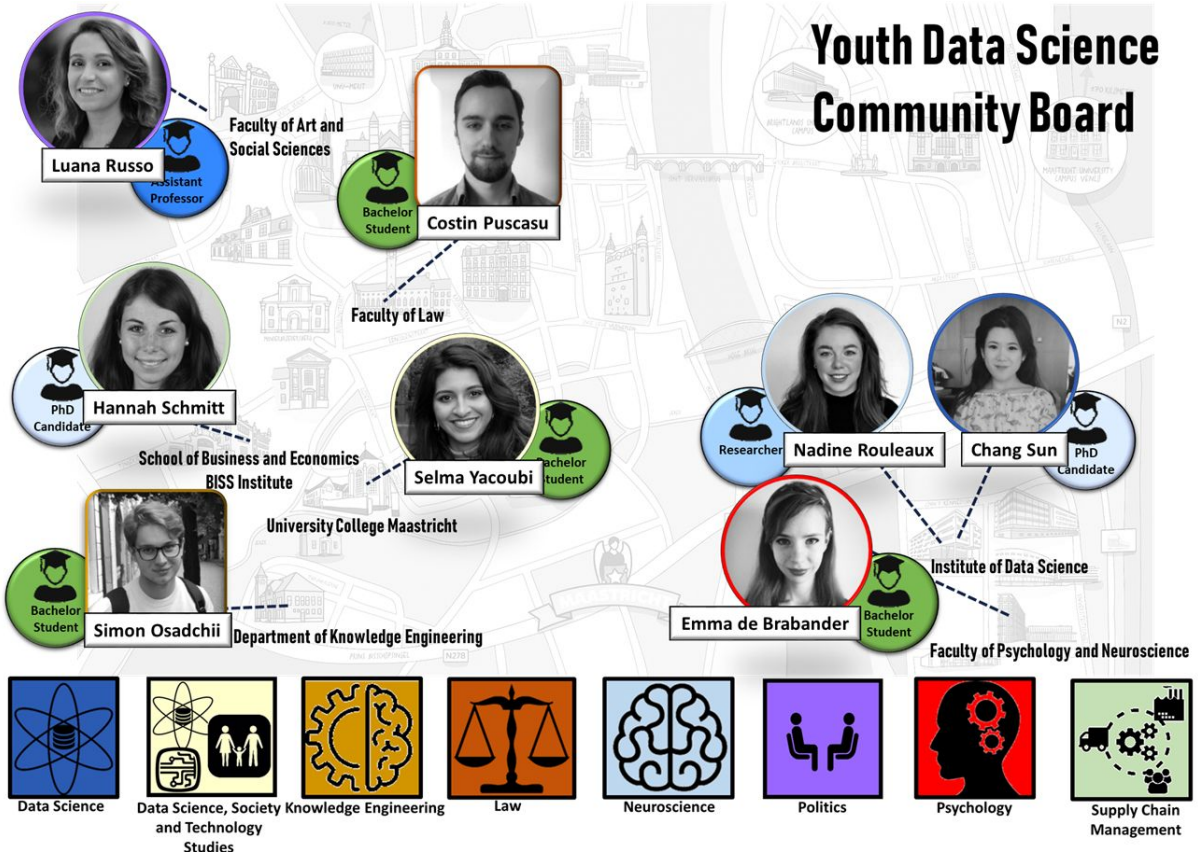
**Philip Bourne** is a researcher in health informatics, non-fiction writer and entrepreneur. He is currently Stephenson Chair of Data Science and Director of the Data Science Institute and professor of Biomedical Engineering and was the first Associate Director for Data Science at the National Institutes of Health, where his projects include managing the Big Data to Knowledge initiative, and formerly Associate Vice Chancellor at UCSD. He is a strong supporter of open-access literature and software and has contributed to textbooks. His diverse interests have spanned structural biology, medical informatics, information technology, structural bioinformatics, scholarly communication and pharmaceutical sciences.

**Frank van Harmelen** is scientific director of The Network Institute and Professor in Knowledge Representation & Reasoning, leader of the Knowledge Representation and Reasoning group and part of the AI research group at Vrije Universiteit Amsterdam. He obtained a PhD on meta-level reasoning, where he co-developed a logic-based toolkit for expert systems and worked on proof planning for inductive theorem proving.
He pioneered in both the development of the Semantic Web and large scale reasoning engines and was co-architect of Sesame. Frank is a fellow of the European AI Society ECCAI and member of the Academia Europa and the Royal Netherlands Society of Sciences and Humanities.




The Institute has previously been advised by an **Internal Advisory Committee** that was formed out of the selection committee of the Distinguished Professor. The role of the Internal Advisory Committee was to provide feedback and advice on the strategy and directions of the Institute in the context of activities ongoing across the University. This Advisory Committee will be reformulated in 2019 as a general forum to share progress and plans for data science research, teaching, and community building across various units the University.

**The Data Science Youth Community Board: Y-DS Board** is an interfaculty board that enables the perspective of youth to be brought to the Institute management. The board will advise on student perspectives relating to research, education, visibility and community building. It will organize activities geared towards engaging youth in data science. The target groups are Bachelor & Master students, PhD students, Post-docs, and young professionals from all UM faculties with an interest in Data Science. The Official Launch will be in Q1 2019. Members of the YDS Board are described in the figure below:




As there are many stakeholders which are highly valued by IDS, in addition to the formal structures described above, diverse multidisciplinary **working groups** (research, education, marketing & communication, HR & recruitment, Research IT, Funding, Finance) have been formed. These help to gather information and advice for governance and decision making by the IDS management team.

To organize and structure the administrative and operational procedures **regularly meetings** in line with the planning & control cycle of the incubator faculty are scheduled.

| **Meeting** | **Frequency** | **Purpose** | **Composition** |
|---|---|---|---|
| IDS Strategy | monthly | Accountability & progress based on assignment- content related | Martin Paul, Michel Dumontier, Claudia van Oppen |
| IDS Operations | monthly | Accountability & progress based on assignment- operations & infrastructure related | Nick Bos, Claudia van Oppen |
| International Advisory Board | annualy | Advise IDS management in terms of strategic positioning, latest developments, collaborative opportunities, visibility, and outsider's perspective | Carole Goble<br>Lucy Suchman<br>Maria-Esther Vidal<br>Frank van Harmelen<br>Phil Bourne |
| University Data Science Board | annualy | Share current and future plans for research, infrastructure. | 1-2 from each faculty<br>To be determined |

| | | | |
|---|---|---|---|
| Y-DS Board<br>Youth Data Science Community Board | quarterly | Advise IDS management on youth perspectives and develop youth engagement activities as it pertains to research, education, and community. | Chang Sun,<br>Nadine Rouleaux,<br>Hannah Schmitt,<br>Emma de Brabander,<br>Selma Yacoubi,<br>Costin Puscasu,<br>Luana Russo,<br>Simon Osadchii. |
| Executive management | weekly | Progress & planning | Michel Dumontier, Claudia van Oppen |
| Staff meeting | weekly | Progress & planning | IDS staff |
| Host Faculty | monthly | Coordination | Thomas Cleij, Bakir Bulic, Michel Dumontier, Claudia van Oppen |
| Faculty outreach | quarterly | To discuss progress and plans as it relates to each of the 5 faculties. | Michel Dumontier, Claudia van Oppen |
| DKE – IDS management | monthly | Collaboration in research & education | Gerhard Weiss, Michel Dumontier, |

# Appendix A Data Science Research Seminars 2018

| Date | Unit | Title | audience |
|---|---|---|---|
| 22 Feb | M4I | Paul van Schayck: Accurate event localisation for direct electron detection using a convolution neural network | 20 |
| 22 Mar | BigCat | Martina Kutmon: WikiPathways | 30 |
| 19 Apr | FASOS | Eli Sapir: Data, linking and the social sciences: end-user-driven procedure for data linking | 27 |
| 17 May | BISS | Deniz Iren, Rudolph Muller: Artificial Emotional Intelligence for Investment Decision Support | 20 |
| 21 Jun | FASOS | Eli Sapir, Tamar Sharon, Sally Wyatt, Michel Dumontier: Big data and the end of theory | 36 |
| 19 Jul | MAASTRO | Leonard Wee: Shared Decision Making in radiation oncology<br>Timo Deist: An empirical comparison of classifiers in radiation oncology | 28 |
| 20 Sep | GSBE | Ines Wilms:  Exploring dependencies in high-dimensional time series<br>Paul Hünermund: The Causal AI Revolution. What's in it for business and policy-making?<br>Mark Graus: Data-Driven Personalization: Adopting a User-Centric Approach | 38 |
| 18 Oct | LAW | Manon van Roozendaal: Algorithms: Teenage Troublemakers of EU Competition Law<br>Marcel Schaper: Network- ing analysis of EU Case Laws | 30 |
| 15 Nov | TGX | Tim Kuijpers: DynOVis: a new tool to visualize | 35 |
| 20 Dec | DKE | Christof Seiler: A Multivariate Mixed Model for Single-Cells Biology<br>Siamak Mehrkanoon: Learning from Partially Labeled Data | 35 |