SIRIUS

ANNUAL REPORT 2017
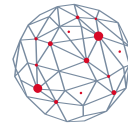
Contents

Summary

SIRIUS is a Centre for Research Based Innovation for Scalable Data Access in the Oil and Gas Domain. It conducts interdisciplinary research leading to innovations that will advance and support digitalization in the oil and gas industry.

SIRIUS started in November 2015 and has now finished two full years of work. 2017 has been spent building up the first tranche of SIRIUS PhD projects and working with partners to define the prototyping and pilot projects that will give the centre its innovation results. This process has been successful, and by the end 2018 we will have initiated an ambitious set of collaborative innovation projects that address pressing industry challenges.

SIRIUS combines public funding for basic research with funding from its industrial partners in an eight-year programme of industrial, research-based innovation. Its partner consortium consists of companies across the oil and gas value chain: including operators (Statoil), service companies (Schlumberger and DNV GL) and IT companies (Computas, EVRY, Dolphin Interconnect Solutions, fluid Operations AG, IBM, Kadme, Numascale, OSISoft and SAP). These companies work with researchers from the University of Oslo, NTNU, University of Oxford and Simula Research Laboratories. These researchers bring expertise in knowledge representation, natural-language technologies, databases, scalable computing, execution modelling and analysis and working practices.

The centre provides the industry with better ways to access and use the massive amounts of data that are generated in exploration, projects and daily operations. The advent of big data, digitalization, the internet of things and data science has made problems with data access more acute. Data access is the bottleneck that can prevent successful digitalization. SIRIUS approaches these problems through interdisciplinary work, as successful innovation requires a combination of methods. Technical innovation is generated through a portfolio of projects approved by the centre's strategic board. Basic technology is developed in laboratory projects. This technology is then moved through prototypes into pilots that solve business problems in industrial applications, focusing on challenging and industrially important beacon projects. The centre has an intellectual property model that is designed to build a core of open knowledge on which commercial solutions can be built.

SIRIUS aims to be an intellectual hub for applied industrial IT in Norway. It has dedicated premises in the Department of Informatics at the University of Oslo and collaborates widely, both nationally and internationally. Many of the results generated by the centre are also applicable outside the oil and gas sector. The centre is therefore also active in communicating and applying its results to other sectors, such as Healthcare, Energy, Earth Observation, Manufacturing and Public Administration, through its cross-domain applications work package.

This report gives an overview of the centre's ambitions, scientific programme and activities for 2017.

The Dilemmas of Digitalization

The world consists of dilemmas. And we in SIRIUS are not exempted.

Knut Sebastian Tungland

Knut Sebastian Tungland is a Senior Advisor in Statoil. He works within the digitalization unit of the CIO office and is Chairman of the SIRIUS General Assembly. He has spent 25 years at the intersection between software and the energy industry: with software products, as a consultant, and with Statoil.

These dilemmas are challenging but can also be opportunities.

One of these is the gap between research and industry. Researchers ask: "Tell us your problems and we will apply research results to find solutions". They expect the industry to be able to articulate its business challenges in a comprehensible language. They also expect the industry to have some intuition about the problems researchers can help solve. On the other hand, the industry – the consumers of technology – asks the researchers: "What can you do for me?" They expect that the researchers can articulate how their know-how and technology can assist industrial operations. Thus we face what is known as an ill-structured problem.

How do we resolve this dilemma? One way is to engage in longer conversations. We need to follow up initial questions with dialogue and open questions that delve deeper into both business challenges & opportunities and know-how & technology. In SIRIUS, we need to strive to express ourselves in terms that can be grasped by the "other party" and write in a language that communicates and inspires beyond its own environment. We also need technology demonstrators that show the capabilities inherent in technology and know-how.

Another dilemma is that improved efficiency creates both winners and losers. The International Energy Agency[1] predicts that widespread use of existing digital technologies could reduce oil and gas production costs by between 10% and 20%, and that technically recoverable oil and gas resources could be increased by around 5% globally. The power sector may save 5% of the total annual power generation cost through use of digital technology. Part of this story of increased efficiency is that some companies will lose business unless they change. This applies to operators, contractors and technology providers.
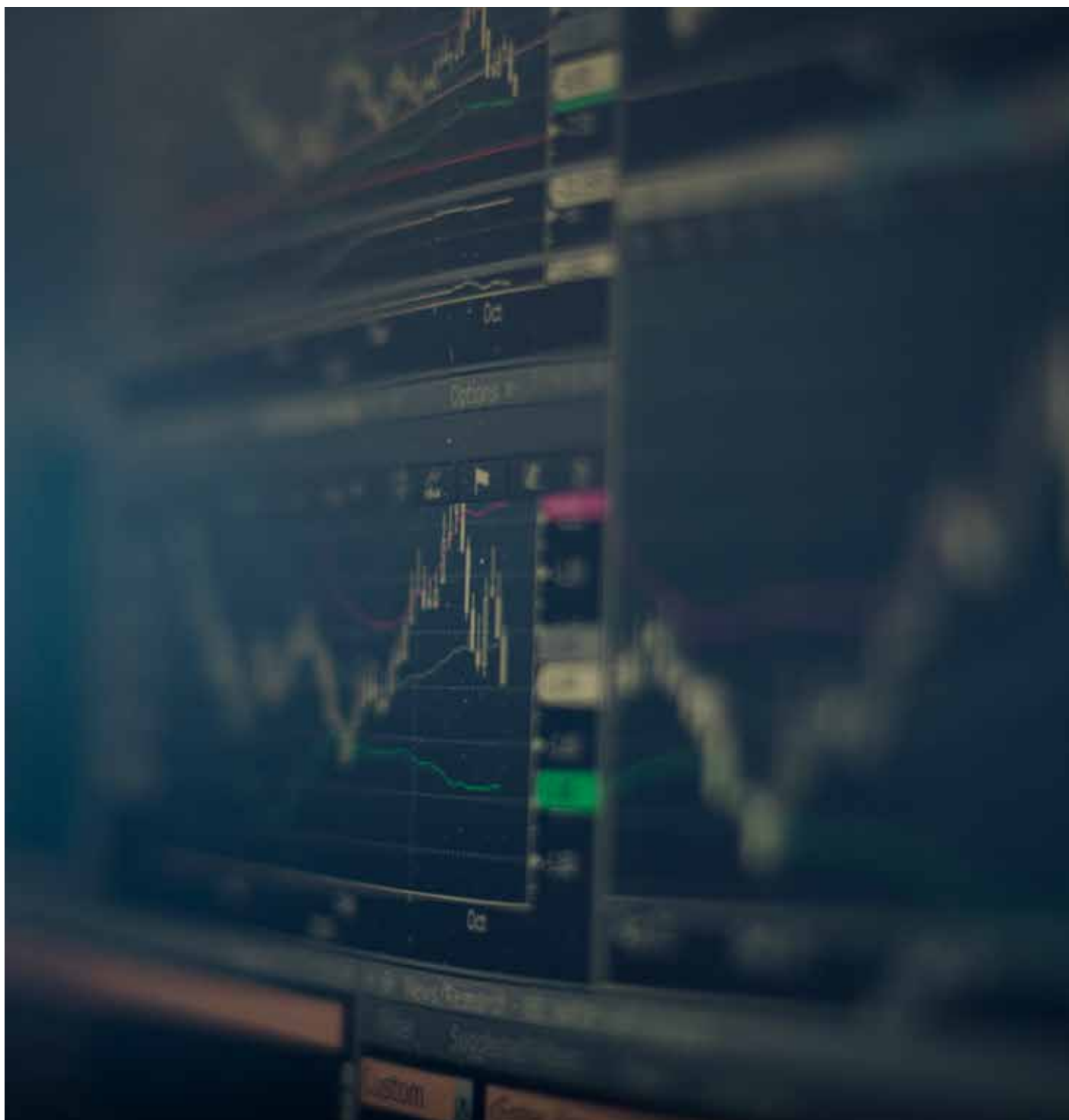
How is this relevant to SIRIUS? We are a group of industrial actors, where each individual is employed by a company with the goal of improving their company's role and position. We need to be aware of this dilemma. Academics need to understand and respect the forces and agendas that drive each company's representative. As industrial partners, we must grasp the opportunities given by new technology and know-how and not cement an inefficient status quo.

We want SIRIUS to bridge the gap between research and industry by being a place where the long and fruitful conversations can take place. We want SIRIUS research partners to engage in showing how research and know-how can improve efficiency in the industry. We want researchers to establish demonstrators that show how things can be done. And we want the industry partners to provide problems and data that allow the research community to discover and develop this new knowledge. At the same time, we want to increase the awareness of knowledge and technology among the industrial partners. This will create useful products and business opportunities in a competitive environment.

[1] http://www.iea.org/digital/

Building Norwegian Competitiveness

In February 2018, KonKraft published recommendations on how to improve the competitiveness of the Norwegian Oil and Gas industry. This report emphasizes improved digitalization through standardization, sharing of data and new modes of interaction. Its recommendations confirm that several projects in the SIRIUS portfolio are spot on in meeting what the industry itself identifies as challenges that need immediate attention.

KonKraft aims to develop national strategies for the petroleum sector and has an industry-wide mandate. The recommendations in the recent report spring out of a process where key stakeholders in the oil and gas industry were consulted. Several SIRIUS partners contributed to the KonKraft work and SIRIUS' leadership engaged in two meetings with the KonKraft working group. Given the current focus on digitalization there is every reason to believe that the KonKraft report (http://konkraft.no/konkraft_statement) will be carefully read in the industry.

One of the detailed points in the KonKraft report stresses the importance of the Norwegian Petroleum Directorate's DISKOS database and recommends measures to improve its usability. In the SIRIUS beacon Access to Exploration Data we attempt to do exactly this. One of our activities is to deploy techniques from ontology-

> KonKraft is at the collaboration arena for Norwegian Oil and Gas, Norwegian Industry, The Norwegian Shipowners Association, the Norwegian Confederation of Trade Unions (LO), The Norwegian United Federation of Trade Unions (Fellesforbundet) and the Industri Energi union.

SIRIUS is exactly aligned with KonKraft's recommendations.

based data access to improve interoperability and access to DISKOS in a project where KADME, one of the suppliers in the current DISKOS operating consortium, plays a key role. On the research side, our new colleagues from the Department of Geosciences collaborate with the Work Practices and Knowledge Representation strands. We also plan to design an information portal for subsurface data in response to demands from Statoil. Both of these projects can deliver innovations quickly. On a longer horizon we believe that SIRIUS' work on using natural language technologies to produce structured data from exploration reports, done in collaboration with IBM and Statoil, can be applied to shared data resources like DISKOS.

In digitalization of field development, the KonKraft report recommends improved standardization and sharing of asset information through use of digital twins. To this end, KonKraft recommends acceleration of the current NORSOK Z-TI digitalization project. A key goal of the Z-TI project is to develop a method for capturing requirements in a structured form that allows effective verification, analysis, and sharing. Z-TI aims to achieve this by deploying cutting-edge knowledge representation techniques, and in particular automated reasoning over OWL ontologies. SIRIUS is deeply involved in the Z-TI project through DNV GL and Statoil, supported by the SIRIUS management and the Knowledge Representation strand. SIRIUS' contribution to Z-TI is at the core of the SIRIUS "Digital Thread" beacon project.

SIRIUS follows up on digital twins in a separate beacon project that targets both design and operations. The "Digital Twin" beacon was initiated in 2017 with several partners

involved, notably Computas, DNV GL and OSIsoft. More detail about this initiative is given on page 62.

On a broader level, the SIRIUS project portfolio is all about using cutting-edge techniques from our research strands to advance digitalization in the industry through improved access to data, sharing of information, and analysis and simulation of complex scenarios. In this, we are exactly aligned with KonKraft's recommendations.

Arild Waaler

Arild Waaler is director of SIRIUS. He has over the last decade been a driving force for establishing collaboration between academic research and industry, both initiating and leading several cross-sector projects including Optique. In SIRIUS his main research interest lies in the Knowledge Representation strand. After finishing a civ.ing. degree in 1989, Waaler graduated as dr. philos. in Philosophy in 1995 within Philosophical logic. He then held research positions at University of Oxford, in logic, and at the Søren Kierkegaard Research Centre in Copenhagen with a project in the History of Philosophy. After a teaching post at Finnmark College, Waaler became professor at the Dept. of Informatics at University of Oslo in 2006, when he started to build up the collaboration that eventually turned into SIRIUS.

Beacons for Industrial Digitalization

2017 was the year in which digitalization took off. You could attend a conference on digitalization nearly every week. Companies appointed Chief Digital Officers and created digital centres of excellence. Breathless speeches warned us of exponential change and disruption. Data was called "the new oil" and *The Economist* **newspaper caught the zeitgeist by putting this on its cover in May 2017. This metaphor is particularly relevant in Norway and has become linked to a deeper debate about the country's economic future. SIRIUS' plans and ambitions lie at the intersection of the "old oil" and the "new oil".**

Digitalization is a vague and all-encompassing concept. Most definitions involve the idea of integrating digital technologies into normal life, business processes and work practices. Gartner Group's definition is very business oriented: "Digitalization is the use of digital technologies to change a business model and provide new revenue and value-producing opportunities; it is the process of moving to a digital business."[1] In SIRIUS, we believe that scalable data access is necessary for this integration. Poor data access can defeat digitalization.

Companies feel that they need to move fast. Management expects agile projects and exponential results. Results are expected after six months. Quick wins are emphasised. How does this fit into a program for research-driven innovation, with an eight-year horizon and built on four-year doctoral projects?

SIRIUS addresses this challenge in two ways. Firstly, we work with the digitalization organizations in our partner companies to identify gaps and challenges in the systems they are using. This feeds into our research agenda in two ways: a fast feedback loop defines prototyping projects with a one-year duration, while a slower loop feeds into our portfolio of doctoral projects. The second way is more strategic. Together with our partners, we have chosen a set of beacon projects that address important issues in the digitalization of the oil and gas industry. Our ambition is that, for each beacon, we will run a pilot project that looks beyond current practice to where the industry and technology will be in 2024.

The beacon projects are aligned with business problems through SIRIUS' work packages. In this way, the projects reflect the special features of each work package. The chosen beacon projects are:

- **Exploration Work Package:** Exploration and Sub-Surface
  - **Digital Support for Exploration Processes.** We are working on providing a geological assistant that will support a geologist in her work flows. See page 50 for more details.

  - **Access to Exploration Data.** SIRIUS is building on the Optique platform for ontology-based data access to demonstrate how repositories like DISKOS can be developed into digital platforms for exploration, research and innovation. More on this can be found on page 48.

  - **Digital Field and Reservoir Management.** SIRIUS is already doing work on reservoir modelling. This is described on page 52.

- **Operations Work Package:** Operations, Logistics and Engineering
  - **Planning** for projects, commissioning, maintenance and logistics. See page 58 for more details.

  - **Digital Twins:** mastering the complexity of relating digital representations of things to reality. SIRIUS' work in knowledge representation, simulation and data science supports this. See page 62 for a description of planned work.

  - **Digital Thread:** keeping track of design and requirements through the entire lifetime of a product, component or system. This is described in more detail on page 8.

- **Cross-Domain Applications Work Package.**
  - **Personalized Medicine:** application of SIRIUS'

David Cameron

David Cameron is the Centre Coordinator for SIRIUS. He is a Chemical Engineer with 30 years of industrial experience in the metals, chemicals and petroleum industries. His technical specialization is the application of simulation to the optimization of operations. He has worked in applied research, strategy and development roles in BHP Billiton, Norsk Hydro, Kongsberg Group and then in consulting and senior business development roles in IBM and Sopra Steria. He earned his PhD in simulator based process analytics from the University of Cambridge.

SIRIUS Beacon Projects

WP1 Exploration	WP2 Operations	WP3 Cross-domain applications
Digital Support for Exploration Processes	Planning	
Access to Exploration Data	Digital Twins	Personalized Medicine
Digital Field and Reservoir Management	Digital Thread	Arctic Earth Observation

WP4 Supporting technologies		
Migration to the Cloud	Sustainable Large-Scale Ontologies	Cross-cutting Technology Stack
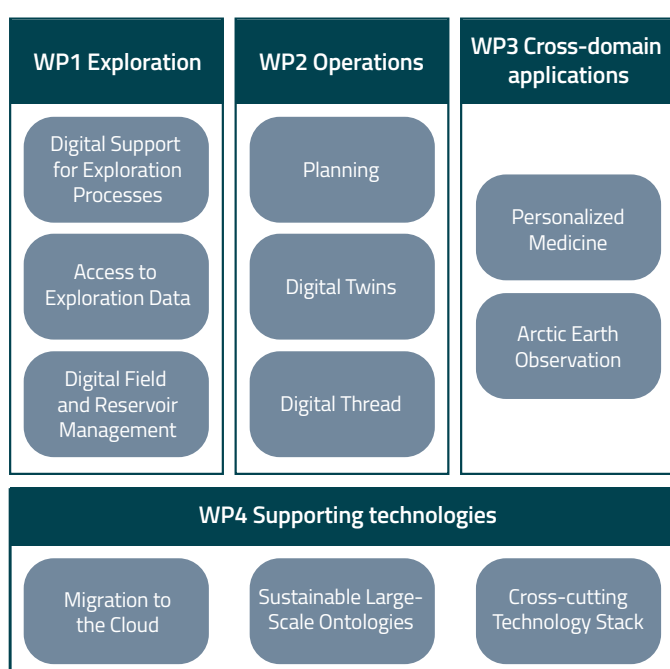
Companies feel that they need to move fast.

methods for scalable data access and analysis to challenges in health services.

- **Arctic Earth Observation:** We are working with the CIRFA Centre for Research-Based Innovation to incorporate scalable data access and cloud computing in their research and pilot systems. See page 68.

Pilots in these business areas will draw on specific focus areas in SIRIUS' methodological research. Three areas have been identified where SIRIUS' researchers can make timely and substantial contributions:

- **Migration to the Cloud.** Successful digitalization needs a well-designed and thought-out use of the cloud. Legacy applications need to be migrated to the cloud or made available as cloud services. Trade-offs need to be made between security, performance and robustness. SIRIUS is developing tools that support design, analysis and modelling for complex cloud computing systems. The ADAPT project, described on page 70, works in this area.

- **Sustainable Large-Scale Ontologies.** Scalable data access needs good models of the data and its meaning. Building and maintaining these models is complex and tedious. We are developing concepts and tools that allow data models to be built up from templates, i.e. reusable modules. Initial work on this topic is presented on page 60.

- **A Cross-Cutting Technology Stack** for scalable data access. Scalable data access methods demand powerful computers. SIRIUS is on projects that provide the best possible interaction between databases, algorithms and high-performance hardware. SIRIUS is working together with the MELODIC EU project and the recently funded eX3 project for Norwegian exascale computing. This will be a main part of the SIRIUS laboratory. See page 40 for more details.

[1] https://www.gartner.com/it-glossary/digitalization/, accessed 19th February 2018.

Vision and Objectives

Vision	
To accelerate the development and adoption of innovative data access technology in the oil and gas industry via broad-based collaboration with a short feedback loop across the whole value chain	Smart ways of finding and getting data from new and existing data sources.
	Oil companies, service companies, IT vendors and universities.
	Industry-near research and innovation: experiments, prototypes and pilots.
	Exploration, field development, operations and downstream.

Objectives	
Accelerate the innovation process for data access in the oil and gas domain	Implement prototype components in industrial pilots. Implement research results in commercial products provided by SIRIUS partners.
Transfer knowledge and expertise via feedback loop in the innovation cycle	Identify constraints imposed by existing tools. Identify opportunities for changes in work practices. Demonstrate the role of tools provided by the partners in prototypes.
Transform end-user work practices	Identify technical, social and cognitive barriers to use of technology. Identify ways to assess operational uncertainties.
Deliver scalable information systems for accessing disparate data sources	Integrate access to text, semi-structured and streaming data. Allow scalable access to large volumes of data, such as seismic data. Allow scalable access to real-time streams of sensor data. Make complex data accessible through end-user interfaces. Reduce the cost and risk of maintaining and changing systems.
Deliver a scalable, efficient and robust computational environment	Allow scalable processing and storage of big volumes of data. Process real-time streams of sensor data. Exploit affordable hardware platforms.
Reinforce mutual understanding and shared vision	Define and maintain shared vision. Establish new collaborations. Track evolving technologies and challenges. Achieve equal opportunity and gender equality within SIRIUS.
Establish SIRIUS as an internationally recognised centre of excellence	Attract additional funding (e.g. from EU programmes). Influence future research directions and funding policy. Influence society on big data, data access and digitalization. Set up a PhD track that combines research and industry skills. Influence the international research community.

Highlights of 2017

1 SAP Joins SIRIUS

SIRIUS gained an industrial partner when SAP Norway AS joined the centre from the beginning of 2018. We now have twelve partners, and with SAP we cover even more of the oil and gas supply chain. You can read more about SAP's ambitions for SIRIUS on page 22.

2 Best Paper Awards at International Semantic Web Conference

In October, the 16th International Semantic Web Conference was held in Vienna, Austria. SIRIUS researchers Evgeny Kharmalov and Ian Horrocks co-authored the paper that won the prize for "Best in-use paper".

3 Keynote and Digitalization Session at Norway-Brazil November Conference

David Cameron from SIRIUS held the keynote address at the Norway-Brazil November conference in Rio de Janeiro. This annual conference is a leading event in academic and industrial collaboration between Norway and Brazil. See page 76.

4 Best Paper at International Joint Conference on Artificial Intelligence

The paper entitled "Foundations of Declarative Data Analysis Using Limit Datalog Programs" and authored by M. Kaminski, B. Cuenca Grau, B. Motik, E. V. Kostylev, and I. Horrocks received the best paper award at the 26th International Joint Conference on Artificial Intelligence (IJCAI'17). This conference was held in Melbourne, Australia in August 2017. The paper was co-authored by SIRIUS researchers Ian Horrocks and Boris Motik. For more details, see page 36.
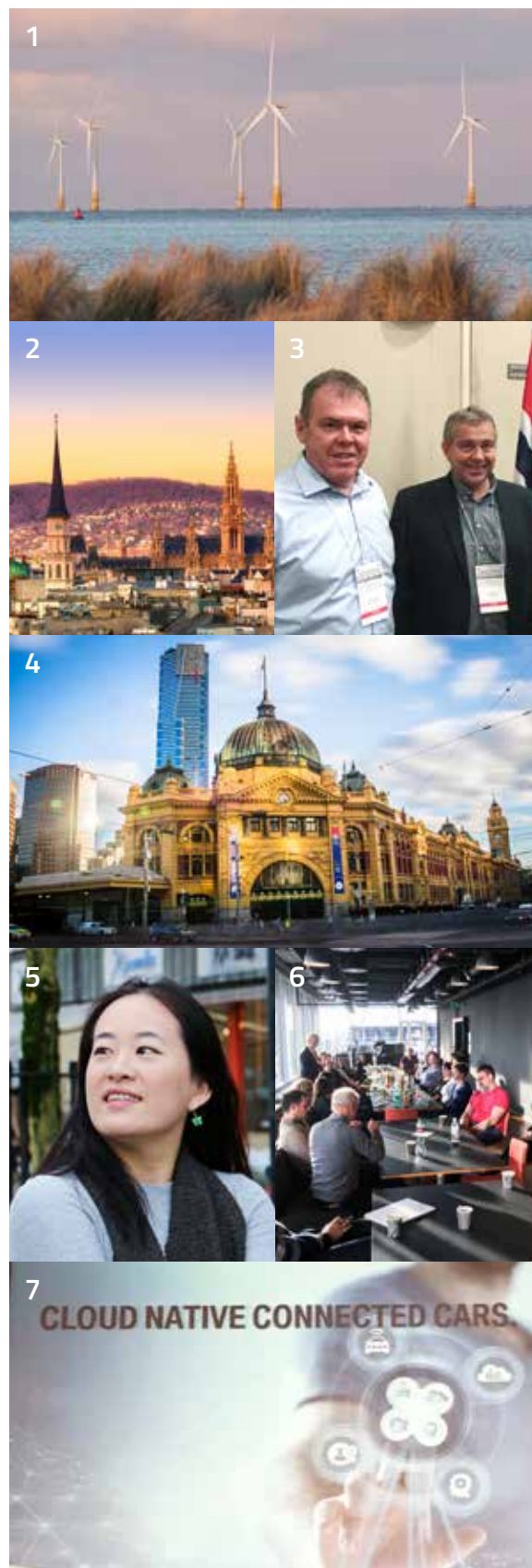
5 Mentoring Program

SIRIUS' mentoring program began in 2017 with ten mentor-mentee pairs. Feedback from both mentees and mentors has been positive. For an overview of the program, see page 78.

6 Industrial Ontology Colloquium

A colloquium has been held regularly that brings together academic and industrial users of semantic technologies in Norway. This colloquium is hosted by SIRIUS and you can read more on page 78.

7 ESOCC Conference

SIRIUS was a co-sponsor and organizer for the European Conference on Service-Oriented and Cloud Computing in September. This annual conference brought together sixty participants from all over Europe. Read more on page 74.

How SIRIUS is Organized

SIRIUS' projects are organized in business-related work packages, whereas its researchers work in technically oriented strands. This matrix organization enables us to obtain the right balance of fundamental research and focus in innovation and industrial applications. This organizational structure was put in place at the start of 2017 and has worked well.

SIRIUS has five work packages:

- Exploration: applications in exploration and subsurface applications. See page 46.

- Operations: applications in the design, construction, operation, maintenance and decommissioning of complex industrial facilities. See page 56.

- Cross-domain Applications: applications in areas outside and beyond the natural resources industries, notably health and earth observation. See page 64.

- Supporting Technologies: generic information technologies that can be applied in projects in several of the business work packages. See page 70.

- Strategy: projects that define the direction of SIRIUS' research and innovation, including education, equal opportunity and dissemination.

SIRIUS' researchers are organized in seven strands:

- Knowledge Representation, see page 32.
- Databases, see page 36.
- Natural Language, see page 38.
- Execution Modelling and analysis, see page 34.
- Scalable Computing, see page 40.
- Work Practices, see page 42.
- Data Science, see page 44.

The idea of a research strand is described on page 31.

Each work package has a manager, with responsibility for the portfolio of projects in that work package. Each strand has a strand leader, who has responsibility for the technical quality of the strand and for the researchers who work in this strand.

SIRIUS has twelve industrial partners and four academic partners. Staff from industrial partners participate in projects in all work packages, although, in the early life of the centre, much partner effort has been used in strategy projects.

SIRIUS' ambition remains to deliver innovation through prototyping and pilot projects that have clear linkages to business problems and our partners' software and hardware. These projects are organised as innovation projects. These projects are governed by a separate project agreement and usually have specific provisions for confidentiality and ownership of intellectual property. This framework allows competing companies to work in SIRIUS without risk of losing intellectual property and competitive advantage. Innovation projects require much effort from both academic and industrial participants. This means that we are working to obtain additional funding from the Research Council, the European Union, joint industry programs and companies for these projects.

The innovation projects build on and feed into the foundation of fundamental projects in the centre. These projects are long-term and are built around PhD fellowships. Our ambition is that all SIRIUS PhD fellows will have a four-year contract, where the fourth year will be used for innovation and centre-related activities. Industrial partners contribute to these projects by supervising and mentoring students and by supplying software, hardware and services as in-kind.
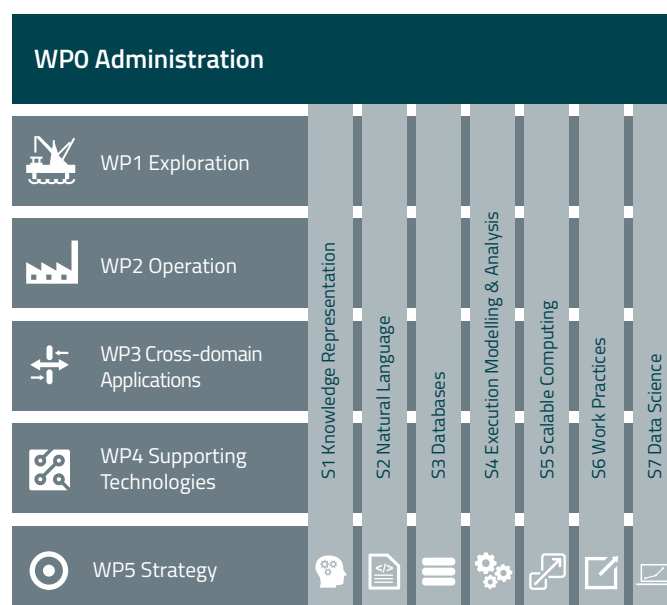
SIRIUS' Partners

Research Partners

SIRIUS brings together three universities and a research institute as research partners. The University of Oslo's Department of Informatics hosts the centre. NTNU, the Norwegian University of Science and Technology, participates through a group from the Department of Computer Science. The University of Oxford's Department of Computer Science is also a partner. Finally, the Simula Research Centre participates in work on high-performance computing and scalable computing.

UiO : University of Oslo

simula

UNIVERSITY OF OXFORD — DEPARTMENT OF COMPUTER SCIENCE

NTNU — Norwegian University of Science and Technology

Industrial Partners

SIRIUS' Industrial Partners together cover the entire supply chain for oil and gas. The consortium is a good mix of operating companies, service companies, information technology vendors and specialized software suppliers.

We are working to increase the number of partners in the centre, bringing in firms that complement the current mix of partners. This process resulted in SAP Norge joining the centre from the start of 2018. There are ongoing negotiations with several leading operating, service and software companies.

Statoil

Statoil has a long history of collaboration with academia and research institutions, and I am very pleased with the close dialogue SIRIUS enables between research communities in academia and our industry. We see SIRIUS as an enabler both for extracting and exploiting our data, and to strengthen innovation based on our data, says Åshild Hanne Larsen, Senior Vice President for IT and CIO in Statoil.

Statoil has a leading role in the SIRIUS centre. As the major operating company on the Norwegian Continental Shelf, Statoil has started an ambitions digitalization initiative. This has a starting point in a vision of how the energy and supplier industry will be in 2025. This perspective aligns well with SIRIUS' eight-year program, ending in 2023.

Åshild Hanne Larsen, CIO and SVP Information Technology, presented Statoil's view of Digitalization and SIRIUS during the launch of SIRIUS' Innovation forum at the Cutting Edge conference in Oslo in September 2017. She presented a long history of digital innovation, of which the adoption in 2017 of a Statoil data platform is but the latest initiative. Her vision of 2025 is a company with widespread use of artificial intelligence, cloud, connectivity, high capacity computing and robotics.

Statoil's digitalization strategy supports the company's three goals: Always Safe, High Value and Low Carbon. Digital technologies will allow Statoil to be an industry leader in safety, security and sustainability. Digitalized processes will give Statoil a cost advantage. Digital operations and maintenance will minimise downtime and make Statoil the most productive operator. Finally, digitalization will change the way the company works.

Scalable data access is one of the pieces in this jigsaw puzzle. SIRIUS is therefore seen as one way of aligning computer science research and innovation with Statoil's ambitious digitalization aims. During 2017, Statoil personnel have worked with SIRIUS researchers in the following scientific areas. In exploration, Jens Grimsgaard from Statoil Research has led the work package. Statoil staff have joined in defining the scope of the Geological Assistant project and projects on applying data science to exploration. In the operations work package, Statoil has been active in supplying data and defining activities in planning and logistics, requirements management and the management of operational data.

Statoil has been generous with their time and resources, and SIRIUS researchers have been brought into fruitful discussions with both senior management and technical specialists. These discussions have resulted in the development and anchoring of SIRIUS' portfolio of laboratory, prototyping and piloting projects. Senior managers in Statoil have also been mentors in the SIRIUS mentoring program (see page 78).

Schlumberger

Per Eivind Solum
Program Manager

Edo Hoekstra
DELFI E&P Environment
Manager

Schlumberger is the world's leading provider of techno-logy for reservoir characterization, drilling, production, and processing to the oil and gas industry. Working in more than 85 countries and employing approximately 100,000 people who represent over 140 nationalities, Schlumber-ger supplies the industry's most comprehensive range of products and services, from exploration through production and integrated pore-to-pipeline solutions for hydrocarbon recovery that optimize reservoir performance.

Schlumberger places strong emphasis on developing inn-ovative technology that adds value for our customers. We consistently invest more in R&D each year than all other oilfield services companies.

Digital technology now stands to bring its own disruption to our industry. Schlumberger is delivering the DELFI cognitive E&P environment for this evolving world. DELFI is a multidi-mensional environment that unites planning and operations. Bringing together advances in technical disciplines such as artificial intelligence, data analytics, and automation-underpinned by decades of unrivaled domain knowledge-the result is an E&P experience like no other.

Schlumberger has a long tradition of engaging with academia to advance technology within the oil and gas industry. We have actively supported SIRIUS from its inception. We believe that SIRIUS has great potential for advancing our industry's general knowledge of data and information technology. We want to contribute actively to achieve this together with the other partners.

Schlumberger

DNV GL is a global quality assurance and risk management company. Driven by our purpose of safeguarding life, property and the environment, we enable our customers to advance the safety and sustainability of their business. We provide classification, technical assurance, software and independent expert advisory services to the maritime, oil & gas, power and renewables industries. We also provide certification, supply chain and data management services to customers across a wide range of industries. By combining technical, digital and operational expertise, risk methodology and in-depth industry knowledge, we empower our customers' decisions and actions with trust and confidence. We continuously invest in research and collaborative innovation to provide customers and society with operational and technological foresight. With origins stretching back to 1864 and operations in more than 100 countries, our experts are dedicated to helping customers make the world safer, smarter and greener.

Rolf Benjamin Johansen, Director

DNV GL is a partner of SIRIUS with an ambition to seek industry-academia collaboration in innovation, development and deployment of new digital technologies and techniques. Scalable, effective and intelligent data access and management are expected to be key enablers in digital transformation – for ourselves, customers and industry partners.

Our partnership in SIRIUS for 2017 involved collaboration in three areas: IPR (intellectual property rights), ontologies and digitalization of requirements and initiation of possible joint industry projects (JIPs).

"With SIRIUS we aim to explore opportunities for new ways of managing requirements, rules and standards, to enable the next generation classification and technical assurance services", says Rolf Benjamin Johansen, Director – DNV GL Oil & Gas.

Joseph Rouhana, Head of
Energy and Natural Resources

Stefano Holguin,
Country Manager

SAP is the global leader in business applications and analytics software as well as a market leader in digital commerce, with over 378,000 customers in more than 180 countries. With more than 150 million cloud users, SAP is the world's largest enterprise cloud company. 76% of all worldwide business transactions touch an SAP system. Our continued growth is attributable to relentless innovation, a diverse portfolio, our ability to anticipate ever-changing customer requirements, and a broad ecosystem of partners.

SAP is at the centre of today's business and technology revolution, where the ability to innovate is the core driver for purpose-driven and successful businesses large and small. Our engagement in building an SAP laboratory at the University of Oslo aims at supporting future professionals to create innovative solutions in a smarter and faster way, and to help scale these innovations to the market. Through our collaboration with SIRIUS and the University we hope to enable and boost entrepreneurship.

"The University of Oslo is a renowned powerhouse for

innovative research and development. It is exciting to build bridges between such an innovative research environment and the enterprise world. We want to create simpler and better solutions for today's challenges and foster future opportunities", says Stefano Holguin, Country Manager at SAP Norway.

SIRIUS focuses on asset intensive industries and also on healthcare – both areas are seen as strategic by SAP as they play key roles in truly improving people's lives and making the world run better, safer and less resource-intense. SAP provides deep knowledge and the state-of-the art technology in both areas.

"Digital collaboration in networks and optimizations within operations planning and scheduling are important game-changers for our customers. Together with our partners we take responsibility to leverage the combination of cutting-edge technologies to the best people and business value", says Joseph Rouhana, Head of Energy and Natural Resources Industries at SAP Nordics.

IBM
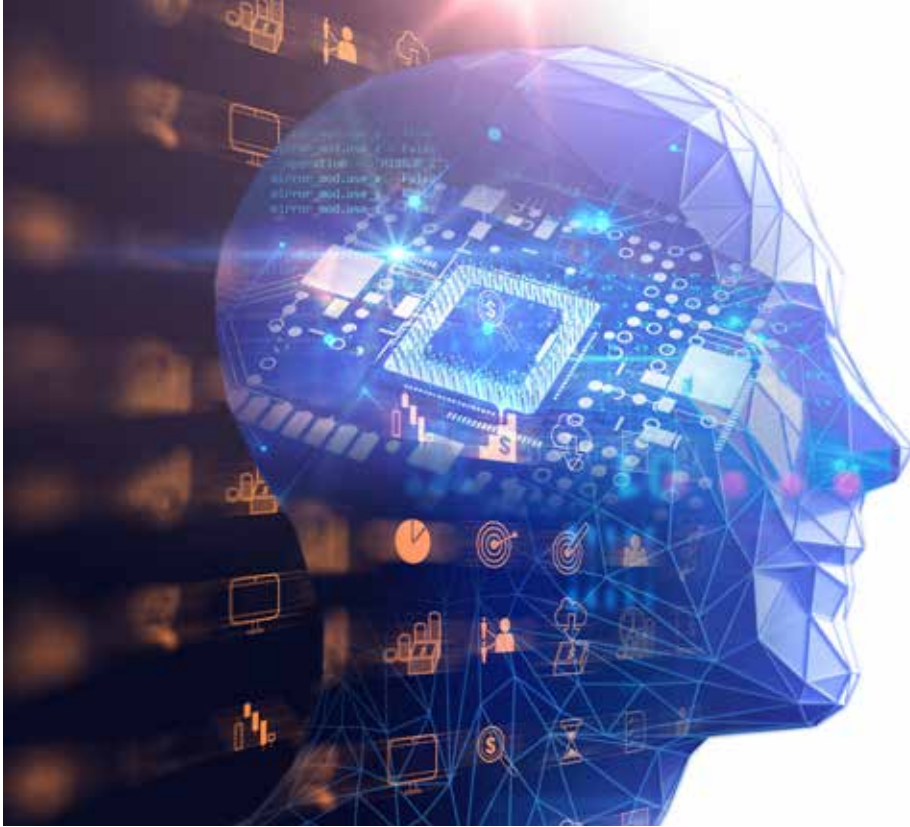
Ulisses Mello, Director of
IBM Research

IBM has been working with SIRIUS to identify opportunities for collaboration around use cases in both oil and gas exploration and production, using technologies like artificial intelligence, optimization, and cloud-based platforms. As a case in point, we look forward be being able to train our systems on relevant real-life business data. We value being able to get a variety of viewpoints on how to solve specific challenges by working with other researchers and partners.

In this collaboration, IBM has engaged with our Research division, in particular the Yorktown Research Centre in the United States and the Brazil Research Centre in Rio de Janeiro and São Paulo.

IBM Research Brazil, led by Ulisses Mello, focuses on research related to oil and gas. This aligns very well with SIRIUS' aims. For example, SIRIUS and IBM are working on a project to develop novel ways to crowdsource the task of annotating geoscience images. Images must be annotated to train machine learning tools. In this way, we want to create better AI-based tools for automatically detecting, segmenting and maybe characterizing geological images in documents. Good annotation needs annotators with high levels of geological knowledge and skill. Our aim is to support annotators with a tool that allows quick, correct and consistent annotation of images, in the first instance, geological maps.

EVRY

Kolbjørn Haarr, EVP EVRY Norway

"The level playing field that business once played upon is gone. Companies that exploit digital opportunities can change the game to their advantage. The constant flow of new digital solutions and business models provides companies with the chance to advance. But choosing the best opportunities can be business critical. At EVRY we partner with customers to create digital advantage. By introducing fresh insight and expertise we help you navigate new opportunities; helping you run your IT Systems – modernize your business – and innovate new, game changing solutions."

With a ten-thousand strong team, based in offices across Europe and Asia, we provide Infrastructure, Consulting, and, most important in the SIRIUS setting, Digital Business Development to the Nordic region. Over fifty years of success has earned us deep industry experience across many business areas, including oil and gas and healthcare. We work with the most innovative players in the market: IBM Watson, Method 'Design Thinking' and Christian Michel-sen Research on IoT. By hiring fresh top-notch candidates straight out of university ("100 digital talents"), we get access to the latest advances in artificial intelligence, robotics, machine learning and cognitive solution. From our Cognitive Centre of Excellence in Ukraine we can deliver solutions that use more powerful techniques and tools than traditional software. EVRY's Digital Transformation team in Bergen can then analyze, develop and program solutions that can drastically change the way you work, the way you interact with customers, or even your whole business model.
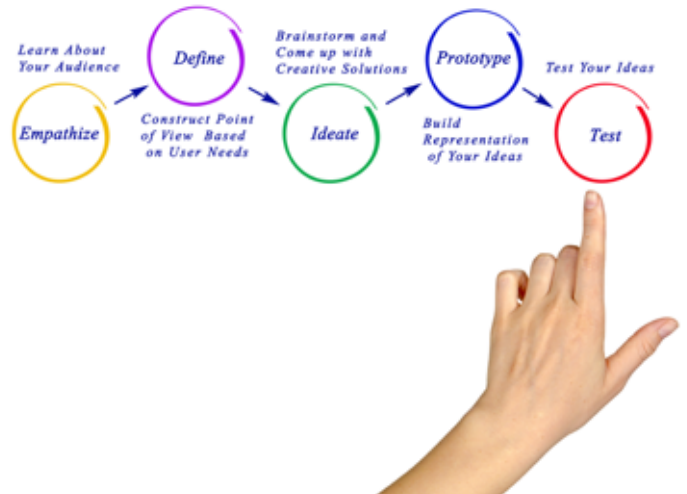
In SIRIUS context we are, together with partners, looking at new ways to exploit the vast amount of data that is stored in the National Data Repository DISKOS. By using insights and technologies from several of the SIRIUS strands, we believe it is possible to extract valuable information, correlations and hints on where to search for new patches of oil and gas. The value of finding a new Johan Sverdrup based on better data analysis of 'old' data in DISKOS, is the driving force for pursuing this holy grail. Furthermore the possibility of exploiting minor finds, better utilization of infrastructure and smarter production strategies are all 'buried' within the databank, waiting to be excavated if you have the right tools. The purpose and benefits speaks for themselves.

EVRY is currently upgrading the pre-qualification procurement database for the Norwegian Continental Shelf. This is called the Joint Qualification System and is hosted by EPIM. Through adding machine learning, intelligent search algorithms and other cognitive solutions, the new system has the potential to enhance and streamline the procurement process between 27 operating oil companies and nearly 4000 suppliers.

These are to name but two examples of the possibilities that SIRIUS opens for the extraction of value in Big Data by enhancing the access and analysis power in the oil and gas domain.

computas

Roar Fjellheim Director, business development

- Artificial Intelligence, Big Data and analytics services
- Digital business transformation and user experience services
- Architecture and project management advisory services

Computas has delivered services and complex solutions to oil & gas companies such as ConocoPhillips, TechnipFMC, Statoil, Engie, Lundin and Aker Solutions as well as to public organizations such as the Food Safety Authority, Courts of Norway, Foreign Office, Police, Welfare and Labor administration.

Computas has a technology independent position in the market and offers competency on a broad range of products and services. The company has extensive discipline networks in areas like data science, cloud architecture, software engineering & testing and project management.

Computas AS was established in 1985 as an "expert system" spin-off from Det Norske Veritas (now DNV GL) and has since built unique competency in the areas of Information Knowledge and Business process management. Computas has nearly 300 employees, of which more than 90% have an MSc or higher qualification. Head office is in Oslo, with offices in Stavanger and Bucharest, Romania. Computas is fully owned by its employees.

Computas is one of Norway's leading providers of IT-services and business critical systems for:
- Workflow, case management and collaboration solutions
- Integrated operations and compliance solutions

In 2018 Computas intends to sponsor an industrial PhD candidate and to contribute to at least one innovation project, in both cases in the Operations domain. The initiatives are well aligned with the company's business strategies and will further strengthen our commitment to SIRIUS.

"Our participation in SIRIUS gives us a firm foothold in academia, helps us to stay ahead in technology, to grow our competency base, and to attract new talent. We look forward to working with SIRIUS partners in exciting new collaboration projects in the operations domain."

Roar Fjellheim, Director, business development

OSIsoft delivers an open enterprise infrastructure that connects sensor-based data, operations and people to enable real-time and actionable insights. As the maker of the PI System, OSIsoft empowers companies in exploration, extraction, production, generation, process and discrete manufacturing, distribution and services to use streaming data to optimize and enrich their businesses. The PI System is used to improve process, quality, energy, regulatory compliance, safety, security and asset health. Founded in 1980, OSIsoft is a privately-held company, headquartered in San Leandro, California, U.S.A, with offices around the world.

OSIsoft became involved with SIRIUS because technological trends like Big Data, Machine Learning and the Internet of Things have increased the quantity of data which can be obtained about operational assets in the oil and gas industry. There is massive potential for creating value from that data. Realizing this potential presents a challenge, which is at the heart of the SIRIUS mission. Thanks to OSIsoft's participation in SIRIUS, the organisation has gained access to, and begun collaborative research with, top academic expertise in promising areas of research. These include semantic technologies and knowledge representation, from the Universities of Oslo and Oxford.

John Matranga, Director of Customer Innovation and Academia, writes:
"OSIsoft has already trained academic staff at the University of Oslo, participated in a number of workshops targeted at

John Matranga, Director of Customer Innovation

specific problem areas, and participated in a tutorial on semantic technologies at the University of Oxford. These have been valuable activities to enhance mutual under-standing of partners' respective capabilities and interests. Over the coming year we expect the PI System to be hosted in the SIRIUS lab for exploring the role of the PI System for Big Data in operations. In addition, a collaboration with OSIsoft's research group will continue, focused specifically on technologies with the potential for contextualizing meta and sensor data at large scales."

fluid Operations AG
fluid Operations® AG (fluidOps), was founded in 2008, is based in Walldorf, Germany and is a leading supplier of semantic technologies. Its product portfolio includes appli-cations for implementing Smart Data, management of data centres, cloud deployments, hybrid clouds and Internet of Things systems.

Users implement pioneering business models based on Smart Data and innovative use cases in data centres and IoT environments. The company and its employees have been honoured with multiple awards for outstanding innovati-ons. Its clients are renowned companies in the automobile, telecommunications, IT, media, healthcare and life sciences industries as well as public organizations.

The Optique EU project used fluidOps Information Work-bench as an integration platform that supports semantic data representations and ontology-based data access. Such a platform will be a valuable component in planned prototy-ping and piloting projects.

Dolphin Interconnect Solutions' eX3 products are used to connect computers and IO systems together to create high performance computing platforms. These clusters have faster response times and transaction throughput than systems built using more conventional technologies.

Dolphin provides the SIRIUS laboratory with a composable infrastructure with high performance and low latency. Graphical processing units (GPUs), storage units using the Non-volatile Memory Express (NVMe) standard and CPU-based systems can be connected together and dynamically configured for high application performance. Dolphin's use of the Peripheral Component Interconnect Express (PCIe) standard allows resources to communicate at native speed without software overhead. The platform

also supports traditional clustering software like MPI and Berkeley socket communications.

Our work in SIRIUS is to continue to develop an infrastructure that supports pools of hundreds of compute resources and CPU systems in a virtualized environment. Fast access to storage will drastically reduce the time needed to load and store data from NVMe devices. Optimized integration of GPUs and HPC libraries will ensure the best possible performance for resource-demanding Big Data applications.

The eX3 infrastructure will be an important prototyping environment for large scale testing and benchmarking of Dolphin's infrastructure with real-world applications from SIRIUS' partners.

Numascale contributes to SIRIUS by providing expertise and software libraries for parallel programming and processing with large data sets. Numascale also manages the operations of a large, cache-coherent shared-memory system that contains 72 IBM servers equipped with NumaChip-1™, Numascale's first-generation cache coherent node controller. This system contains 1728 CPU cores and 4,7 Terabyte main memory (DRAM) connected in a 3 x 6 x 4, 3-D Torus topology. It can be used as a single partition running an instance of Linux or in smaller, logically separated, partitions. The system was initially operated as a prototype for the European PRACE initiative and has provided valuable experience for Numascale through pioneering work with Linux on such a large number of CPU cores.

"We are excited to see that our technology in the hands of bright and eager young PhD students and post-docs can contribute to development of methods, software and tools that provide overview and insight in large volumes of unstructured data" says Einar Rustad, CTO in Numascale.

KADME

KADME AS

The tag line for SIRIUS is "Scalable Data access in the Oil and Gas domain". This tag line also perfectly represents the core of KADME business and is the driver behind our Whereoil technology. KADME stands for "Knowledge And Data Management Expertise", and our team is composed of experts that have been working to improve access to data within the Oil and Gas industry, even before the creation of the company in 2002. Some of us have been involved with DISKOS, the E&P National Data Repository of Norway since its inception in 1993. Whereoil provides modern and efficient methods for managing the increasing volumes of data that is collected by this industry every day. To unlock the full potential of this data, companies need to digitally transform their business and get ready to take advantage of new technologies such as machine learning, cloud computing and artificial intelligence. Whereoil gives our customers a rapid path to achieve their digitalization efforts, with effective methods to search, access and connect this data making it available for interpretation and analysis by both humans and machines. Companies that use Whereoil include Petronas, Lukoil and Lundin. Maybe the most relevant and visible example of how KADME is helping the oil industry to modernize management of E&P data is the DISKOS data repository in Norway, which is built on a Whereoil-based solution operated in collaboration by CGG, KADME and EVRY.

As a small and successful company, KADME is constantly bringing innovation quickly into our roadmap while at the same time working hard to please the more specific needs of our current customers and business partners. In this scenario, participation in the SIRIUS centre is very important, as it gives us access to a wider forum in which to exchange and validate research ideas. In 2018 KADME will be involved in several SIRIUS projects. Our primary focus will be on a more efficient use of data from National Data Repositories and on improved methodologies to create relationships between information items and the work processes that have generated them.

To quote from Gianluca Monachese, Founder and Director of Business Development: "The collaborative environment in SIRIUS is ideal to speed up the type of research that is of common value to multiple partners and provides access to some of the best academic researchers in the many universities that participate in the project."

Gianluca Monachese, Founder and Director
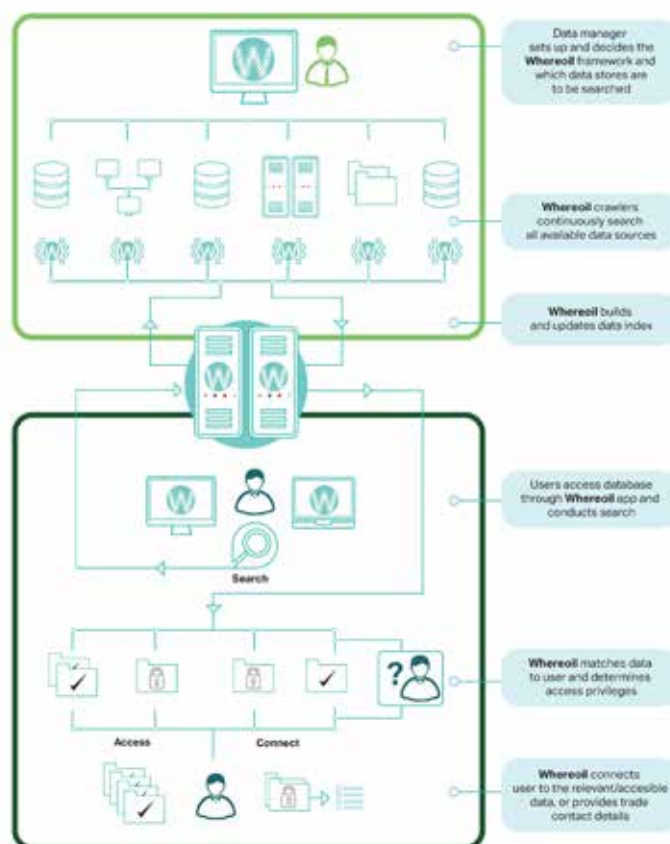
The Whereoil system delivers a seamless data solution for data managers and end users.

General Assembly
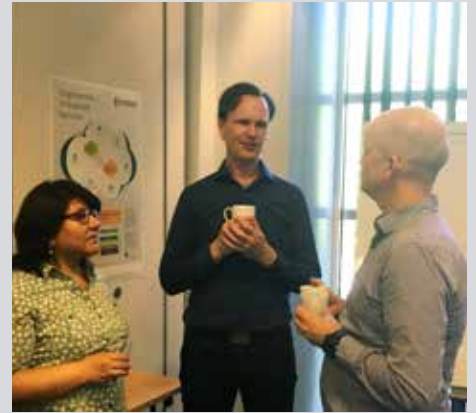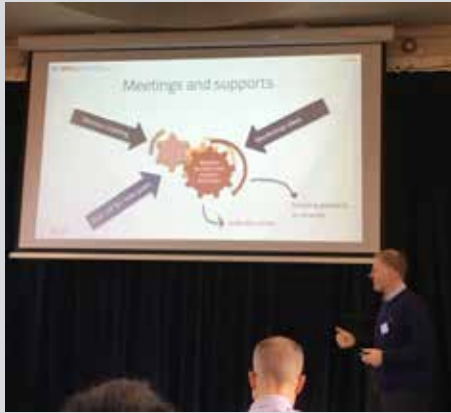
Spring General Assembly

The Spring General Assembly was held at Lysaker on 22nd and 23rd May 2017. The first day was a business meeting, while the second day was a technical meeting. On the afternoon of the first day, Ian Horrocks talked about what scientific excellence means for SIRIUS. The technical meeting focused on what the SIRIUS researchers could offer. Presentations were made about digital geosciences, planning and scheduling and data science. PhD students and post-doctoral researchers prepared posters and there was a very active discussion session after lunch. Finally, Arild Waaler and Ian Horrocks talked about two initiatives for commercializing software from SIRIUS.

Autumn General Assembly

The Autumn General Assembly was held in Oslo on 28th November 2017. This meeting had two purposes. The first of these was to approve the accession of SAP as a new partner. The rest of the meeting was built around reviewing and approving the contents of the work plan for 2018.

Presentations were made about digital geosciences, planning and scheduling and data science.

The Research Council held a visit to the centre in April 2017. This was a half-day meeting with the centre management, the University of Oslo management and SIRIUS' industrial partners. During this meeting, it was observed that the general assembly was the only formal means for partner supervision of the centre. This forum was too large and met too seldom to allow effective control of the centre's strategy and operations.

It was therefore proposed that an Executive Committee be set up, to whom the General Assembly could delegate its authority. The committee consists of four industrial representatives, one from Statoil (as chairman), one from Schlumberger, one from the large system integrators and another from the smaller vendors. The academic representatives are the Centre Director (Arild Waaler), and the Pilot Strategy Coordinator (Eric Monteiro).

The Executive Committee was elected at the May General Assembly and had the following representatives: Knut Sebastian Tungland from Statoil, Edo Hoekstra from Schlumberger, Frode Myren from IBM and Einar Rustad from Numascale. Deputies were Per Eivind Solum from Schlumberger and Nicolas Peels from OSIsoft.

The Executive Committee meets monthly to review project status and focus on chosen topics: individual projects, initiatives and strands.

SIRIUS' Strands

Each SIRIUS researcher has his or her own core research area and linkage to a research group. This specialization is needed to ensure the academic depth of their work, but it can be a barrier to solving complex industrial and innovation problems. Industrial challenges require interdisciplinary solutions. Our experts need to work together across group boundaries. At the same time, each researcher needs a deep technical grounding in their specialization. We attempt to resolve this tension by organizing our researchers into technical strands, while projects are organized into business-oriented work-packages. Each researcher has the

academic home they need, while being open to the business and innovation challenges of the SIRIUS projects.

Each strand in SIRIUS has developed a technical strategy. This is based on a vision of how their technology will have developed in 2024. This technical vision can then be used to stimulate business visions for the same time. In the following sections each strand explains what they do, what they have achieved in 2017 and what their vision is for 2024.

Knowledge Representation

Data is always related to things or ideas. Getting access to data is made much simpler if we are able to link that data to its underlying thing or idea. The things and ideas related to a certain discipline or system, together with the linkages between them, constitute the domain knowledge for that discipline. Given a model of the relationships and connections between these things and ideas, we can reason about the related data, make predictions and check for errors and inconsistencies. This discipline, which lies in the borderlands between philosophy and computer science, is called knowledge representation.

To use a common and simple example: a pump is a widely used thing in the oil and gas industry. It is linked to other things, pipes, fittings, a motor and its various sensors. Each sensor produces measurements: data on things like speed, torque, liquid flowrate, inlet pressure, outlet pressure and temperature. The pump object is also linked to ideas like specifications, requirements and model numbers and things like maintenance orders and invoices. The data about a pump is stored in many databases: data historians, ERP systems and design databases. Knowledge representation consists of creating a useful model of things like this pump and using this model to improve access to data about this pump. This approach is essential to mastering the data produced from the Internet of Things and from complex, unit-operation based systems like oil platforms. The theory and methods necessary for these applications are supplied by the Knowledge Representation strand.

What we do

The Knowledge Representation strand specialises in machine-processable representations of domain knowledge. Often, these representations take the shape of an ontology, a tractable representation of a domain vocabulary (things and ideas) and certain kinds of facts (linkages between things and idea) concerning the domain. We are interested in using ontologies in a variety of contexts:

- Data integration, where ontologies mediate between data sources by lifting from the level of data to the level of information.

- Data access, where ontologies can be used to present information to end-users.

- Requirement management, where ontologies are used

Knowledge Representation uses digital domain models to link data to the real world

to represent, check, and solve complex combinations of definitions and constraints, e.g. on engineering designs.

We develop ontology-based tools and methods for these and other applications. A particular focus is on tools and methods needed to set up and maintain necessary artefacts, like ontologies and mappings, that control the remaining tool chain.

Knowledge Representation 2024 – Our Ambitions

Integration of information sources. From previous research, we know that ontology-based techniques have a strong potential as enabling technology for data integration. In coming years, we will mature this approach and verify its applicability in a range of industrial settings. This will include further insights into entity alignment, schema (ontology) alignment, the treatment of combinations of streaming and static data, and the connection of semantic technologies to language technologies.

Construction and maintenance of ontology-based systems. Ontology-based tool chains rely on a number of artefacts like ontologies, mappings and query collections, which are rather high-level and declarative, compared to conventional solutions. Still, these artefacts need to be created and maintained, like other software. Our goal is to provide a proven set of technology and methods for the engineering of ontology-based solutions.

Ontology-driven User Interfaces. Ontologies formalise a

domain as end-users think and speak about it. That makes ontologies suitable as a driver for end-user interfaces. We have used this idea for several systems that present information or elicit information needs. We envisage that similar principles can be applied to any situation where end-users interact with information systems, be it information dashboards, entry forms or wizards. We will explore how different kinds of user interaction can be driven using the same ontology-level control.

Highlights in 2017

The OTTR template mechanism has been successfully tested in an industrial ontology application. It dramatically simplifies the generation and maintenance of ontologies based on other representations of domain knowledge.

The visual query system OptiqueVQS has been thoroughly refactored to make it easier to integrate with other systems, and easier to extend and adapt to new use cases. A scalable indexing structure was developed that restricts options presented to users in a more intelligent way.

Initial ideas for the ontological representation of requirements (e.g. legal requirements on the documentation of equipment) have been successfully prototyped and will form the basis for the next generation of the NORSOK Z-TI standard.

The KR strand is involved in several new projects in the Exploration work packge. One aims to establish a data access solution for the Diskos National Data Repository for Petroleum data, based on an ontology that captures geological domain vocabulary and knowledge. Another project will build a tool that uses an ontology to drive a manual image tagging process in a way that is compatible with a domain ontology.

Martin Giese

Martin Giese is a professor of informatics at the University of Oslo. He has a PhD in logic and automated reasoning from Karlsruhe University, and has spent time as a researcher at Chalmers University in Gothenburg and the Austrian Academy of Sciences before coming to Norway. For the past 10 years, he has been working in projects to bring ontologies and other knowledge representation techniques to applications in industry and the public sector. He was part of the team that attracted the Optique project, and acted as the project's Assistant Scientific Director during 2012–2016.

Execution Modelling and Analysis

The oil and gas industry is full of complex systems. These systems are technical, such as cloud systems, or organizational, like maintenance programs. Simulation and formal analysis can improve and optimize the performance of these systems. This makes operations safer, more efficient and more reliable.

Cloud computing is changing the traditional business model for Information Technology. It offers users on-demand infrastructure and applications over the Internet with pay-as-you-go pricing. This changes how software suppliers build and deliver their programs. Cloud-based applications must be adaptable, modular and quickly accessible. Furthermore, container technology allows applications to be run in communicating, yet self-contained, environments that fully exploit the elasticity of cloud. These new possibilities and opportunities come with the cost of a more complex workflow. Development decisions are intertwined with the application deployment issues. Designing a complex system is challenging: many possible settings and parameters can be tuned. Poor choices can result in system failure, high costs and displeased customers.
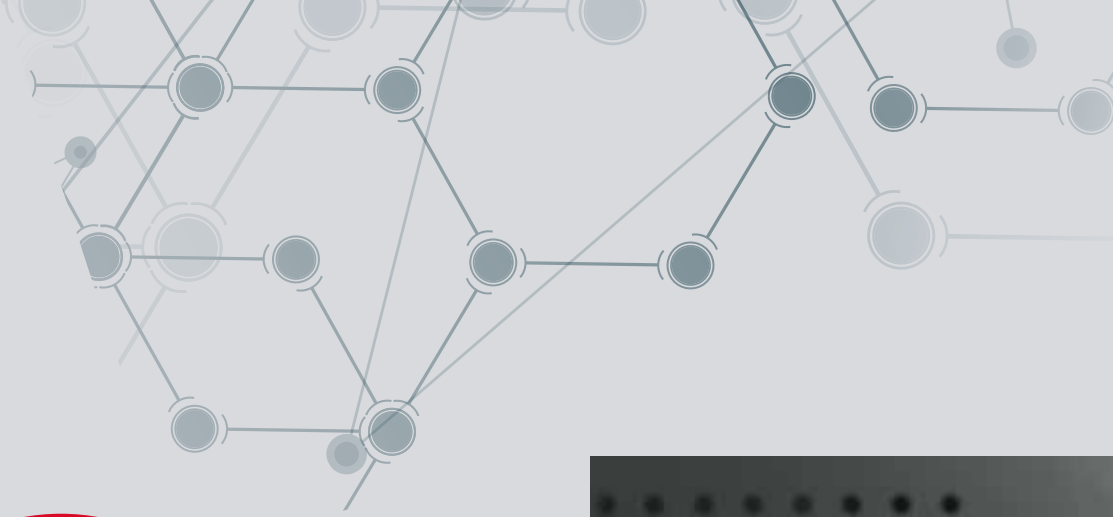
Designing a complex system is challenging: many possible settings and parameters can be tuned. Poor choices can result in system failure, high costs and displeased customers.

In 2017 we worked on supporting the people who develop and deploy applications to find the best design of their applications for use in the cloud. We developed a methodology for optimizing deployment decisions that combines optimization with system modelling. This approach involves creating an executable model of the system and using the model to search for suitable settings using an optimization-based, automatic parameter configurator.
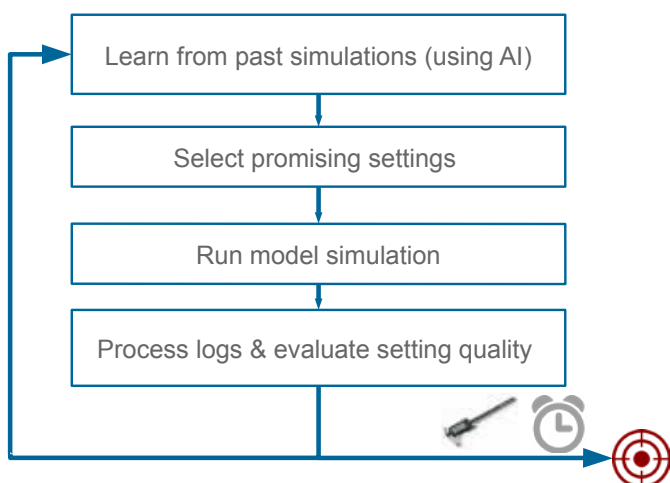
As shown in the figure, the decision process consists of a cycle, where artificial intelligence techniques are used to select better system parameters on the basis of past simulations. The model is then run with these new para-meters. This process is completed until a best possible set of parameters is found. These optimization and simulation calculations need powerful computers. We have used the Numascale cluster in the SIRIUS laboratories for this purpose. Further work is planned in 2018 to improve the speed of our tools.

Our methodology has been applied successfully in the Horizon 2020 HyVar project (**http://www.hyvar-project.eu**) to a system updating the software in motor vehicles. This

project developed and piloted a micro-service system, deployed on the Amazon cloud, that allows a supplier of on-board computers for cars to monitor and update their software over the internet. Naturally, this must be done at an opportune time to avoid inconvenience to the customer and maintain safe operation. Our tools are used to ensure that this large and real-time system meets its performance requirements.

The ABS modelling and optimization tools that we develop can also be applied to complex organisational processes, such as planning of maintenance, commissioning and logistics. This work is described in more detail on page 58.

We are starting several projects in 2018. One of them, with a high profile, is the ADAPT project (see page 70). This project aims to develop a new technique to improve data locality in parallel processing. The idea is systematically to coordinate task scheduling and data allocation, using predicted patterns of data access to memory, i.e. reads and writes access to locations in memory. The approach will combine formal models of parallel systems with basic research in programming language theory to (1) capture abstractly the interaction of workflows with dynamically created tasks and memory locations on parallel computers, and (2) combine formal analysis with model-based simulations to explore how to systemically calibrate schedulers and allocators for specific applications. The result of this project will complement EMA's current toolbox on model-based prediction.

Ingrid Chieh Yu

Ingrid Chieh Yu is Associate Professor at the University of Oslo. Her research interests include the design and analysis of concurrent, distributed, context-dependent and evolving systems. Yu worked at DNV GL as a Senior Researcher and returned to the University as Associate Professor in 2011. Yu is the leader of the Execution Modelling and Analysis strand and is also the Mentoring and Education Coordinator of SIRIUS. Moreover, she is currently one of the principal investigators of CUMULUS, an RCN funded FRINATEK research project, and the site leader of the EU H2020 HyVar project, in which she leads the design of the scalable cloud infrastructure for monitoring and individually customizing software upgrades for remote devices.

Databases

The Database strand in SIRIUS builds on observation that scalable data access is dependent on good and scalable technologies for data storage and retrieval.

The exploration and operations functions in the oil and gas industry rely heavily on Relational Database Management Systems (RDMS) and will continue to do so for the foreseeable future. Semantic technologies such as ontology-based data access can facilitate data access, but performance (of back-end repositories) may be a limiting factor. New technologies such as RDFox offer many exciting opportunities, but we need to demonstrate practicality in real use cases.

Our ambitions are to extend the capabilities and performance of database and semantic systems:
1. Increase the scale of data that can be processed.

2. Increase the complexity and heterogeneity of data that can be handled.

3. Increase the speed with which data can be handled. Use this functionality and scalability in the SIRIUS pilots.

In 2017, the database strand is still mainly focused on the development and deployment of graph databases, and in particular on RDFox. Some highlights from this year:

Declarative data analytics

Motivated by applications in declarative data analysis, we have been studying extensions of Datalog with arithmetic functions using integers (DatalogZ). This language is known to be undecidable in general, but we have identified two interesting decidable fragments: limit DatalogZ and stable DatalogZ, the latter of which is even tractable (polynomial time) in data complexity. Stable DatalogZ is interesting not only due to its tractability, but also because we have shown that it can express many useful data analysis tasks. This work was published at IJCAI-2017, where it was awarded the IJCAI Distinguished Paper Prize for best paper. Stable DatalogZ will also be the basis for implementation in the

Datalog

Datalog is a programming language for logic programming. It is also used as a language for writing queries. It is declarative, not procedural, and is well suited to computing using first order logic and working with triple-store or graph databases, i.e. databases that store triples of subject, object and predicate.

The language looks like this. Facts are expressed as triples.
`controls(PC23001A,PT23001)` states that PC23001A controls the value of PT23001
`measuresTemperature(PT23001,V23001)` states that PT23001 measures the temperature in V23001.

You can also write rules:
`controlsTemperature(X,Y) :- controls(X,Z),measuresTemperature(Z,Y)`, i.e. X controls some signal Z and Z measures the temperature of Y.

Finally, you can write queries:
`?- controls(PC23001A,X)` asks the question, "which signals does PC23001A control?"

Scalable Data Access is
dependent on good and scalable
technologies for data storage
and retrieval.

RDFox system, and we hope to be able to show that it can provide a solid foundation for declarative data analytics.

Incremental reasoning in Datalog

We developed a new algorithm for incremental reasoning in Datalog that combines the well-known DRed and Counting algorithms. The former algorithm used to be widely regarded as state of the art in cases when the rules are recursive, whereas the latter algorithm is very efficient but can be applied only to non-recursive rules. Our new algorithm combines the two in order to get the best of both algorithms. The work was published at the AAAI 2018 conference in New Orleans in February 2018.

Stream reasoning

We developed a new approach to reasoning with Datalog over streaming data. This work does not define a new language for stream reasoning; rather, the main contribution is describing how to operationalize Datalog reasoning over infinite streams. This work was published at the AAAI 2018 conference in New Orleans in February 2018. The SIRIUS PhD student, Alessandro Ronca was a co-author here.

Estimating the cardinality of conjunctive queries over graph databases

Cardinality estimation of conjunctive queries plays a key role in database query planning. We have developed a novel, principled technique for addressing this problem in graph databases. Our technique is based on graph summarisation, and its prominent feature is being able to estimate the probability of estimation error. The work has been accepted for publication at the WWW 2018 conference.

RDFox and HPC

We have conducted tests of RDFox on a Numa-0-0 computer, which has 67 nodes (each node has 6 processing units), which relative distances between that go from 10 (intra-no de distance) to 200. The tests used the Claros dataset, which is particularly interesting as the materialisation of the database increases the size of the graph more than 20 times. Performance depends on node distances, but when bound to near nodes is comparable to the SPARC T5-8 which we used in earlier tests (see AAAI paper).

Spin-out commercialisation activity

In March 2017 Oxford Semantic Technologies was set up to commercialise RDFox. The company was established with the help of Oxford University Innovation (the University's exploitation company), and with investment from Oxford Sciences Innovation (a company whose role is to invest in Oxford University start-ups). The company already has three full-time employees, is working on several commercial projects, and has successfully deployed RDFox in a large German technology company.

Boris Motik

Boris Motik is leader of the SIRIUS database strand. He completed his PhD at the University of Karlsruhe under the supervision of Prof Rudi Studer, and he is currently a Professor of Computer Science at the University of Oxford. His research interests include knowledge representation and reasoning, ontologies, automated theorem proving, databases, and the Semantic Web. His research encompasses both theory and practice, including studying and developing knowledge representation languages and reasoning algorithms, as well as practical implementation of reasoning systems. He is the main author of the HermiT reasoning system, a popular and widely used OWL 2 DL reasoner, and RDFox. He played a key role in the development of the OWL 2 DL ontology. He has an extensive collaboration network with industry partners, with whom he is working on practical applications of semantic technologies in areas as diverse as health care, energy, and travel.

Natural Language and Language Technology

Text documents make up a vital part of the information in the oil and gas industry. Computer processing is needed to use this information efficiently. Moreover, computers must have some sort of understanding of the texts. For a human, it is obvious that the words "oil", "gas", and "hydrocarbons" are related. Thus, a question regarding whether a certain field contains "hydrocarbons" should be answered affirmatively from a document which tells us that the field contains "oil". For the computer, however, this is not obvious at all. The words "oil" and "gas" are seen simply as strings of letters, which are no more related than the pair "oil" and "cheese".

Language technology aims to bring out the similarities in meaning between linguistic expressions that look dissimilar on the surface. A common approach to this is to observe a word's distributional patterns in large collections of texts. Each word is represented by an object in a mathematical model, (technically, as a vector in a multi-dimensional vector space). This representation is based on the word's distribution in the text corpus. This implies that words with similar distributions, like "oil" and "gas", will have similar (but not identical) representations in the model, while a word like "cheese" will have a quite different representation. State-of-the-art models, called word embeddings, are trained on very large collections of texts, like the English Wikipedia, with more than three billion words (roughly corresponding to 20 000 books) or the Google News corpus with 100 billion words. Word embeddings are built using unsupervised machine learning, without human intervention.

These word embeddings have brought language technology a large step forward the last few years. At the same time, we observe that the meaning of words varies from one domain of use, say health, to another domain, like oil and gas. A technical domain, like oil and gas, may make frequent use of terms which have a much lower frequency and different distributional patterns in general text. It may also use everyday terms with a different technical meaning, like "christmas tree" for an assembly of valves, spools and fittings. We therefore expect that domain-specific word models for the oil and gas domain will perform better than general models when processing oil and gas documents. We have therefore collected a corpus of scientific articles from the oil and gas domain and trained our own word embedding

model. We then tested the model on the Schlumberger oilfield glossary, which contains about 4000 terms that are manually marked for synonymy and similar relations. Our model was trained on only about 100 million words, which is a small number for machine learning. Despite this, it performed better than general models trained on corpora several orders of magnitude larger.

Even though our model did better than the general models, the evaluation scores were quite low. We therefore carried out a manual evaluation of the word embedding model. This work was done together with specialists from Statoil, who had the technical knowledge needed to evaluate the meanings of the various terms. This evaluation showed that the similarity results from the model reflected similarity in word meaning quite well.

We use other tools besides automatically created word embeddings. Word meanings can also be modelled in humanly created dictionaries that show, for example, synonyms. The Schlumberger Oilfield Glossary, which is licensed as an in-kind contribution from Schlumberger, is one such tool. Is it possible to combine the automatically created word embeddings with information from handcrafted dictionaries? To answer this, we experimented by using a dictionary to improve the word embedding model using a technique called retrofitting. We tested these models on a classification task where text documents regarding exploration had been labeled with content type. It turned out that the addition of the manually created dictionary improved the results on the classification task. The results from these experiments will be presented on the Language Resources and Evaluation Conference, 7-12 May 2018 in Miyazaki (Japan).

Jan Tore Lønning

. . . .

Jan Tore Lønning is professor of informatics at the University of Oslo. After obtaining a dr. philos degree from the Department of Mathematics at UiO in 1989, he was appointed associate professor and later full professor at the Faculty of Humanities at UiO. There he had the main responsibility for building up a study program in computational linguistics (natural language processing/language technology). In 2006, he, together with the program, moved to the Department of Informatics. His main research interest is the semantics of natural languages broadly understood, applying logic as well as machine learning techniques. He has also carried out research on machine translation.

In the autumn, we shifted attention to relation extraction from texts. Besides word meaning, an important step in understanding text is to understand who did what to whom. To localize the "who" and "whom"-s in a text is called entity extraction. After they are localized, the next step is to determine the "what" between them; e.g. in the oil and gas domain, the entities could be a specific oil company and a field, while the possible relations could include terms like HAS_LICENSE, HAS_PRODUCTION, IS_DRILLING, SEISMIC_EXPLORATION.
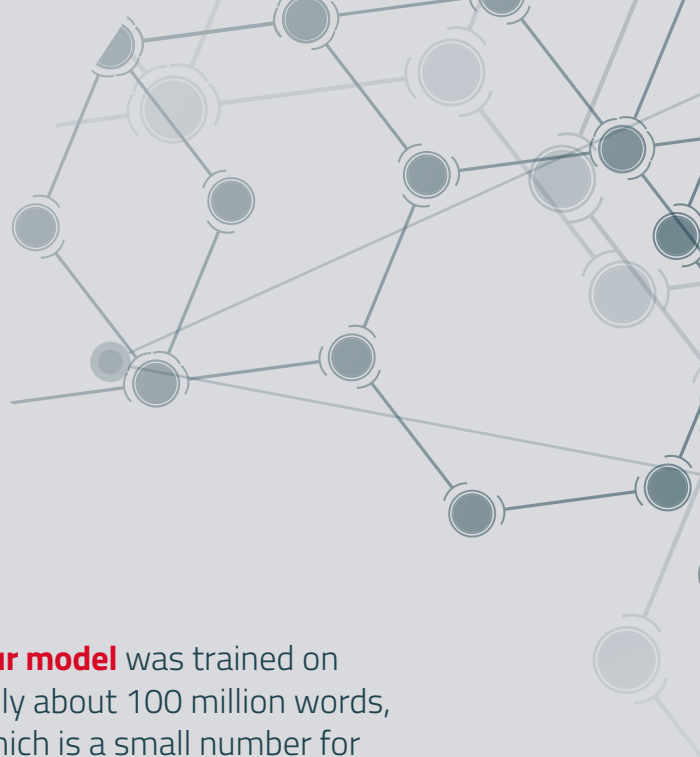
Our model was trained on only about 100 million words, which is a small number for machine learning

In language technology, many different approaches can be applied to similar tasks, and it can be difficult to evaluate the merits of the different approaches, since the conditions for how the tools are trained may vary to a degree that influences the overall results. For this reason, competitions are organized around so-called shared tasks. Here, different research groups try to solve exactly the same task starting with the same training material. A benefit of participating in such tasks, besides getting feedback on the feasibility of your approach, is access to hand-labelled data for training and evaluating machine learning-based approaches. We therefore participated in a shared task, the SemEval-2018 task 7 "Semantic Relation Extraction and Classification in Scientific Papers". Our approach, based on deep learning with convolutional neural networks, did well. We participated in three different subtasks and got the third best result in all the three subtasks. Given that there were up to 28 participating teams in the competition and some of the participating groups have a much longer experience with similar tasks, we are very satisfied with our results. They show that what we are doing is state-of-the-art. This work will be presented in New Orleans in June 2018.

The scientific papers in the SemEval task were from a different field than oil and gas. However the methods developed can be applied to other domains as well. The only prerequisite for using them for other domains is the availability of training data, i.e. documents manually annotated with entities and relations. Our plan for 2018 is to get annotated data for the oil and gas domain so that we can transfer and develop our successful methods further.

Scalable Computing

Geir Horn

Dr. Geir Horn is Head of European ICT Research at the Faculty of Mathematics and Natural Sciences. He represents the Faculty on the SIRIUS strategic board and leads the SIRIUS Scalable Computing strand.

Geir holds a Cand. Scient. degree in cybernetics and a PhD in computer science on mathematical learning in combinatorial optimization problems, both from the University of Oslo. He started his career at the Centre for Industrial Research in Oslo working on embedded software and field-bus sensor systems, leading him on to distributed and parallel computing. Geir's current research interest is on stochastic combinatorial optimization and how to handle complexity and services choreography for large-scale distributed applications through adaptation, autonomic decisions, self-awareness, and emergence.

Geir has previously held positions as senior scientist and research director at SINTEF in Oslo, before spending 4 years in more basic research at the SIMULA Research Laboratory. He has been working with European research for 20 years and has been coordinating 16 European collaborative projects ranging from coordination and support actions to large integrated projects. Geir has participated in multiple proposal evaluations for IST/ICT in FP5, FP6 and FP7 and has also been on the scientific review panel for several ICT projects within the areas of Cloud, communication technologies, and the Internet of Things.

The Scalable Computing (SC) research strand is about making data access faster. It aims to providing more efficient processing to SIRIUS projects using the full range of computing resources, from High Performance Computing (HPC) up to elastic deployment in the Cloud. Business value lies in the SIRIUS applications and scalable computing is the way we create more of this value in a shorter time.

Main achievements

This year we have set up the testbed infrastructure of the SIRIUS Laboratory. We then conducted initial evaluations with the software intended to be used for maintenance planning and scheduling in the Operations work package. In addition, the RDFox database developed by the University of Oxford in the Supporting Technologies work package has been ported to the HPC testbed and is ready for use with real data. Finally, work has started on reservoir simulation code, supporting the Exploration work package.

The High-Performance Computing test beds

SIRIUS researchers have access to a shared memory computer that was delivered by Numascale to the University of Oslo some years ago. Even though the system is some-how dated, it is fully operational and serves as a conceptual test bench for using HPC to improve application performance. SIRIUS gains three benefits from using this system. First, the machine is already installed and can be used immediately. Second, SIRIUS is the primary user of the machine. Finally, the machine reflects the architecture that can be realised with latest technology from Numascale. This final factor is important as shared memory systems are particularly attractive for large databases of structured data. Here performance improvements are expected due to being able to hold larger parts of the database in memory and not on disks. Numascale is working within SIRIUS on their next generation technology. This supports the recent Intel UltraPath Interconnect. Prototypes will be available for testing in 2019.

Dolphin Interconnect Solutions offers a different approach to high performance computing with their Peripheral Component Interconnect express (PCIe) solutions. These allow fast and transparent interconnection of any PCIe device and server. This opens some interesting opportunities for

SIRIUS in that it is possible to start with a small computing cluster and add more servers as needed. Furthermore, any device in the system can be connected as a "local" device on a server. Consider, for example, a server that has several graphical processing units (GPU) attached for fast stream processing of data. One or more of these GPUs can be "borrowed" by another server in the cluster if it needs additional stream processing power. In the same way any disk storage attached to the system can be mounted as a local disk on any server. This is particularly useful for handling unstructured data, like huge document collections. Under daily operation one may benefit from symmetric use of the disks in the system, with one disk attached to each server responsible for a sub-collection of the documents. However, for larger and more complex data-mining exercises, the server running the job may "borrow" all disks and access them as if they were local discs without the overhead of going via a remote server. SIMULA Research Laboratories are now setting up and testing a cluster of servers using Dolphin's technology, and it is expected that this cluster will be available to SIRIUS application testing during the first half of 2018. Its first application will be to test parts of a scalable Cloud infrastructure using the technologies from the MELODIC Horizon 2020 project. The MELODIC project is coordinated by the University of Oslo.

Applications

Software, developed in the Horizon 2020 project HyVar for solving parameter optimisation problems in large scale planning, was ported to the Numascale testbed by researchers from the University of Oslo. This software was then subjected to intensive testing in setups involving the use of 221 184 processes using up to 104 computing nodes. In this case it was found that the application structure gave a better speed-up by limiting the total number of cores used simultaneously. This could be due the operating system of the machine having to schedule many relatively small jobs over many computing nodes according to each node's load and availability. This effectively makes the operating system the bottleneck for performance improvements.The RDFox database was ported to the NUMA machine by workers at the University of Oxford. A cultural heritage data set, from the British Museum, was used as test data. This database had 18.8 million initial facts. These were materialized to 533 million facts after reasoning. Contrary to what was originally expected, it seems that shared memory was not essential for the speed-up for the processing of this data set. Locality of data seems to be crucial. We are working further on an approach that views the NUMA machine as a cluster. It would then make sense to partition the query graph with minimum cut-set partitions. The initial results therefore seem to indicate that the second test bed, the PCIe cluster at SIMULA, could be a better platform for this data set. It will be exciting to see if this hypothesis can be verified in this year, and if this conclusion remains the same when industrial data sets become available from exploration and operations projects.

Work Practices

High volumes of digital data can be used to improve organizational performance. However, if only a small subset of the data is used, the expected benefits of generating the data remain largely unreleased. This is the situation SIRIUS seeks to resolve, with novel and innovative technologies for scalable data access in the oil and gas industry. However, from experience with previous research projects over the past decade, we know that professional communities within oil and gas companies use available data only to the extent that it supports existing work practices or solves previously intractable practical problems.

Avoiding Answering the Wrong Questions

In 2017, we have therefore focused on supporting the other technical strands to prevent what we call Type III errors. A Type III error occurs when a researcher answers the wrong question using the right methods. This kind of error is a real and recurring problem in technologically oriented research projects. Oil and gas company employees are justifiably wary of well-built software solutions to what they perceive to be non-problems. Many of them have been exposed too often to novel software that has little or no practical use in their daily work. Our goal is to help the technically oriented strands in SIRIUS to not only identify researchable problems but to identify problems that are also important to the centre's industrial partners.

Supporting the Digital Geological Assistant

We have pursued this through several projects in 2017. Here we describe three of the projects. In close collaboration with Schlumberger and Statoil, we have used our approach to identify and develop the concept of geological history reasoning. Based on the two industrial partners' joint observation that the digital tool set available is not always a good match for exploration geologists, we conducted a series of interviews with exploration specialists to test, substantiate and possibly explain this observation. We found that existing tools predominantly focus on building structural understanding of the subsurface by reducing the noise and uncertainties in seismic images. Geologists, on the other hand, use data to build a processual understanding of the subsurface and how it has become this way through series of geological events. Based on the concept of geological history reasoning, our aim for 2018 is together with the Knowledge Representation and Execution Modelling and Analysistechnical strands to develop a demonstrator of how geologists' work could be better supported. See page 50 for more details.

Improving the Use of DISKOS

As a second example, the data available in DISKOS, Norway's national data repository, is an invaluable asset for the oil and gas industry. SIRIUS is looking at how to even better exploit the value of the data available here.

> **DISKOS is an invaluable asset for the oil and gas industry.**

While exploration specialists are the end users of DISKOS data, project data managers (PDMs) are often the interface between DISKOS and the data user. We therefore conducted a series of interviews with PDMs to find out the challenges of accessing DISKOS for those who do so regularly. PDMs solve the practical challenges of finding all relevant data. They are also critical in verifying and cleaning the data they serve geoscientists from DISKOS. PDMs have different and more nuanced needs than data consumers. Their role must not be underestimated in solving the scalable data access problem.

Radical Change in Managing Engineering Requirements

Finally, we have taken part in the centre's ongoing engagement with digital requirements handling. Managing the complexity of large-scale construction projects is critical for the cost and economic viability of the resulting installations. This cost can quickly spiral out of control. Requirements handling is a central mechanism for managing this complexity. There are ongoing industrial initiatives towards digitalising today's document-based requirements processes. Our role has been to highlight the large-scale sociotechnical challenges inherent in changing how an entire ecosystem of stakeholders produces, maintains, propagates and consumes requirements in the EPC industry. While our role here has been small this year, we will expand our engagements as we employed a new PhD student, Mina Hagshenas, to work on this.

> **Our goal** is to help the technically oriented strands in SIRIUS to not only identify researchable problems but to identify problems…

Thomas Østerlie

Thomas Østerlie leads the SIRIUS technical strand on work practices. He is a senior research scientist with a shared position between the Department of Computer Science, NTNU and Studio Apertura, NTNU Social Research. He holds a PhD in information systems from NTNU from 2009 and has 10 years of experience as software professional prior to that. He specializes in doing empirical research in close collaboration with industry actors. Østerlie has worked extensively with the oil and gas industry since 2009, and currently heads Studio Apertura's strategic initiative on innovation in organizations.

Data Science

Accessing data, the overarching theme of the SIRIUS centre, is of interest only if the data is ultimately used to create insight and guide decisions. In addition to data access, we need data analytics, comprising topics like data analysis, machine learning and visualization. SIRIUS has teamed up with the BigInsight SFI to form the DataScience@UiO cluster.
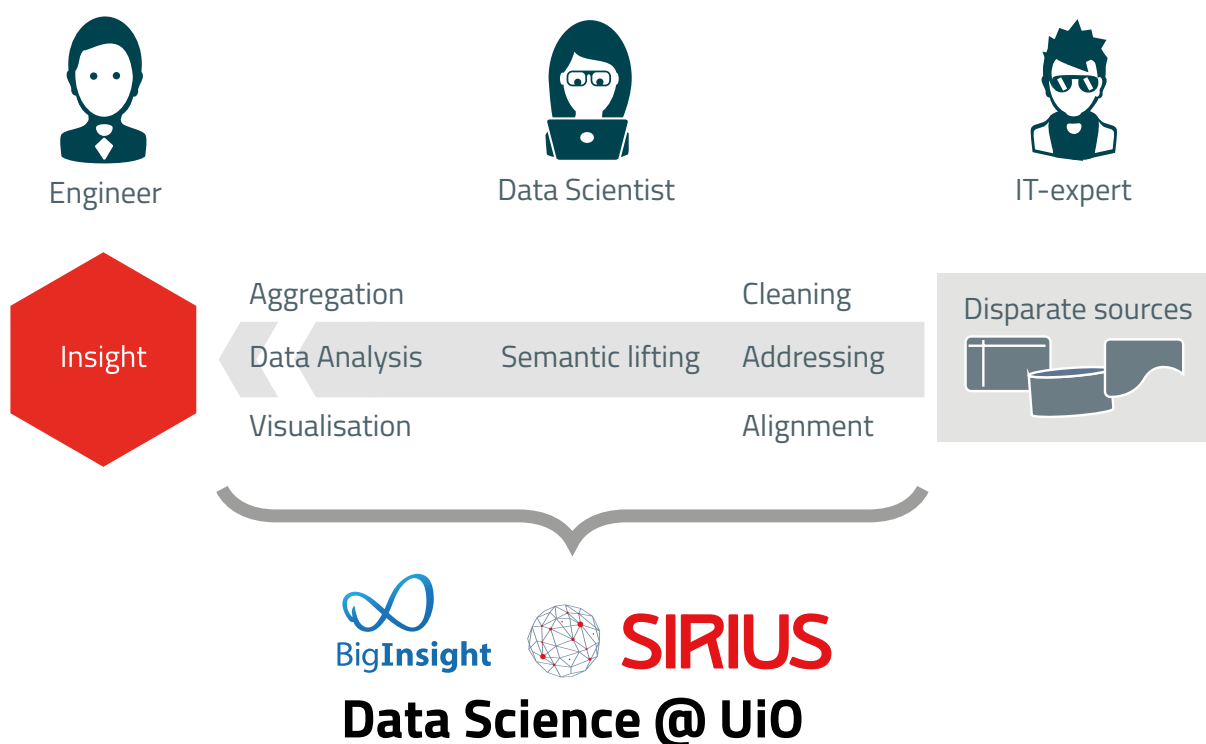
What we do

The Data Science research strand in SIRIUS is concerned mostly with the connections between data access and analytics. In a nutshell, research activities address one of the following questions:

- How can data access help analytics?
- How can data analytics help access?
- How can everything work together?

A particularly interesting aspect is that data access (the work in the Knowledge Representation strand) is very much concerned with structured, e.g. graph-shaped or tree-shaped, data, while current statistical methods in data analytics (e.g. machine learning) are very much based on vector-shaped data. Establishing a solid connection between these two types of data will open countless new possibilities for research and application.

Data Science 2024 – Our Ambitions

- An end-to-end understanding of "Data Science" that encompasses data access, reasoning and statistics for a variety of structured and semi-structured data types.

- To establish a framework and methods to describe phenomena by a combination of statistical and ontological methods.

- To develop a maintainable pipeline of tools to link data access and data analytics.
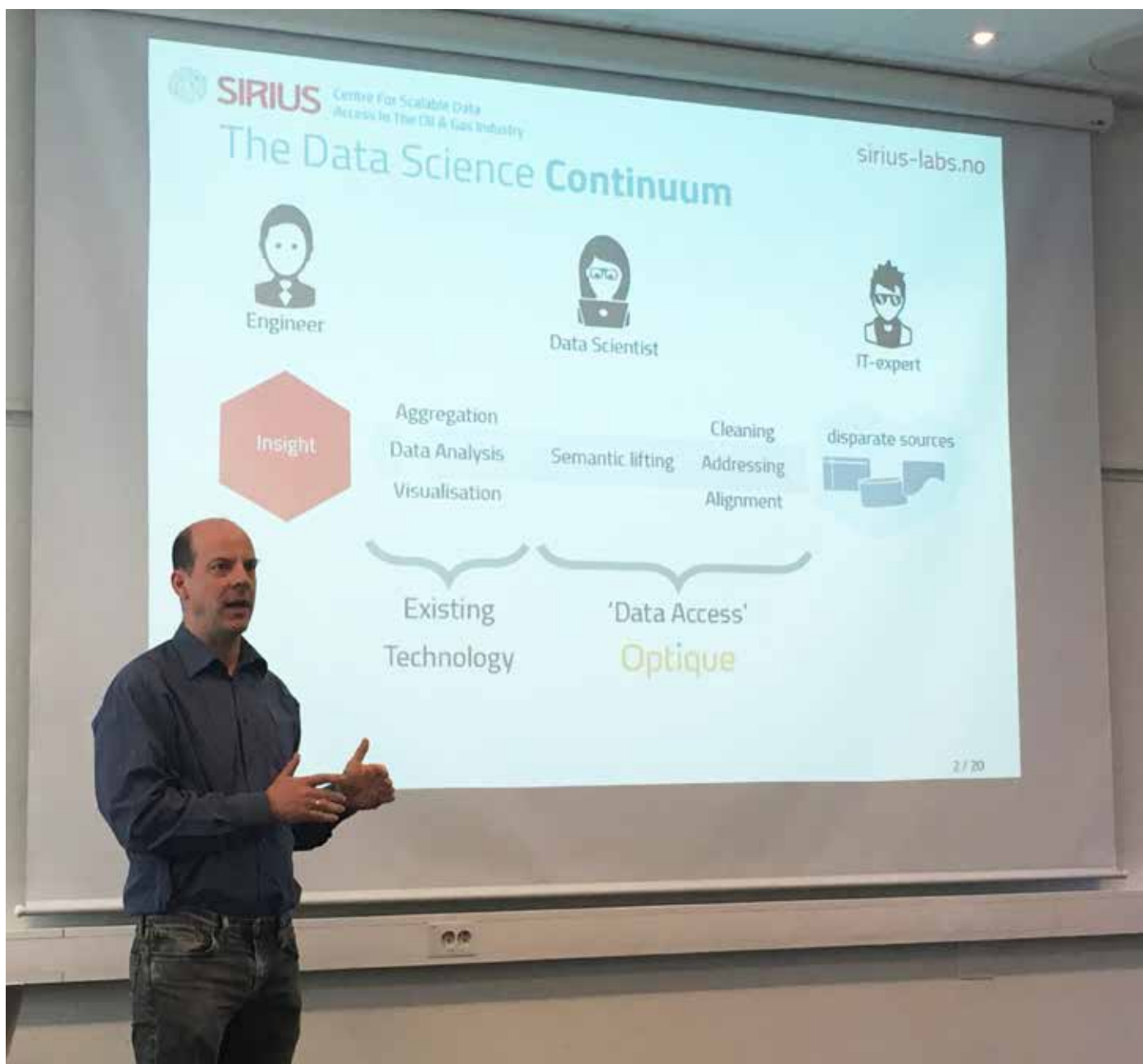
Martin Giese explains the data science continuum to the Spring General Assembly.

Establishing a solid connection between these two types of data of will open countless new possibilities for research and application.

Highlights in 2017

- Investigated statistical descriptions of tree-shaped information, tuned to the case of installation descriptions. The goal is to represent the relationship between the bill of materials (BOM), the actual installation built, and the information added into e.g. an SAP database, to pinpoint and repair mistakes in the database.

- Worked on data gathering for a project to predict reservoir parameters for early prospect evaluation.

- Defined a PhD project and enrolled a student (with external funding from NIVA) who will work on the engineering aspects of building a data access and analytics pipeline.

- Started on student projects that will use machine learning techniques for ontology engineering.

- Initiated an interaction with the BigInsight SFI to consider ranking of user options in our visual query system OptiqueVQS as a recommendation system.

Exploration

The Exploration work-package is responsible for defining and running a portfolio of experiment, prototype and pilot projects that support business processes in the sub-surface part of the oil and gas business: exploration, geosciences, reservoir modelling and wells.

Jens Grimsgaard

Jens Grimsgaard is manager of the Exploration work package in SIRIUS. He is a geophysicist based at Statoil's offices in Bergen. His master's degree is in Physics/Geophysics from NTNU (NTH) in 1983 and he has worked in the oil industry, in Norsk Hydro, Statoil Hydro and now Statoil, since 1984. He worked for the first 15 years with geo-applications, computer operations and database programming. He has then worked as project leader in research since 1999, applying new technologies within geosciences. His main research interests are visualization, virtual reality, natural language processing and machine learning.

Scalable data access can increase the speed of decisions in field assessments through giving end-users easy access to data from multiple databases. These databases have relational data models with thousands of tables and tens of thousands of columns. They are often in silos, with each discipline maintaining its own databases. Data is also often a mixture of structured data and unstructured text and reports. Seismic data, in particular pre-stack data, also has massive volume.

Our ambitions in the Exploration work package are to:
- Open up exploration data so that it is readily accessible to end-users.

- Manage the large variety of data representations used in the domain.

- Ensure that large-volume data, such as seismic data, can be managed, stored and processed efficiently.

- Break down silo-based work practices.

- Develop easier and cheaper methods for ensuring data quality.

- Focus on end-users' needs.

During 2017, a scoping activity with Schlumberger has identified a portfolio of project ideas. These, along with project ideas identified during the exploration workshop held in 2016, will be refined and worked up into project proposals through a common feasibility project called Feasibility Activities for Exploration Portfolio.

Another scoping activity was organized in September 2017 with KADME around Effective Use of Data from National Data Repositories. The main focus of this workshop was to explore ways to make national data repositories, DISKOS in particular, a more active tool in exploration. Project ideas developed in this workshop are now put

forward as a common feasibility project. These project ideas will be refined and worked up into fundamental and innovation projects in 2018. Furthermore, workshops will be organized in 2018 with partners to explore more R&D possibilities in utilizing national data repositories.

A project called "Data demand in Early-Exploration" had SIRIUS researchers working with Schlumberger and Statoil to define an innovation project. This activity resulted in the Geological Assistant project 2018. You can read more about this work on page 50.

A FORNY innovation grant and a faculty innovation grant were used to explore and define pilot applications of ontology-based data access (OBDA) with Statoil and DNV-GL. This work was to be done as verification for a software product for OBDA. The proposal for a full FORNY project was unsuccessful, by we are continuing work with DNV GL to improve the commercial maturity of our OBDA software.

Two doctoral projects have started in 2017 in the exploration work package. Andreas Thune is working at Simula and the University of Oslo on HPC support for reservoir simulation (page 52) and Summaya Mumtaz, (a DataScience@UiO fellow), is working on the application of data science processes to analogues.

The main focus of this workshop was to explore ways to make national data repositories, DISKOS in particular, a more active tool in exploration.

Finally, two project ideas identified at the at the November 2016 and September 2017 workshops have been chosen for specific feasibility and innovation work in 2018 because of their fit to partner interests and the SIRIUS technical profile. These are:

- Developing and implementing an Information Portal for exploration workers.

- Tracking and visualizing data provenance and quality during ontology-based data access.

SIRIUS has teamed up with Geosciences

Adnan Latif

Adnan Latif joined SIRIUS in August 2017 as Project Leader for the Exploration work package. He has a background in Geoscience and Business administration. He holds a master degree in Geophysics from the University of Oslo and a MBA degree in marketing and finance. Since 2007, he has worked in the oil and gas industry as a geoscientist and has a broad experience within the oil and gas exploration, technical sales and project management. His main roles in SIRIUS are to manage projects in the Exploration work package, industry relations, networking and research assistance.

Digital Geosciences

Digitalization is about information technology solving problems for people. This means that a successful innovation requires collaboration between the domain experts– the people who know about a business or a subject – and the IT specialists. SIRIUS researchers are IT specialists. They are rewarded for developing and programming generic methods, software systems and hardware. Driving innovation in oil and gas requires knowledge of specific disciplines, such as geosciences, reservoir engineering, planning and logistics or process automation engineering. How can SIRIUS gain access to this knowledge? Our industrial partners can and do help, but they have their day jobs. To solve this problem, SIRIUS has adopted a policy of working with academic domain experts in our chosen areas of innovation.

For the exploration work package, a source of deep domain knowledge lies a kilometre away from SIRIUS' offices in the University of Oslo's Department of Geosciences. They have many years of experience in using computers to analyse exploration data: They are academic partners in DISKOS, the National Data Repository. Dr. Michael Heeremans works at the Department as their Data Manager. In this role, he faces

challenges similar to the ones faced by data managers in petroleum companies. He uses DISKOS and other databases to prepare data sets that students can use in classes and research. He and his colleagues have domain knowledge that is needed to build a useful and usable system for ontology-based data access.

A collaboration between SIRIUS and Geosciences was first proposed in 2016. Peter Nielsen, then Chief Geologist in Statoil and an Oslo alumnus, suggested that the quickest way to further develop the Optique platform for accessing exploration data was to apply it to databases at UiO. This proposal was met with enthusiasm at the Department of Geosciences, and so in 2017 SIRIUS and Geosciences set up a collaboration. Using funding from the Statoil Academia program, we employed a researcher, Irina Pene, to work with a team of computer scientists (Dag Hovland, Ernesto-Jimenez Ruiz, Martin Skjæveland and Jens Otten) on implementing the Optique platform. Irina is a geoscientist, with long work experience in Repsol, a Spanish oil company, and oil service companies. In addition, we employed Adnan Latif, another geoscientist, to lead projects in the exploration work package.

Around the same time, a chance encounter with the head of the PPDM standards organization led to a broadening and internationalization of this collaboration. PPDM had been collaborating with Mara Abel, a professor at the Federal University of Rio Grande do Sul (UFRGS) in Porto Alegre. Professor Abel has a masters' degree in geosciences and a doctorate in computer science. She leads a research group in petroleum informatics, with a commercial offshoot and collaboration with Petrobras.

We followed up this contact with a video conference and agreed that our research programmes were complementary. Financing for the necessary travel to set up a collaboration was a challenge, so we prepared a program for educational collaboration and applied to two SiU UTFORSK calls – one of which was a joint call with the Brazilian organization CAPES. Both applications were successful and the academic exchanges will begin in 2018.

The first installation of Optique software was done at the Department of Geosciences in late 2017. This provides a foundation for an ambitious program of semantic modelling of the geoscience domain and demonstration of the benefits of scalable data access. We will also be integrating the results of this work into the teaching at the Department of Geosciences and developing a short course as part of the UTFORSK projects. The vision is that a student can, by themselves, find data for a prospect from a wide range of data sources, perhaps over more than one National Data Repository.

The availability of geoscience expertise and work on semantic modelling is helping other SIRIUS projects. There is cross-fertilization of people, ideas and models with the Geological Assistant project (see next article). Access to geoscientists has also helped PhD students Farhad Nooralazaheh and Summaya Mumtaz with their work on natural language processing and applications of data science methods to geological analogues. Finally, this work is a valuable test bench for our fundamental work on ontology templates (see page 60).

Towards a Digital Geological Assistant

Hallgrim Ludvigsen is a geoscientist who works as a Digital Innovation Champion in Schlumberger, a SIRIUS partner company. In 2017 he has worked with SIRIUS to scope out collaborative projects and set up a promising program of work related to digitalizing workflows for exploration geologists.

Schlumberger is in transition. The oil price shock of 2013 coincided with the rise of digitalization. Schlumberger chose to meet the market challenge with an accelerated move towards a new digital platform. This plan was announced at an Investor Conference in May 2014 and achieved a milestone with the launch of the DELFI cognitive E&P environment in September 2017. As part of this initiative, certain staff in Schlumberger have been given responsibility to see how further digitalization can support Schlumberger's

business and provide new business opportunities. Hallgrim has one of these champion positions.

Hallgrim first became engaged with SIRIUS when he was visiting Schlumberger's Software Technology Innovation Centre in California. The development manager there discussed a message from SIRIUS' contact person (who, amusingly, sat in the next office to Hallgrim in Stavanger) on how Schlumberger could engage with SIRIUS. On the flight

home, Hallgrim wrote down a bullet list of things that the company needed.

He was inspired to work with SIRIUS by two possibilities. The first of these is the long-term view that SIRIUS represents. An eight-year project is almost absurdly long in a business driven by quarterly results. Or is it? Schlumberger's business is built on knowledge and software with many decades of development behind them. If we look at, for example, deep learning in neural nets, we see methods that have been worked on in academia for decades but only recently matured enough to be widely used in applications enabled by advances in elastic cloud computing and big data sets. Perhaps there are similar technologies lurking in academia that we can find and exploit in a similar way?

The second possibility is related to DELFI and Schlumberger's move to service-based delivery of software. Previously, most of Schlumberger's technical software was installed on a computer that was owned and run by the customer. This made it difficult for Schlumberger to innovate and experiment with the software and how it is used. A move to software-as-a-service means that the software is run on computers under Schlumberger's control. This means that, using DevOps, Schlumberger can try out and prototype new ideas at a much greater pace, with better access for systems like those that we are developing in SIRIUS.

On returning to Norway, Hallgrim found that his "shopping list", written on the plane, needed to be modified if it was to be used with SIRIUS. Together with Per Eivind Solum, the SIRIUS main contact, he concluded that a successful project in a consortium like SIRIUS cannot simply meet the short-term needs of one partner but must arise from perceived benefits or incentives in Schlumberger, the academic partners and the other industrial partners in the centre.

This required a lot of talking with academics, Schlumberger colleagues and people in Statoil, but it was worthwhile as it produced a portfolio of project ideas that have clear benefits for all parties in the project.

The first project to come out of this process is called The Digital Geological Assistant. This is a collaboration between Schlumberger, Statoil, NTNU and the University of Oslo. It involves academic staff from the Work Practices, Knowledge Representation and Execution Modelling & Analysis strands. Staff from the Department of Geosciences at the University of Oslo are also involved. Thomas Østerlie and his colleagues from NTNU interviewed geologists in Schlumberger and Statoil and found out that where it was obvious how work flows could be digitalized, they had been digitalized. This applied to most work flows where sensor data and measurements were used for simulation or visualization. For example, it is relatively straightforward to apply a neural net to pick out a horizon on a seismic image. On the other hand, there were other work flows where there was little or no support by digital tools. These "black holes" cover areas where knowledge is discrete, experience-based and intuitive. Pictures and analogies are used. This lack of numerical precision makes computer support difficult. The understanding of the geological history of a prospect is vital to the assessment of a prospect but is not supported well by computer tools.

A working hypothesis was made that a tool for supporting geologists would need to use resources and ideas from knowledge representation and formal methods, and this became the technical vision for the project. However, a successful project needs early prototyping to demonstrate the idea and see whether it addresses geologists' problems. For this reason, a small part of the problem, namely the ordering and timing of geological events, was presented to team of eight computer-science students at NTNU as a project challenge. These students worked on this project as part of their normal course work and used gaming technology to produce a working prototype. This prototype was presented to Schlumberger geologists and was well received.

Hallgrim's ambition with the ongoing project is to develop a geological reasoning engine. This engine will support a deep dive into formally describing geological history for a prospect so that it can assist an exploration geologist with automated reasoning. This will combine massive computing power with formal methods of analysis and a logical understanding of how things and ideas in geology interact. Exploration geologists are highly skilled but are currently limited by their reasoning capacities. Geological analysis is a complex mental process. An automated tool could be a valuable supplement and support in generating better and more accurate geological assessments of prospects.
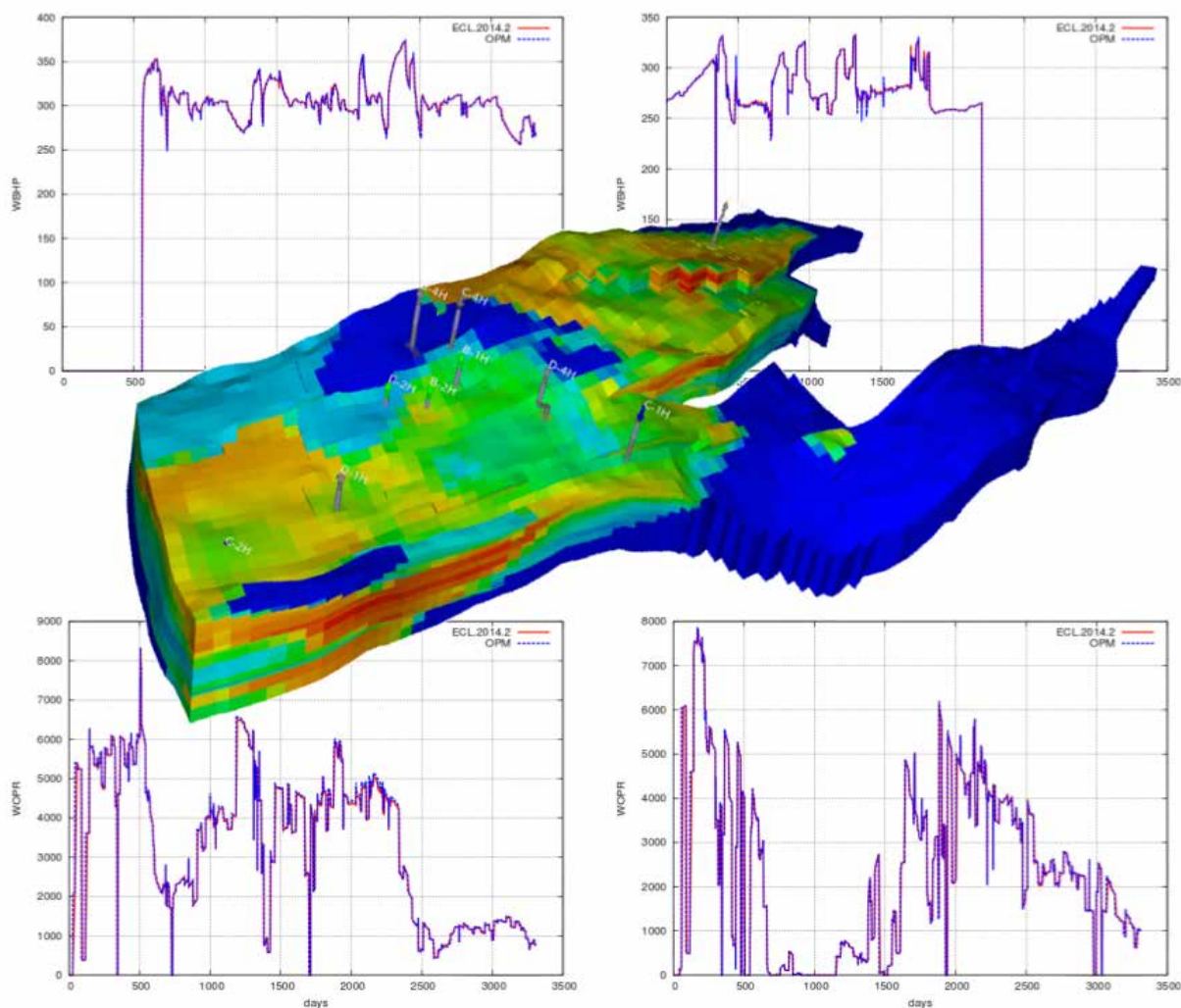
High-Performance Computing for Reservoir Simulation

Vastly heterogeneous geological features are known to exist in subsurface reservoirs. The multi-phase flow in these reservoirs can thus have very complex patterns, for which a mathematical description involves a system of nonlinear partial differential equations. High resolution is needed in numerical reservoir simulations to resolve the complex flow phenomena, leading to large amounts of computation that can only be properly handled by parallel computing platforms. Moreover, various types of uncertainty in the geological properties and the need for considering different scenarios will require detailed reservoir simulations to be run repeatedly. The importance of high-performance computing (HPC) is therefore self-evident.

There has traditionally been a culture of lack of openness in the domain of reservoir simulation, reflected in proprietary

Andreas Thune

Xing Cai

Being a part of the Scalable Computing strand, this PhD project will provide important experience on programming and utilizing various modern platforms of parallel computing...

reservoir software and data. This prevailing culture has also hindered a swift development of HPC techniques for this domain. The OPM initiative (https://opm-project.org), however, represents a new community effort in promoting openness in reservoir simulation research. Apart from making a number of real-world data sets openly available, OPM also coordinates collaborative development of open-source software for simulating porous media processes. An extensive software framework has already been developed in OPM.

However, a closer look at the parallel performance of the OPM reservoir simulator shows that its current implementation is not yet capable of achieving the full potential of modern parallel computing platforms. This inefficiency is due to both imperfect components in the numerical scheme and mismatches between parts of the implementation and modern hardware architectures. A newly started SIRIUS PhD project (student: Andreas Thune, supervisors: X. Cai, A.-B. Rustad, T. Skeie, E. G. Gran) thus aims to greatly

enhance the HPC capability of OPM's reservoir simulator. During the second half of 2017, extensive numerical experiments were carried out to identify the numerical and software bottlenecks. Inefficiency in the mesh partitioning and some of the numerical components of the parallel simulator has been pinpointed.

The future work on the PhD project, in close collaboration with domain experts from Statoil, will address the following aspects:

- Improved, hardware-matching domain decomposition strategies.

- Efficiency-oriented adaptation of numerical techniques in the parallel simulator.

- A software overhaul of OPM's parallel implementation for better utilizing modern computing hardware.

Being a part of the Scalable Computing strand, this PhD project will provide important experience on programming and utilizing various modern platforms of parallel computing, including systems delivered by the HPC industrial partners of SIRIUS. Such experience is valuable for many of the other activities in SIRIUS. Moreover, the improved HPC capability of reservoir simulation will speed up the turn-around time and improve uncertainty quantification associated with the workflows of planning and operations in the oil & gas industry.

Elena Parmiggiani

Bringing Context into SIRIUS

Elena Parmiggiani is a postdoctoral researcher in the Department of Computer Science at NTNU in Trondheim and holds a PhD in Information Systems (IS) from NTNU. She came to Norway almost by accident. When she was growing up in the town of Spilamberto, near Modena in Italy, she wanted to experience living in the Nordic countries. This chance came during her Master's studies at the University of Modena and Reggio Emilia. Through a connection of her Italian supervisor at KTH in Stockholm, she found a relevant project to participate in for an exchange period. However, it turned out that the project was led by Professor Monica Divitini at NTNU in Trondheim. This brought her to enjoy the winter in Trondheim (down to -27 degrees that year) in 2010 and a project on building a context-based recommender system for new students to learn about a city like Trondheim.

Elena's primary qualifications are in computer science and the project that brought her to Trondheim used ontologies and semantic technologies. However, she became interested in the user factors and context behind the systems she was working with. Having finished this exchange, she moved back to Italy, but only for a short while. In February 2012 she was employed in Professor Eric Monteiro's group at NTNU as a PhD student in the Digital Oil project. The

Elenas primary interest is the study of the sociotechnical aspects involved in technology development, implementation and use.

studies of technology in practice, which inherit their research methodologies from the social sciences. There are differences in method and approach and it is not straightforward to bring them together. Fortunately, there is a well-established tradition in doing so in Science and Technology Studies. Both of Elena's external examiners for her PhD worked in this field. One of them, Helena Karasti from the University of Oulu, has become a collaborator. Elena will be spending six months in 2018 working with Professor Karasti in addition to contributing to SIRIUS projects.

In 2017, Elena worked with Eric Monteiro to develop and pilot a course in Digital Service Innovation at NTNU. She is now responsible for teaching the course, which has 40 students this year. The course covers the methods used to digitally transform a business and explores the role of platforms (such as the Apple Store or Schlumberger's Ocean) in business. She is experimenting with digital service delivery in her teaching. The course uses different teaching approaches, such as frontal lectures, student presentations, interactive quizzes like Kahoot and a new interactive tool called Beekast.

Elena's role in SIRIUS has been spread over both the exploration and operations work packages. She has contributed to defining the Geological Assistant project, described on page 50. Another project in the exploration work package is with the consortium partner KADME: This work looks at how the data flow and the work flow can be better integrated in an oil company. The main idea in this work is to focus data provenance, or pedigree: Where did this data come from? Is it reliable? Do I trust the source?

The mentoring program (see page 78) has been valuable and enjoyable. Elena's mentor was Derek Lacey, a senior manager in IBM Norway. She feels that the program has helped her to mature as a researcher, taught her how to handle stress and prepared her to be a leader at a later stage in her career. Sessions in the program also raised important issues around work-life balance, communicating across disciplines and handling conflict. In addition, the program resulted in the definition of a collaboration project between IBM, the University of Oslo and NTNU around the design and delivery of complex systems for cloud computing.

In her view, the mentor program is an unprecedented arena for industry-academic collaboration. Perhaps the next version of the program can include training in how researchers can take account of the societal requirements of national and European funding bodies. These organizations require impact from their funding. The definition of impact varies according to local needs and circumstances. It would be instructive to explore this aspect of defining research and career development.

condition on the contract was that she learned Norwegian. This happened through a mixture of evening classes and deep immersion through field work in an oil company to study an environmental monitoring project. This resulted in Elena speaking an interesting domain-adapted Norwegian: full of oil and gas technical terms. By 2014 she could conduct field interviews in Norwegian.

Her primary interest is the study of the sociotechnical aspects involved in technology development, implementation and use. Her primary field is Information Systems, with influences from Computer-Supported Collaborative Work and Science and Technology Studies. This required her to learn a new approach to method, theory and language as she moved from pure computer science to sociotechnical

Our ambitions
are long term...

Operations

The operations work package covers the parts of the oil and gas supply chain that do not involve sub-surface data and exploration. It concentrates on facilities: the data about them and the data produced by them. We take a lifecycle view of the facility and are interested in data from the first conceptual studies through engineering, procurement and construction, commissioning, operations, maintenance and decommissioning.

2017 was spent working with the SIRIUS partners to narrow down the focus of this work package. Digitalization of oil and gas operations is an area where many actors are working hard. Vendors like DNV GL, GE, Kongsberg Digital and Siemens are offering oil companies competing and complementary platforms for managing and analyzing data. The Aker group, through Cognite, is integrating operational data for AkerBP and Aker Solutions. Statoil established a Digital Centre of Excellence and its own internal platform for data analytics for operational improvement. SIRIUS must define a role for research to support all this feverish activity. Our ambitions are long term: we are seeking to enable a vision of computer-enabled operations that applies in 2024. Our plan is to work with and observe the current efforts in operational digitalization to find "pain points" that need to be addressed by our research-based innovation.

2017 was a year for scoping and feasibility projects, in which we built up understanding with partners and developed projects for prototyping and piloting from 2018. There were projects in the 2017 work plan that were discontinued after a short time because it became clear that they didn't match our partners' needs or our research profile. It soon became clear that there were three areas of operations where SIRIUS research could contribute:

- Applying formal methods of analysis and simulation to planning problems in commissioning, supply-chain logistics and maintenance. This work is done in collaboration with Statoil.

- Finding better ways of building and delivering systems that use operational data for decision support. One of the buzzwords of 2017 was digital twins. SIRIUS partners OSIsoft, Computas and DNV GL are working with the University of Oslo and University of Oxford on methodological projects related to accessing data efficiently and handing time-series data effectively for machine learning.

- Implementing and standardizing better ways of managing requirements throughout the lifecycle of the facility. This process is called a digital thread in the aerospace industry. This work has been described by Arild Waaler on page 8.

Each of these areas has been proposed as a SIRIUS beacon project. This means that our ambition is, by the end of 2019, to have begun substantial, externally financed projects that pilot the application of SIRIUS' technologies in each of these areas.

Planning:
Formal Models of Vessel Movement

SIRIUS' planning project has started by developing a case study that looks at vessel movements and cargo transport in the North Sea. The goal is to improve the workflow of planners at Statoil by providing a better overview of the bottlenecks that could delay overall progress, the load on different vessels, and the quality of their logistics operations. We hope to improve both the utilization of vessel capacity and on the timely delivery of material. In the initial phase of the project, we have used Real-time ABS as a modelling language to simulate and visualize the actual logistics operations. Compared to the tools currently used, Real-time ABS simulations provide a different level of overview which helps a user to gain precision in the decision-making phase.

This initial case study already illustrates the general usefulness of Real-time ABS modelling, beyond the realm of computing systems. Thus far, we have only used Real-time ABS for simulations in this case study. The case study has also driven the development of new input and output facilities for the Real-time ABS simulator, to better facilitate the interaction between model simulation and real data about Statoil's vessels and cargo transport. These new tools allow customizable visualization of output from the ABS simulator.

Figure 1: Raw operational vessel data

We are working with industrial data from different parts of a complex supply chain and integrating these into a uniform ABS model. The data covers transport plans for a large number of vessels moving between a supply base and installations, with logs for bulk and cargo delivery covering a twelve-month period, as illustrated in Figure 1. In this case, ABS is used to define a general framework for modelling transport plans by means of for vessels, containers, containers, bulk cargo, route segments, and delivery deadlines.

The model is populated by specific data, representing a concrete plan. This is currently done by moving the data from Excel into a SQL database, then generating ABS data structures that correspond to the industrial data set. Thus, the industrial data set acts as the driver for the simulation of the ABS model. The modeller specifies a time window, and data for this time interval is automatically extracted from the SQL database, converted into ABS data structures and fed into the model as the concrete plan is simulated. This allows the ABS model of the concrete plan for the given time window to be simulated. The planner is presented with a graphical view of the simulated plan, as depicted in Figure 2. This graphical view is dynamically generated in-browser from JSON data fetched as output from the simulation. The display can be easily adapted by a frontend developer. No knowledge of ABS is needed to create different views of the simulation data.

A practical challenge with this case study, in addition to the data cleaning required to convert operational data to fit with the modelling framework in Real-time ABS and the interaction of the simulation backend and the SQL database, was the conversion of calendar data to model time. Real-time ABS represents time using rational numbers. We calibrated the model with time 0 representing midnight on the first day of the simulation. Subsequent dates were numbered 1, 2, …, with the fractional part representing the time of day. This approach gave us sufficient resolution to model real time using abstract time units.

We intend to combine these simulations with stronger

Figure 2: Visualization of time series data depicting vessel movements
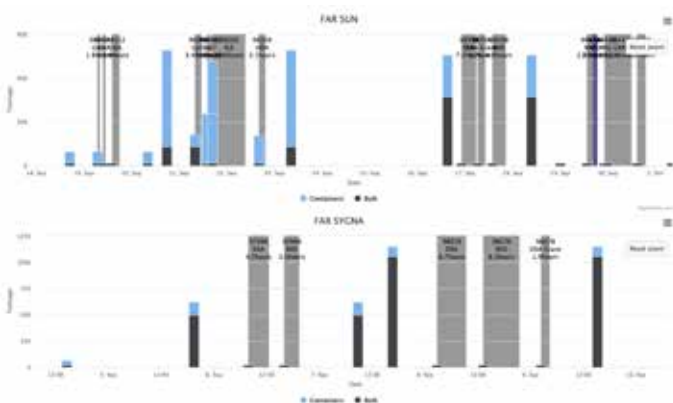
analyses to generate solutions and verify their correctness with respect to requirements such as resource restrictions, safety regulations and space limitations. This long-term perspective is illustrated in Figure 3, which shows the current work (in black) and some possible directions (in red) that can be explored to extend the current case study; these future directions provide concrete scientific challenges as well as added value in the planning processes.

Figure 3: Possible extensions of the current work
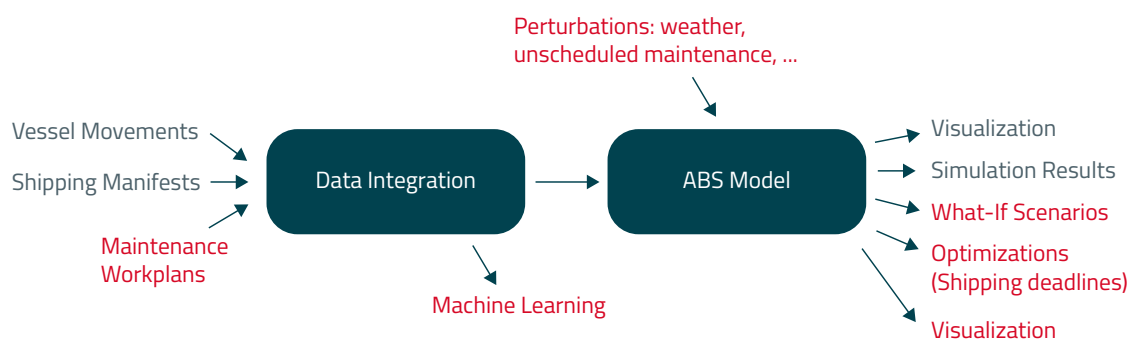
Einar Broch Johnsen

Einar Broch Johnsen is a professor at the Department of informatics of the University of Oslo and the Deputy Director of SIRIUS. His main research interests address the design and analysis of parallel and distributed systems, including resource-restricted and context-dependent behavior. He was the coordinator of the EU FP7 project ENVISAGE about the design of resource-aware services for deployment on the cloud and the analysis of service-level agreements for such systems. He is the scientific coordinator of the EU H2020 project HyVar, which combines sensor data, scalable cloud infrastructure and software product lines to enable context-dependent software evolution of embedded devices on cars. He leads the Frinatek project CUMULUS on formal foundations of cloud computing, funded by the Research Council of Norway, and he has previously been site leader for three EU projects (UpScale, HATS, and Credo) and led the Norwegian project CONNECT. Johnsen is or has been the supervisor of 14 PhD students and has published more than 100 internationally peer-reviewed research papers.

Sustainable Modelling of Complex Ontologies

Building ontologies is complex and laborious. The languages used to represent ontologies are expressive but are correspondingly complex. They require knowledge of mathematical logic and are often very verbose. A description of even a simple physical concept or object can require many repetitive formulas. This is error-prone and leads to models that are difficult to understand and maintain. This again is a barrier to adoption of knowledge representation approaches in industrial applications. It is thus a barrier to the successful piloting of the SIRIUS beacon projects.

SIRIUS is developing a solution to this problem: Reasonable Ontology Templates (OTTR). This is a language for representing reusable patterns for semantic modelling. These patterns can then be used to simplify the creation, maintenance and use of ontologies. The ontology templates work is being done in collaboration with the SIRIUS partner DNV GL and other companies (Aibel, Acando and Siemens). Successful prototyping experiments have been performed in commercial projects with encouraging results.

As noted above, building sustainable, high-quality, large-scale ontologies is hard. Part of the problem is the lack of established tool-supported best-practices for ontology construction and maintenance. We will illustrate this by a simple example, dear to the heart of any computer scientist: pizza. An ontology is built in three distinct steps:

1. Understanding the target domain for the ontology. In this case the domain of pizza.
2. Identifying the abstractions that apply in this domain. For example, "Margherita is a particular Italian pizza with only mozzarella and tomato".
3. Formulating the abstractions in a logical language such as description logics.

$$\text{Margherita} \sqsubseteq \text{Pizza} \sqcap \exists \, \text{hasCountryOfOrigin}.\{\text{Italy}\} \qquad (1)$$
$$\text{Margherita} \sqsubseteq \exists \, \text{hasTopping}.\text{Mozzarella} \sqcap \exists \, \text{hasTopping}.\text{Tomato} \qquad (2)$$
$$\text{Margherita} \sqsubseteq \forall \, \text{hasTopping}.(\text{Mozzarella} \sqcup \text{Tomato}) \qquad (3)$$

These formulas state that (1) Margherita is a pizza from Italy, (2) it must have mozzarella and tomato topping, and (3) it can only have mozzarella and tomato topping.

The ontology templates project targets the third task and particularly the large gap that exists between how domain knowledge facts are naturally expressed and how the same information must be recorded in the Web Ontology Language (OWL), the de facto standard ontology language.

The cause of the gap is the fact that OWL at its core supports only very basic model building blocks like (unary) classes and (binary) properties and offers no real mechanism for user-defined abstractions with which reoccurring modelling patterns can be captured, encapsulated and instantiated. It is possible in OWL to combine classes and properties to form arbitrary complex statements, such as those shown above, but there are no means to effectively record the complex statement as a pattern. The effect of this is that every domain statement must be broken down into OWL's small building blocks and no longer remains a coherent unit. This makes the resulting ontology harder to comprehend for the user and more difficult to maintain for the ontology engineer.

With OTTR templates the modelling pattern can be recorded similar as to a macro: the modelling pattern is represented as a parameterized ontology which may be instantiated by providing arguments that match its parameters. The result is a replica of the modelling pattern where parameters are replaced by the arguments. The following shows an OTTR template of the pizza pattern used above. The first line contains the name of the template and its list of parameters, each with a type and cardinality. The next rows contain the body of the pattern, which comprises other templates that represent basic OWL expressions.

$$\text{Pizza}(?\text{Name} : 1 \text{ class}, ?\text{Country} : ? \text{ individual}, ?\text{Toppings} : + \text{ class})$$
$$\quad :: \; \text{SubClassOf}(?\text{Name}, \text{Pizza})$$
$$\qquad \text{HasValue}(?\text{Name}, \text{hasCountryOfOrigin}, ?\text{Country})$$
$$\qquad \text{SomeValuesFrom} \mid \times (?\text{Name}, \text{hasTopping}, ?\text{Toppings})$$
$$\qquad \text{AllValuesFrom}(?\text{Name}, \text{hasTopping}, b_1)$$
$$\qquad \text{UnionOf}(b_1, ?\text{Toppings}) \; .$$

Instances of the template are created by referring to the Pizza template with matching arguments:
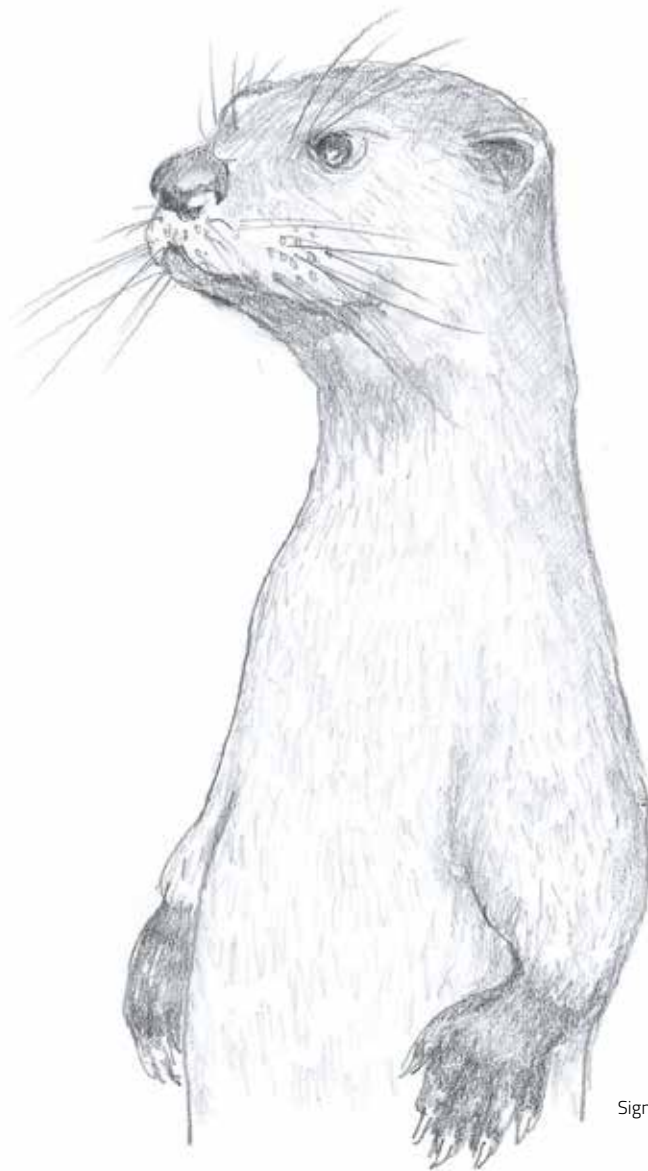
Signe H. Skjæveland

$$\textsc{Pizza}\big(\text{Margherita, Italy}, \langle \text{Tomato, Mozzarella}\rangle\big)$$
$$\textsc{Pizza}\big(\text{Grandiosa, Norway}, \langle \text{Tomato, Jarlsberg, Ham, SweetPepper}\rangle\big)$$

The benefit of building and maintaining ontologies in this way is that the design of the ontology, which may now be captured completely by templates, is clearly separated from the bulk content of the ontology, represented by template instances. The ontology expert can, using the tools of her choice, create and compose the templates necessary for

The languages used to represent ontologies are expressive but are correspondingly complex. They require knowledge of mathematical logic and are often very verbose.

representing domain statements that make sense for the domain expert. The domain expert in turn formulates instances of these user-facing templates using familiar desktop tools like spreadsheets, without worrying about the intricacies of formal logic. However, the fact that these templates are inherently logical can be exploited by reasoning techniques to discover redundancies, overlap, and other relationships between templates. These tests are crucial for efficiently using and maintaining a large set of templates.

Special care has been taken to ensure that the OTTR technology is easy to use. Formats for presenting templates and template instances to fit different purposes are available, including an OWL vocabulary. A public library of core and example OTTR templates has been published online, together with a web application and a desktop application for displaying and interpreting templates. All of this may be found at the project website **http://ottr.xyz.**

Digital Twins

A digital twin is a digital replica of an object, facility or system. It aims to give a description of its corresponding reality at a specific time. This means that the digital replica must be updated to reflect the current state of the referent: this requires alignment with data obtained from the object and the updating of configuration and design information as changes are made to the real object.

Digital twin was one of the buzzwords of 2017. Its premise is that digitalization of industry can be promoted by using models of products and processes with sensor data to improve operational decisions and enhance products. SIRIUS agrees. This trend is a mainstreaming and maturing of decades of work in simulation, model-based control and computer-aided design. Recently, vendors have been bringing to market digital twins that build on their specific expertise and products. Thus, simulation vendors refine their on-line simulators. CAD suppliers build virtual reality based on design models. Product and system vendors bring their design and analysis models on line. Control and database vendors offer integration platforms for twins, as do database and ERP. A facility owner faces the challenge of building a useful and sustainable digital twin from a multitude of competing, partial and inconsistent solutions. How can we increase the speed of implementation of these systems? How can we reduce system cost? How can we achieve interoperability between systems so that they support interdisciplinary decision making? How can we avoid overwhelming users with unnecessary complexity?

SIRIUS believes that a research agenda is needed to support and ensure the success of digital twins. We believe that our interdisciplinary program in scalable data access is a core part of this agenda. A supporting infrastructure that brings together knowledge representation, language technologies, formal methods, data science and high-performance computing can allow open interoperability between systems and task-oriented access to information.

The operations use case in the Optique EU project provided a better data integration system for a digital twin of turbomachinery. Support engineers used process data and simulations to monitor and troubleshoot turbines. Ontology-based data access allowed simpler use and access to the digital twins for each turbine by removing the need to know specific details about the database structure for each turbine. In 2016 and 2017 we have sought to apply this approach in proposals to the European Union. Unfortunately, neither of these proposals were successful.

In 2018 we are addressing this problem in two ways. Firstly, we have set up two prototyping projects that address fundamental issues that hinder the construction of digital twins. One of these projects is working on improving the process of building asset models for accessing data and interfacing a digital twin to operational data. The second project examines the role of time series data in digital twins. These projects are a collaboration between the University of Oslo, the University of Oxford, DNV GL, Computas and OSISoft.

These two projects support a strategic initiative, where SIRIUS brings together end users, integrators and suppliers of digital twins to define a research agenda for digital twins in the period up to 2024. This agenda fits well with the European Union's ambitions to demonstrate pilot applications of cognitive process operations. We plan to use this fit to position SIRIUS as a partner in a successful application for funding in 2019.

Cross-Domain Applications

The Cross-Domain Applications work package in SIRIUS contains projects that transfer technology between the oil and gas sector and other important economic sectors. This transfer is bidirectional. Scalable data access is a challenge in all complex organizations, be it in health, public administration or energy. Methods and skills developed in one domain are readily applied in other areas.

Personalized Health Care
BigMed

The BigMed personalized medicine lighthouse project started in 2017. SIRIUS researchers are building an IT framework that supports clinical decision-making in rare monogenic diseases and heart disease. We have completed relevant tasks towards the development of IT tools that are summarized in the following paragraphs.

We have worked in 2017 with understanding of the needs of doctors and lab technicians. This was done through structured meetings to find out what information is available for each clinical/laboratory problem. These meetings provided feedback and insight into information needs and decision-making processes in order to develop IT support for clinicians.

One of the main goals in BigMed is to improve the capture of clinical phenotypes and improve communication between healthcare personnel and laboratory technicians. A phenotype is a description of a patient's observable characteristics or traits. We examined the use of the PhenoTips, a tool that uses the Human Phenotype Ontology (HPO) to collect and manage phenotype data. We then worked on building knowledge in the use of specific medical ontologies: HPO for phenotyping, OMIM and Gene Ontology for genetics and the Disease Ontology for diseases. We examined how these could be aligned. Ernesto Jimenez-Ruiz compared different algorithms for aligning these ontologies. We also started work on a simple ontology for the information, including

> **SIRIUS** researchers have developed a visual interface to linked data to make it easy for non-technical users to search and browse datasets.

phenotype, that is used in the genetic requisition cycle workflow.

The DIPS Electronic Health Records (EHR) system, used in Oslo University Hospital, makes use of Open EHR archetypes. A successful personalized health system will require archetypes and ontologies to work together. Laura Slaughter reviewed the large amount of existing work in this area as a foundation for a possible IT solution in BigMed.

Work was also done on tools and infrastructure. RDF Surveyor, described in more detail below, is a generic tool for browsing triple stores. It was demonstrated at the BigMed plenary meeting June 2017). A specific tool has also been developed for entering HPO terms. This is simpler and easier to use than the PhenoTips system. A public server has also been set up to allow testing of semantic web and linked data applications.

The BigMed project issued a position paper entitled "Big data management for the precise treatment of three patient groups" in January 2018. This summarizes the findings of the first year of the project and presents a summary of the obstacles to personalized medicine in Norway. Barriers are financial, legal, technological and organizational. Success will depend on (a) hospitals being rewarded for innovation, (b) patient consent and privacy being managed in a way that allows access to big data analysis, (c) hospitals and health care organisations sharing data and (d) the implementation of a flexible, open and modern ICT infrastructure.

Laura Slaughter

Laura Slaughter is Work Package Manager for Cross-Domain Applications. Her PhD was received in 2002 in Information Science from the University of Maryland. This was followed by postdoctoral training at the Department of Biomedical Informatics, Columbia University, New York. Her main focus is in the area of healthcare and biomedical informatics. She has worked on a wide variety of healthcare projects covering various types of healthcare systems, including patient-oriented personal health records (PHRs) and clinical decision support. In general, she is interested in healthcare standards, biomedical ontologies and terminologies and use of semantic web technologies within the health area. Her current research centres on querying and exploring large integrated healthcare datasets, and the reuse, development and evaluation of biomedical ontologies.

What SIRIUS offers for personalized health-care research

SIRIUS's skills and tools are useful in healthcare where data is spread over different legacy systems, and various work processes need access to this data.

SIRIUS' research strands provide skills and tools that support needs in clinical care and precision medicine. Here we review areas where we can contribute, strand by strand.

The Knowledge Representation strand offers ways for clinicians and clinical researchers to find and use data more effectively. Semantic Web technologies are designed to allow open access to databases and datasets. To be accessible, this data must be made available as linked open data. Fortunately, there is much useful linked open data available in healthcare, such as data on drugs, chemical structure related to drugs and genetic disease datasets (OMIM). A catalogue of these datasets has been compiled by the Bio2RDF project. Linked data is presented as a in Resource Description Format (RDF), which is a standard defined by the W3C consortium.

SIRIUS researchers have developed a visual interface to linked data to make it easy for non-technical users to search and browse datasets. This tool allows a medical user to get an overview of the data, analyse the contents of the database and navigate between classes in the data set. Such a tool should be generic and not require any installation on an end-user's computer. It should also work even with large datasets, containing billions of data values. To the best of our knowledge, there is no tool that meets all these requirements. For example, LodView4 is a semantic browser that presents RDF data as tables and allows clicking on links. However, it does not give any overview of the dataset and navigation between classes is not supported.

SIRIUS researchers have created several visual query tools, like PepeSearch, OptiqueVQS and SemFacet, to support faceted search over RDF datasets. These tools can be used to explore this data in order to get a clear picture of its size and content. PepeSearch and OptiqueVQS present a flat list of the classes in the repository as a starting point, while SemFacet asks for a set of keywords as input in order to build a faceted interface to the data.

Previously developed query tools required installation of software on the user's computer and were not able to cope with large datasets, such as DBpedia. To address these problems, as part of the work related to needs of healthcare practitioners, we designed a web-based tool called RDF Surveyor. This takes a complex, large RDF data source and generates a web-based, navigable overview of each entry in the data source, with all its attributes and links. A demo of this tool is found at **http://tools.sirius-labs.no/rdfsurveyor/.**

Practitioners then need a tool for linking these general tools to real systems with patient data. This can be done effectively using Ontology-based Data Access (OBDA), as implemented in the Optique EU project. Using OBDA means that it is easy to implement reasoning using the data at the time of data access. It is therefore easier for an average non-technical user to make precise queries using her own terminology, without waiting for a database technician to formulate a complex SQL query.

SIRIUS' skills and tools are useful in healthcare where data is spread over different legacy systems and various work processes need access to this data. We can demonstrate how complex queries across these datasets can support work processes in healthcare. We can also provide workflow-relevant user interfaces for non-technical health-care personnel. In addition, researchers from the Work Practices strand, social scientists from NTNU, are available to help with design and assessment of how work flows and processes can be made more efficient through the use of new information technologies.
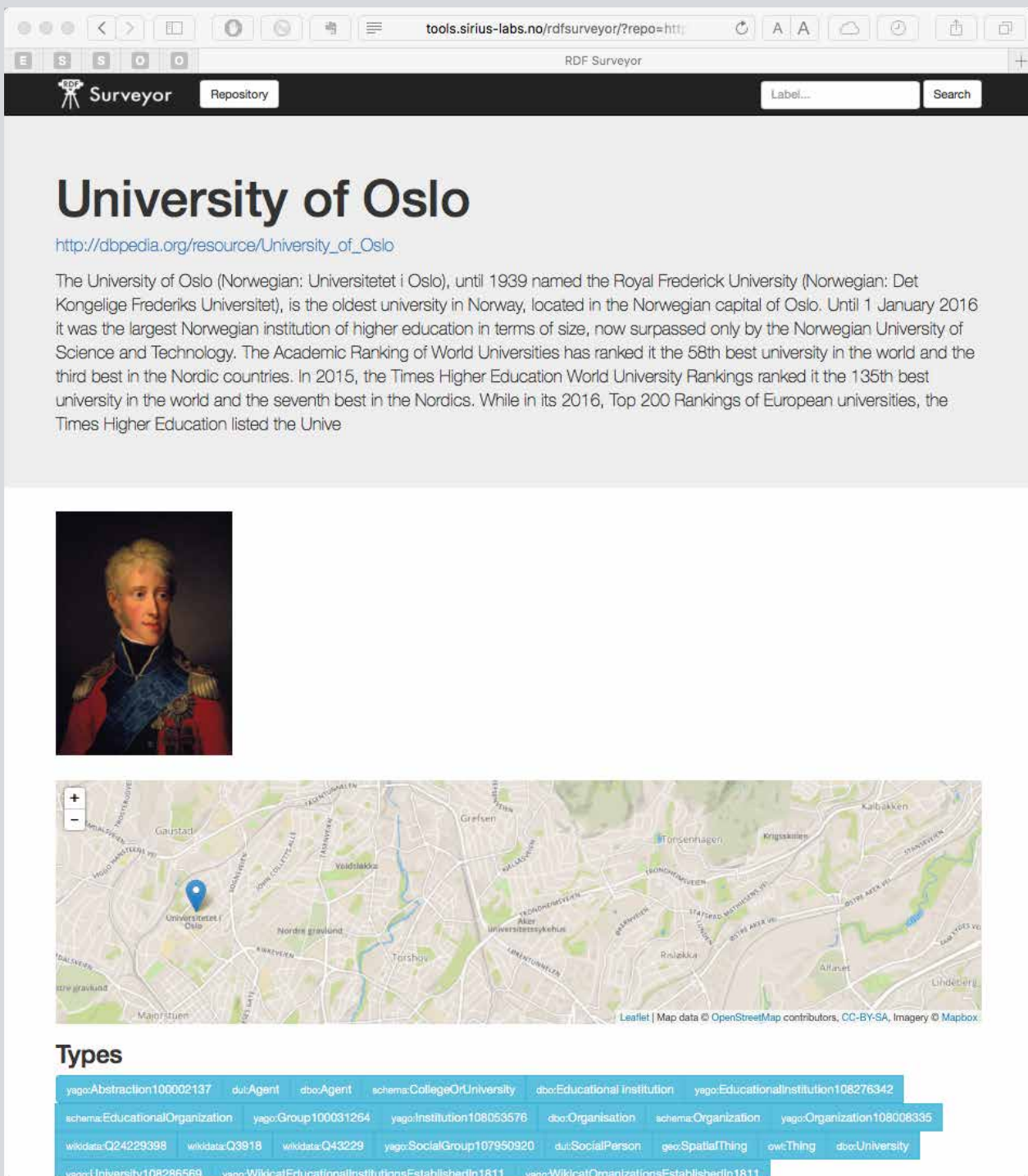
The Natural Language strand in SIRIUS has proven skills in processing domain-specific language and Norwegian language documents. The goal of the strand is to improve natural language processing for specific languages and technical sub-languages, like the ones found in medical records. Healthcare documentation of patient care contains narrative texts with many abbreviations, Latin terms and jargon, and conventional grammar is often not observed. Much work has already been done in English, but existing tools need to be adapted to Norwegian. Further research in this area focuses on machine understanding of the texts and integration with structured data. This allows clinicians to extract information from texts and then interact with the structured data and knowledge available. The Natural Language group in Oslo is one of the leading research centres for Norwegian language processing.

The Data Science, Databases and Scalable Computing strands in SIRIUS can provide precision medicine projects with skills and infrastructure for:

- Statistically analysing and modelling data.

- Storing and retrieving data from high-performance RDF databases. The RDFox database developed by SIRIUS partners at the University of Oxford is a fast and powerful database for storing linked data.

- Running hard, large and complex calculations quickly.

- The SIRIUS laboratory provides a testbed for these projects.

Finally, the Execution Modelling and Analysis (EMA) strand works with the simulation and analysis of complex plans and systems. We can determine the optimal deployment of computers in a cloud implementation of a health system so that applications function as expected when put online. We can also determine optimal deployment of health workers and equipment in a hospital. This is accomplished by modelling and analysing common hospital operations, ensuring that operating theatres and staff are available, and at the same time, allowing for replanning in the event of unforeseen changes. SIRIUS can contribute to projects run by healthcare management organizations, infrastructure providers and IT providers who want to understand and improve their operational performance. On the local institutional level, healthcare facilities can receive help on how to execute delivery plans.

SIRIUS researchers have created several visual query tools

Einar Broch Johnsen, Rudi Schlatte, Manolis Kourabakis and Dag Hovland in Tromsø for the SIRIUS-CIRFA workshop in September.

Arctic Earth Observation

The CIRFA Centre for Integrated Remote Sensing and Forecasting for Arctic Operations, based in Tromsø, is another of the Research Council's Centres for Research-based Innovation. Its aim is to develop new and better ways of observing the environment in the Arctic; this being a prerequisite for safe industrial operations in this area.

Remote sensing and forecasting produces large amounts of data in many formats. Satellite and aerial pictures and radar images are an important primary source of information. However, unstructured data, like reports, is also important. Processing, modelling and forecasting can consume large amounts of computer power and require intelligent management of data and results. CIRFA's Work Package 7 has the ambition to develop a pilot Integrated Remote Sensing Service that will allow oil companies and other interested parties to use data for daily decision support in their applications.

SIRIUS and CIRFA share a common partner company, namely Statoil. Discussions between Statoil's contacts, CIRFA and SIRIUS suggested that SIRIUS may be able to provide research-based IT solutions that could improve the performance and usability of the CIRFA pilot. An initial meeting was held in Tromsø in June 2017. This meeting identified two areas where SIRIUS and CIRFA could work together. The first was in cloud deployment. The CIRFA pilot will use simulating and processing components produced by CIRFA researchers and deployed on a cloud computing platform. Ensuring good resource use and reliable performance is non-trivial. SIRIUS'

Satellite and aerial pictures and radar images are an important primary source of information.

tools and methods for analysing and designing complex distributed systems could be of use. The second area was around data management. SIRIUS' work in ontology-based data access could improve the sustainability and usability of the CIRFA pilot.

This meeting was followed up with a two-day workshop with SIRIUS and CIRFA researchers, held on 11th-12th September 2017. A SIRIUS collaborator from the National and Kapodistrian University of Athens, Manolis Kourabakis, was invited to present his work on semantic web applications in earth observation, as implemented in several European Union projects. SIRIUS presented their design tools for cloud deployment and CIRFA presented their technology stack and needs for novel IT solutions.

It was decided to recruit two PhD students to work on SIRIUS-CIRFA problems, with shared supervision. One will work on knowledge representation problems related to remote-sensing decision support systems. The other will work on efficient cloud deployment of the CIRFA pilot. In addition, CIRFA and SIRIUS were partners in an application to the European Union's Marie Curie program. This application was coordinated by Professor Kourabakis in Athens and was submitted in January 2018.

Supporting Technologies

The Supporting Technologies work package organizes SIRIUS projects that work on problems which are relevant to several of the business work packages. These projects are all experimental work and are mostly PhD projects. However, larger fundamental projects in methodology for, for example, high-performance computing are also managed in this work package.

These projects are also used to provide recruiting positions that build up SIRIUS' long term delivery capacity. The projects are aligned with a strand. This means that progress on the projects is reported in the relevant section for each strand.

Here we will describe two fundamental activities that cut across work packages and strands. The first of these is a project called ADAPT, which was awarded a Young Research Talents grant by the Research Council of Norway. This project develops and applies formal methods at the intersection of database technology and high-performance computing.

The second area of fundamental research relates to the inner life of the SIRIUS centre: The social and technological studies skills of the Work Practices strand. Successful innovation from our centre requires a well-designed and thought out approach to collaboration and technology transfer. The centre itself is a laboratory for technology studies.

SIRIUS' New FRINATEK Young Research Talent

In December 2017 there was jubilation on the 8th floor of the Informatics building when it was announced that Silvia Lizeth Tapia Tarifa, a SIRIUS researcher, had been selected for a FRINATEK Young Research Talent grant. This gant, one of only 10 made across all the natural sciences and technology and the only one in computer science in Norway, gives four years' financing for her work and pays for a PhD student to work with her on her project.

Lizeth first came to the University of Oslo in March 2009. She had met Einar Broch Johnsen, now SIRIUS' Deputy Director, during a research internship at the United Nations University in Macao. Because of this contact, she had been invited to visit the University of Oslo as a guest student and research assistant on a one-year contract. This one-year contract was first truncated by visa delays but in the end turned into a Msc, a PhD, participation in various EU projects and now nine years in Norway.

Brazil was Lizeth's first choice for postgraduate education after completing her Bachelor's degree in Computer Science and Engineering in her homeland, Peru. However, a summer school in Peru run by the UN University resulted in an invitation to go instead to Macao for a year. In between these academic stays, she worked as a programmer and system analyst, the last in a start-up company supported by the Inter-American Development Bank, the largest source of development financing in Latin America.

Lizeth's home town is called Arequipa, a town that lies 2335 metres over sea level at the foot of an active volcano. As noted above, all her undergraduate education was in Peru. Having been in Norway for nine years, she enjoys its nature and walking possibilities around Oslo. She noted that she likes to do this with friends: not like a solitary activity as typical for Scandinavian culture.

First row: Silvia Lizeth Tapia Tarifa & Irep Gözen.
Second row: Morten Dæhlen, Svein Stølen & Inger Skrede.
Back Row: Tobias Rolf, Philipp Häfliger og
Hans Kristian Kamfjord Eriksen.
Photo: Bjarne Røsjø, UiO.

The FRINATEK project granted to Lizeth is called *ADAPT: Exploiting Abstract Data-Access Patterns for Better Data Locality in Parallel Processing.* It is a project that is done in collaboration with two SIRIUS partners: Numascale and the University of Oxford. Collaborators outside SIRIUS are AIA Science (a spin-off from Numscale), Imperial College London and CWI in Amsterdam. The project's advisory board has representatives from Numascale, Statoil and ARM Microsystems.

The project won support at the second attempt. Lizeth identified two things that made the successful application better than the first. The first of these is the case study provided by SIRIUS built around using software from the University of Oxford on infrastructure from Numascale. The second improvement was in finding a better balance between ambitious goals and realistic scope in the project plans.

Put simply, Lizeth's project looks at how a complex computer program can be run on a parallel or distributed computer so that it works optimally. Good performance for these systems depends on two things: where the different tasks of a computer program are run and where the data is stored in memory. Load balancers determine where the computer program is run, while memory allocators decide where data is stored. Finding the best set up for load balancing and memory allocation is, at present, an art rather than a science. Design is based on trial and error, rules of thumb and experience. Lizeth aims to find a systematic way of designing load balancing and memory allocation for a program. Her approach is based on programming language theory and model-based simulations, aiming to develop a novel and advanced static analysis technique to find information that can be used to improve load balancers and allocators. To

take a simple example: often all the information in a system about an object is grouped together in memory. This can cause inefficient data access, where only certain parts of this data are used often in calculations happening in various computation nodes. It may make sense to collect this information from all individuals in the system and store it together in memory, near the computation nodes. Or, consider another case. Here two computation processes each want to use the same data. The load balancer and memory allocator must cooperate to find better ways for these two processes to share this data.

This approach will be validated using RDFox, a triple-store database system from the University of Oxford. Lizeth and her PhD student will examine how this new static analysis technique can be used to design the running of the data-base on Numascale hardware. This project will give benefit to SIRIUS as a whole, as results of this novel technique will complement the tools of the SIRIUS laboratory using model-based predictions and also because RDFox is a core part of the SIRIUS laboratory and can be a useful tool for the SIRIUS pilot projects that use semantic technologies.

Lizeth is also participating in the mentoring program that started last year. Her mentor is Åshild Hanne Larsen, the CIO and SVP Corporate IT of Statoil. The mentoring has been valuable for Lizeth. It has been good to able to talk with a smart and experienced woman with high responsibility and long experience in managing people. Lizeth received practical advice on how to lift her vision up from just the technical work, to look around and see things in an integrated way. Another important message was the need to have a long-term plan. At this stage, this plan is focused on academia, to the benefit of SIRIUS and the University.

Work Processes in Consortia

One of the interesting things about SIRIUS is that its internal processes are themselves a subject for research and innovation. It is challenging to define an industrially-relevant program of research in computer science for innovation in the oil and gas sector. We need to bridge the gaps in understanding and incentives between IT academics, IT vendors and the oil and gas business. This means that we must find projects and ideas that create benefits for all three groups. This is a complex process of creative thought and negotiation. It takes time and requires fora where shared understanding can be generated.
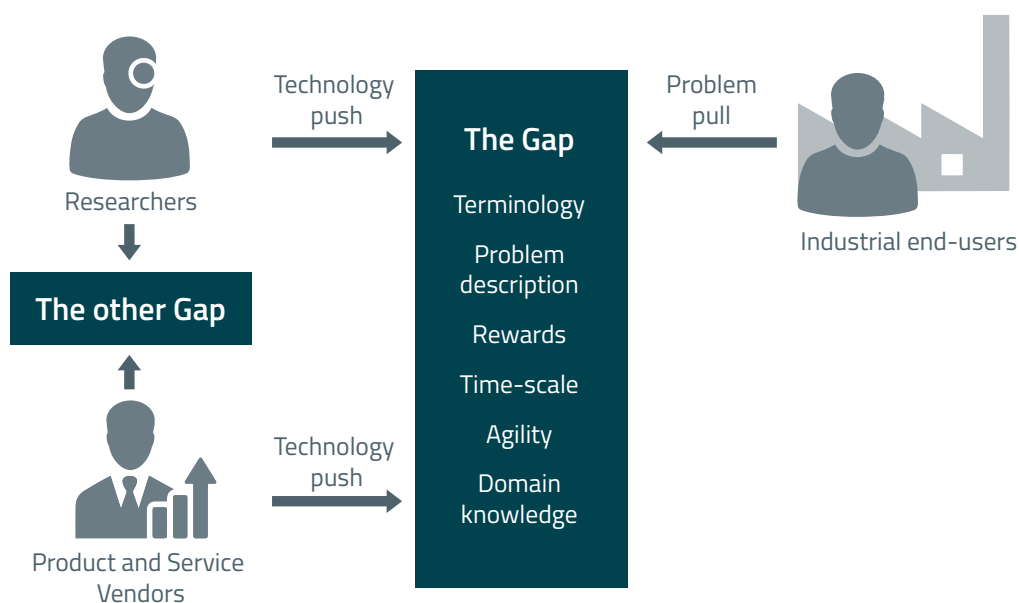
Often an academic comes with a defined research program that reflects their skills and network. This program has a tenuous linkage to business problems and a communication style that is difficult to understand for non-specialists. The IT vendor comes will a set of finished, ready-to-use software and hardware products that can be configured into a solution for a business problem. Academic computer scientists, however, are not rewarded for configuring solutions from products. Business representatives come to discussions with "shopping lists" of short-term problems that may or may not have research-based solutions.

This starting point leads to mutual misunderstanding and projects that don't meet the aims of a centre for research-based innovation. The academics revert to normal practice: publishing within their field of expertise. Business representatives are frustrated because they cannot see the link between the academic work and their daily needs. Vendor

companies are frustrated because neither the businesses nor the academics can find a use for their products.

The innovation pipeline in SIRIUS is designed to address these pitfalls and build a framework where the interests of all participants are advanced. This is the experiment-prototype-pilot framework.

The Work Practices strand in SIRIUS has its roots in the departments of Computer Science and Social Research at NTNU in Trondheim. Their research interests include how complex organizations work to implement digital transformation, technology implementation and adoption. Researchers from this strand are active in many of SIRIUS' projects, gathering requirements, interviewing stakeholders and running workshops. In addition to doing this work, they also examine how the social and innovation processes in SIRIUS are working and how they can be improved. SIRIUS
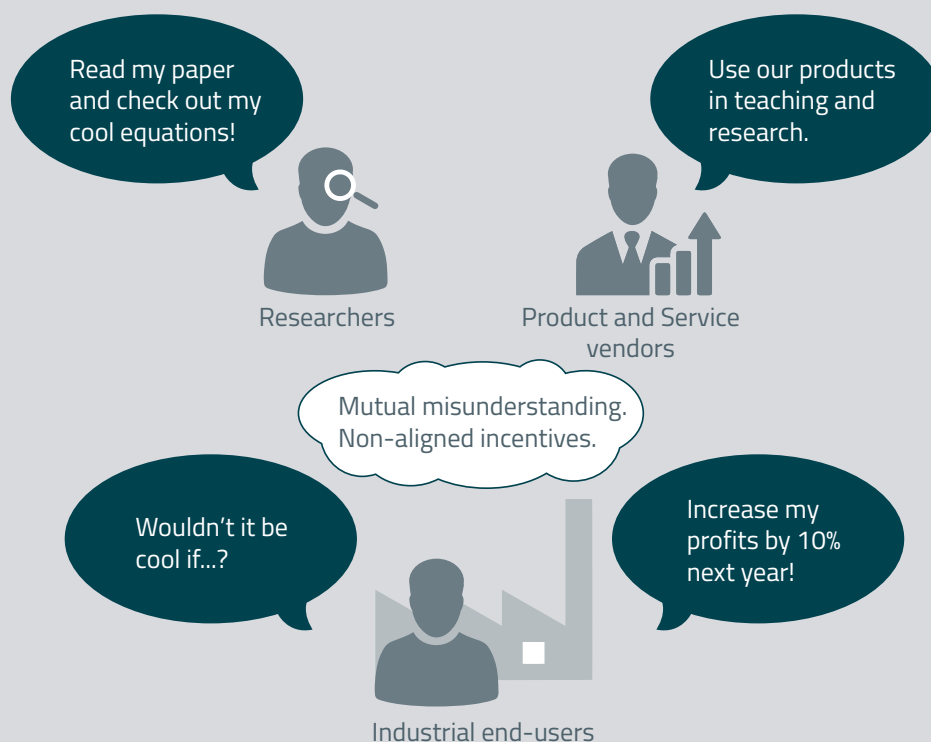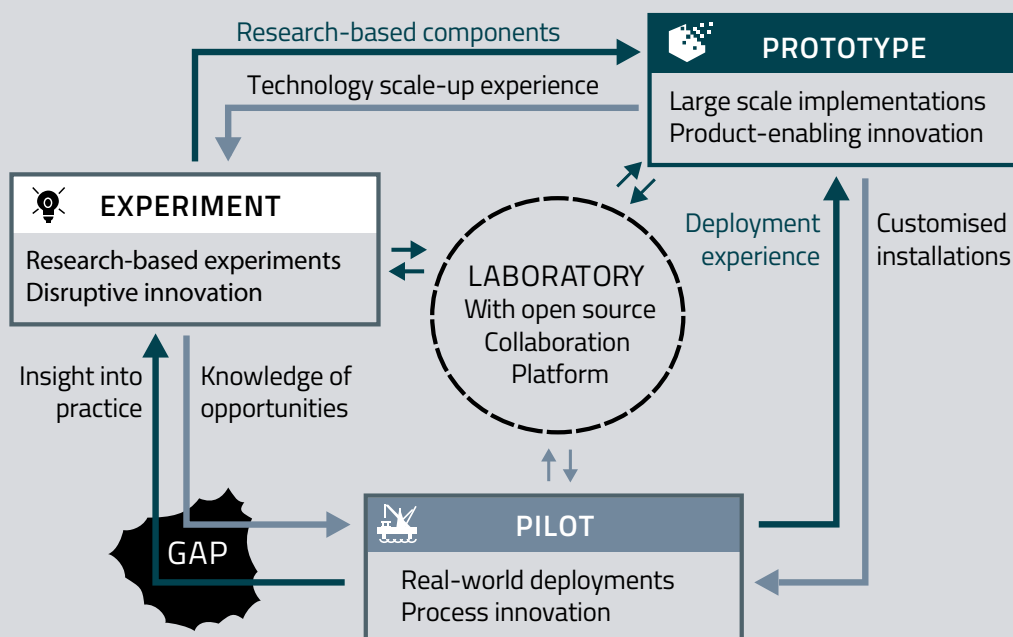
itself becomes a laboratory for social science research around digital and technical innovation in the oil and gas industry.

For example, Thomas Østerlie, the strand leader for Work Practices has published a conference paper[1], together with Elena Parmiggiani and Eric Monteiro, where they use the work processes in a SIRIUS innovation activity, suitably anonymized, as a basis for applying their model of how explo-ration geologists do their work. This type of publication will be an important part of SIRIUS' academic output. We also believe that this introspection is necessary for us to meet our technical ambitions.

---

[1] "Information infrastructure in the face of irreducible uncertainty", 5th Innovation in Information Infrastructures Workshop, Rome, November 2017.

International Activity and Dissemination

ESOCC 2017

The European Conference on Service-Oriented and Cloud Computing (ESOCC) is about advances in the state of the art and practice of Service-Oriented Computing and Cloud Computing in Europe. Cloud computing is a game changer for European software industry, allowing products to move from expensive on-site installations to flexible, service-oriented models.

ESOCC is an annual technical conference gathering researchers from academia and industry. It is held annually at different locations in Europe. In 2017 it was organized in Oslo, from September 27-29th, with support from SIRIUS, the University of Oslo and the Research Council of Norway. Our primary aim in hosting ESOCC 2017 was to organize the sixth conference in the conference series in a way that contributed to the further development of the series and to the cloud computing and services community in Europe. ESOCC 2017 attracted around 60 participants.

Our further objectives for hosting ESOCC 2017 in Oslo were:

1. **Expose international research to the Norwegian community.** ESOCC 2017 involved the University of Bergen, SINTEF and SIRIUS in the organization.

2. **Promote and further strengthen strategic connections (both research and education)** with leading international groups. We made many new, interesting connections and are developing some of these for future collaborations and EU projects.

3. **Inspire students and the public by showcasing groundbreaking research and software verification technology.** ESOCC 2017 covered a broad range of research, spanning from theoretical results to applications and case studies. Alexander Lenk, from BMW's research division in Munich, gave an inspiring talk on their work on connected cars, combining sensor technology and the cloud. A special track on industrial experience highlighted industrial take-up and facilitated an exchange of experiences between academia and representatives from different industries. This panel included Alessandro Rossini from EVRY, who held a keynote address, and David Cameron from SIRIUS.

ESOCC's industry keynote was given by Alexander Lenk from BMW.

Industry panel at ESOCC 2017

ESOCC 2017 included a PhD symposium that highlighted PhD students and their research. The conference attracted industrial participants, which ensured a good networking occasion for students.

Industrial Ontology Colloquium

The Oslo Industrial Ontology Colloquium has been a regular event at the University of Oslo in 2017. It brings together

researchers and practitioners in the application of ontologies and semantic technologies in industrial practice. In 2017 the seminar was organized by Andreas Nakkerud (from the Department of Mathematics and the DataScience@UiO cluster) and Johan Klüwer from DNV GL. These seminars have been attended by staff from companies including DNV GL, EPIM, PCA, Aker Solutions, Aibel and Sopra Steria. The topics of discussion have been a mix of industrial case studies and introductions to fundamental methods.

European Projects Completed or Near Completion
2017 was a year in which SIRIUS researchers completed work or follow-up on several European Projects. Final presentations for the Optique project were held in January 2017 in Luxembourg. Optique had formally ended November 2016 and had met its project goals already at the end of 2015. The project received a very favourable review.

The HyVar project had its last full year of operation and will finish in early 2018. In this project, SIRIUS researchers have been applying the ABS simulation tool to solving complex cloud and Internet of Things deployments, with an application in the automotive industry.

The Envisage project, which developed ABS as a tool to design large and complex cloud-deployed systems (such as web shops) also ended in 2016. The ABS tools developed there are available at the Envisage website (**www.envisage-project.eu**).

European Private-Public Partnerships: BDVA and A.SPIRE
We believe that Horizon 2020 is a natural source of support and financing for SIRIUS' beacon projects. Successful participation in Horizon 2020 requires effort and investment in building networks and alliances with European companies and researchers. SIRIUS has been a full member of BDVA, the Big Data Value Association, since 2016. This private-public partnership (PPP) brings together large IT companies and the IT research institutions in Europe to build a strategic road map for data-based innovation in Europe. This PPP is focused on the provision of IT tools to solve generic problems in European society and industry. In many ways, it represents the technology push part of our innovation plan.

It is a challenge in Horizon 2020 that technology pull from business is organized by a number of different private-public partnerships. Thus, the EFFRA PPP organizes the manufacturing sector around Factories of the Future, while AIOTI organizes businesses and researchers around the

Internet of Things. The process industries have their own PPP: called A.SPIRE. This PPP is aware of the need for research and innovation around the digitalization of process operations and has proposed a project to meet this need in the 2019 program for Horizon 2020. SIRIUS' beacon projects are defined by technology pull and fit well into A.SPIRE's ambition. SIRIUS is therefore participating in the A.SPIRE PPP from 2018.

Subsea Valley
SIRIUS has continued its collaboration with the Subsea Valley cluster. The Faculty of Mathematics and Natural Sciences at the University of Oslo is partner in the cluster and SIRIUS wrote a support letter for Subsea Valley's successful bid for National Centre of Expertise status.

The annual Subsea Valley Conference is an important event for SIRIUS. It provides the centre with a forum for communicating with the oil, gas and engineering companies in South East Norway. As in 2016, SIRIUS ran a master class at the conference in 2017. This class had the title "Digitalization of Engineering" and brought together three lectures. David Cameron provided an overview of the topic while Bjørn Berli from EPIM introduced the application of digital technologies to the management of requirements in engineering. Johan Klüwer from DNV GL then followed this up with a practical demonstration of the use of knowledge representation tools to describe and enforce requirements.

Cutting Edge Festival and SIRIUS Innovation Forum
SIRIUS participated in the Cutting Edge festival in September 2017 with a workshop on "Renewing Norwegian Industry". This workshop brought industrial perspectives from Åshild Larsen, CIO in Statoil and Hans Kristian Danielsen from DNV GL. These perspectives were complemented by presentations by David Cameron and Arild Waaler from SIRIUS. The event was moderated by Bergljot Gundersen from the University of Oslo.

This event was also used to launch the SIRIUS Innovation Forum. This was conceived as a way for SIRIUS to be able to collaborate with start-up firms and small enterprises. The forum provides a way for an interested firm to sign up for our mailing list and to participate in SIRIUS' technology transfer activities.

This launch did not generate the interest we expected. This may have been due to the launch being held in a closed conference with an admission fee. We are planning

The simple ontology demonstrated at the Subsea Valley Conference
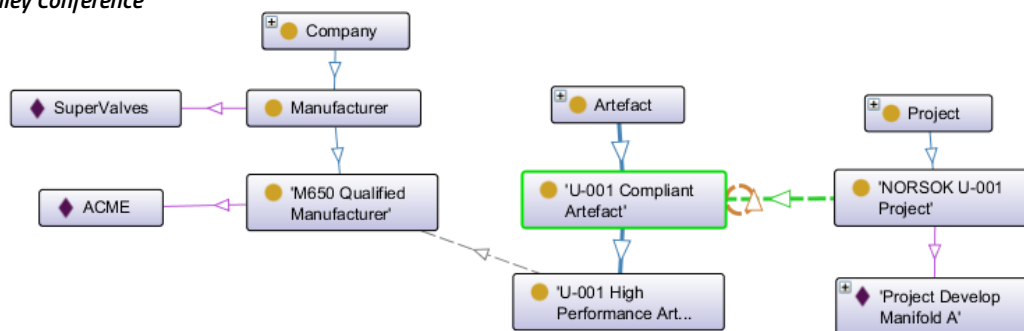
a re-launch of the forum at the Subsea Valley conference in March 2018.

Alan Turing Institute

The Alan Turing Institute (ATI) is the United Kingdom's national institute for data science. It was founded in 2015 by five British universities: Cambridge, Edinburgh, Oxford, University College London and Warwick. SIRUS participates in the centre through the involvement of Ian Horrocks' group at the University of Oxford.

Two SIRIUS researchers have been selected to be visiting scholars at the ATI in 2018. Evgenij Thorstensen will be working there in the second half of 2018 and Ernesto Jimenez-Ruiz will also be working there while maintaining a role in SIRIUS projects. Both of these workers have interests in the linkage of data science and knowledge representation.

Norwegian Energy Partners (NORWEP)

SIRIUS funds the University of Oslo's partnership in Norwegian Energy Partners (NORWEP), which was formed in 2017 by the merging of INTSOK and INTPOW. We do this because membership provides us with a forum for communicating SIRIUS' research interests to two important groups: firstly, foreign oil and gas companies and secondly, the supplier companies who are primary members of NORWEP. Other research groups in the university, in informatics, mathematics and geosciences can use this partnership.

All the main Norwegian Universities are members of NORWEP. We use our membership to find ways in which our research and education can follow Norwegian suppliers out into the international market.

A presentation was made of SIRIUS program and results at a French-Norwegian thematic day on Digitalization of the Oil and Gas Industry. This meeting was held in Paris La Defense on 31st May 2017. The presentation was made at a conference attended by leaders in the French and Norwegian oil and engineering industries. Other presentations were made by senior managers in Total, TechnipFMC, Aker Solutions and AkerBP.

After the conference, a Norwegian delegation went to

Toulouse for a study tour of digitalization in Airbus. This involved a morning seminar on digitalization and data integration initiatives in Airbus, followed by a tour of the A380 factory.

Meetings have also been held with NORWEP's country managers for the United Kingdom and Brazil.

Brazilian Collaboration

SIRIUS has followed up the visit to Brazil with the Minister for Education in 2016 with participation as a Keynote Speaker and session chair at the 5th November Conference for Norwegian-Brazilian Cooperation in Energy Research, held in Rio de Janeiro in November 2017. David Cameron held a keynote address with the title "Digitalization: How can we mix the 'new oil' and the old oil?". The parallel session on digitalization contained presentations from the University of Oslo, NTNU, IBM, Petrobras, Federal University of Rio de Janeiro, Federal University of Rio Grande do Sul and CBPF (the Brazilian research institute for physics).

During 2017 SIRIUS has established a set of promising collaborations with universities and companies in Brazil. Following the November conference, David Cameron participated in meetings with Petrobras and other oil companies in Rio de Janeiro. In addition, contact via the PPDM standardization organization resulted in contact with the group for Petroleum Informatics at the Federal University of Rio Grande do Sul (UFRGS). This group, led by Professor Mara Abel, has done work in applying knowledge representation technologies to problems in oil and gas exploration. Their approach complements that of the SIRIUS team. This collaboration began with the preparation of proposals to funding bodies in Norway and Brazil for three educational exchange projects. All the proposals were successful. This means that SIRIUS will be starting an ambitious program of academic exchanges with UFRGS in 2018, financed through the UTFORSK and UTFORSK-CAPES programmes run by SIU, the Norwegian Centre for International Cooperation in Education. This work will be complemented by an INTERNABROARD project that allows University of Oslo students to do their Masters' project in collaboration with Petrobras' Libra project organization.

Recruitment, Education and Equal Opportunity

Recruitment

SIRIUS recruited three categories of personnel in 2017: PhD fellows, post-doctoral researchers and technical-administrative workers to support the innovation program.

New PhD fellows recruited in 2017 were, at the University of Oslo, Sigurd Kittilsen and Andreas Thune. NTNU recruited Mina Haghshenas, while Temitope Ajileye started at the University of Oxford.

Jiaoyan Chen was recruited as a Postdoctoral Research Associate at the University of Oxford. An additional position was advertised but was not filled. It has now been re-advertised. At the University of Oslo, we have chosen to use SIRIUS financing to retain experienced postdoctoral researchers. Thus, in 2017, Lizeth Tapia and Violet Pun were financed in this way.

SIRIUS also employed a technical programmer, Shahila Retnadhas, this year. Her role is vital for the Centre's innovation program. A professional programmer is needed to provide the services and framework necessary for academic software to be used in innovation. The technical programmer runs the configuration management and software quality systems in SIRIUS and assists researchers in improving the robustness and quality of their software.

SIRIUS also employed two workers in the field of Digital Geosciences. Initially, these workers are financed by the Statoil Academia program. Irina Pene is employed at the Department of Geosciences as a joint researcher with SIRIUS. She is a geophysicist with long industrial experience. Adnan Latif has been employed at SIRIUS as a project coordinator and participant in the Exploration work package. He also has a geosciences background.

Education

SIRIUS started its recruitment of Masters' students this year. For the first time, SIRIUS proposed projects as a separate research group at the Department of Informatics, whereas previously SIRIUS academics supervised candidates through their former research group. The candidates recruited will be starting their research projects in the second half of 2018. As mentioned on page 76, we have received financing for joint supervision of Master's students with the Federal University of Rio Grande do Sul in Brazil and for industrial

Ian Horrocks lectures on Semantic Languages at the Semantic Technologies tutorial

Masters' projects with Petrobras. In these programs, the student will be able to visit Brazil for up to three months. We will recruit students who begin in summer 2018 into these programs.

Focus in 2016 and 2017 has been on PhD education, with the SIRIUS Mentoring program as its main tool. The mentoring program is described in more detail below. In 2018, we will transfer focus to the MSc program.

Tutorial on Semantic Technologies and Reasoning in Oxford

In the last week of March 2017, a one-week course was held at the University of Oxford on semantic technologies and reasoning. This was co-organised by SIRIUS and had industrial participation from SIRIUS partners and Norwegian industry. The instructor on the first day was Ian Horrocks, SIRIUS' scientific coordinator. This was followed up by practical training in the Optique platform. Dag Hovland from SIRIUS led the practical work. The course was jointly organized by SIRIUS and the British DbOnto platform project. Industrial practitioners from EDF, Aibel and OSIsoft particpated.

Per Eivind Solum, an experienced manager in Schlumberger, is mentor for UiO researcher Ernesto Jimenez-Ruiz. The SIRIUS centre is exploring new ways of advancing cooperation between business and academia.
by Gunhild M. Haugnes

The new Vice-Dean for Innovation in the Faculty of Mathematics and Natural Sciences, Kristin Vinje, talked to a SIRIUS lunch in November.

Mentoring: Building Bridges Between Industry and Research

"*I have extensive experience with mentoring and coaching internally in Schlumberger. With us, it is often the case that a technical expert is mentor and advisor for a junior employee. However, being a mentor for a researcher is something completely new and is very exciting*", says Solum and nods towards Ernesto Jimenez-Ruiz, who has been his mentee since last April.

"*In academia it is often the case that a professor is mentor or advisor for doctoral students and junior research workers. Having a mentor from industry gives a completely different set of insights and a new contact network*" replies Jimenez-Ruiz.

Ten mentor-mentee pairs

SIRIUS, which is a centre for research-based innovation (SFI), has set up a project with ten mentor-mentee pairs. The University of Oslo leads the Centre and Schlumberger, a large international oil service company, is one of the industrial partners in the centre.

Jimenez-Ruiz is a researcher in semantic web technologies. Solum also has an IT background but in a different field to the UiO researcher. However, Solum can introduce him to colleagues with similar background and skills.

"*It will be very interesting to see if there is a group that I can work with. Perhaps we can develop something new, or set up a new collaborative project. Nevertheless, it is good to learn about what people are concerned about and how work is done in industry. In academia we are often very focused on the next publication*", says Jimenez-Ruiz.

He believes that the mentor-mentee program can be good for his career. Even though he enjoys academia, he doesn't reject the idea of a job in business at a later stage. "Building networks with key people in industry gives me a unique skill and makes me more attractive on the job market".

Per Eivind Solum, Ernesto Jimenez-Ruiz and Ingrid Chieh Yu hope to contribute to personal development for researchers, build networks and create innovation through a unique mentoring programme. Photography: Gunhild M. Haugnes/UiO

It is important for us in industry to get an insight into new research and how researchers work.

Can lead to more long-term thought

Solum noted that many in industry know little about what happens in university research groups.

"*It is important for us in industry to get an insight into new research and how researchers work. We have a busy daily routine. There is lots of day-to-day work, often to meet short-term ends*", says Solum, who believes that a mentor program can contribute to more long-term thought.

"*We can step back and dive deeper into development and challenges that can be important for us in the long term*." This applies to everything from software development to more fundamental changes in methods. We see the value of developing teams that consist of people with different backgrounds and perspectives.

Personal development of mentees

"*Personal development and career planning for doctoral students and postdoctoral researcher is, in general, something we see very little of in academia. This is central in this rather rare mentor-mentee program*", says Associate Professor Ingrid Chieh Yu, who is responsible for running the project.

She also hopes that innovative projects can be conceived because of the mentor–mentee program.

"*SIRIUS is a centre for research-based innovation, and it is therefore central for us to find new ways of stimulating innovation, not least finding ways for researchers and industry to do this together. At the same time, we hope to be able to build an interdisciplinary network between industry and research*."

Equal Opportunity

The mentoring program is a vehicle for promoting and training future female research leaders. The 2017/2018 program had six of ten female mentees. The program also benefitted from the participation of senior female executives in Statoil as mentors.

About a third of our junior researchers are female. This is better than the average for computer science departments in Norway. Unfortunately, there are only two females in the leadership of the centre.

She has just held a meeting with mentees from UiO and NTNU, where they shared their experience of the program.

"*What has the experience been thus far?*"

It is still a bit early to say. We will do a detailed evaluation in April, but we are very pleased thus far. She concluded this summary by emphasising that this initial project is a pilot project.

Communication is the key

Ingrid Chieh Yu noted further that good communication is often the key. This can be a challenge when the mentor and the mentee come from different countries and cultures. "*We have worked a lot with this and have used external experts to help. It is important to gain experience on what is needed to build a good partnership and how we can get people from different cultures to work together*".

The aims of the Mentoring Project

- Offer researchers a personal development strategy and cultivate future research leaders.

- Exchange of expertise, values, perspectives and attitudes, with network building.

- Encourage collaboration, engagement and building of skills on both sides of the mentoring relationship.

- Develop and exploit the diversity of culture, age, gender and skills in SIRIUS.

SIRIUS' mentor-mentee program

Mentors are senior managers and technical leaders from Statoil, Schlumberger, IBM, EVRY and Numascale. The mentees are PhD students and postdoctoral research fellows at UiO and NTNU.

- Mentor-mentee program lasts 12 months.

- Pairs are arranged based on profiles and meet 1-2 times each month.

- Other activities are seminars on personal development, arenas for networking and social interaction with the other pairs to share experience.

- A Mentor master class was held to prepare the mentors before the programme started.

Governance

General Assembly

The General Assembly is the body that makes final decisions in the centre. It consists of one high-level representative from each partner. It meets physically twice a year, with video meetings as otherwise needed. The spring General Assembly focuses on the strategy of the centre, whereas the autumn meeting discusses and approves the centre's work plan. Two meetings were held in 2017, both in Oslo.

Executive Committee of the General Assembly

An Executive Committee was elected by the General Assembly in May 2017. Its role is to act on behalf of the General Assembly to provide regular supervision and review of the centre's operations and strategy. This committee has been described in more detail on page 29. It meets monthly.

Strategy Board

The Strategy Board defines the strategic plans for the SIRIUS centre and for approving the projects in the SIRIUS project portfolio. It is chaired by the Centre Leader and consists of the Scientific Coordinator, Intellectual Property Manager, Pilot Strategy Coordinator, the Faculty Research Strategist and the leader of the Strategy Work-package (WP5)

Operations Board / Centre Management

The Operations Board (or Centre Management team) is responsible for day-to-day operation of the centre and the definition of work plan that implements the decisions of the Strategy Board. It is chaired by the Operations Manager and consists of the Centre Leader, Administration Manager, Mentor and Education Coordinator, Work Package Leaders and Strand Leaders.
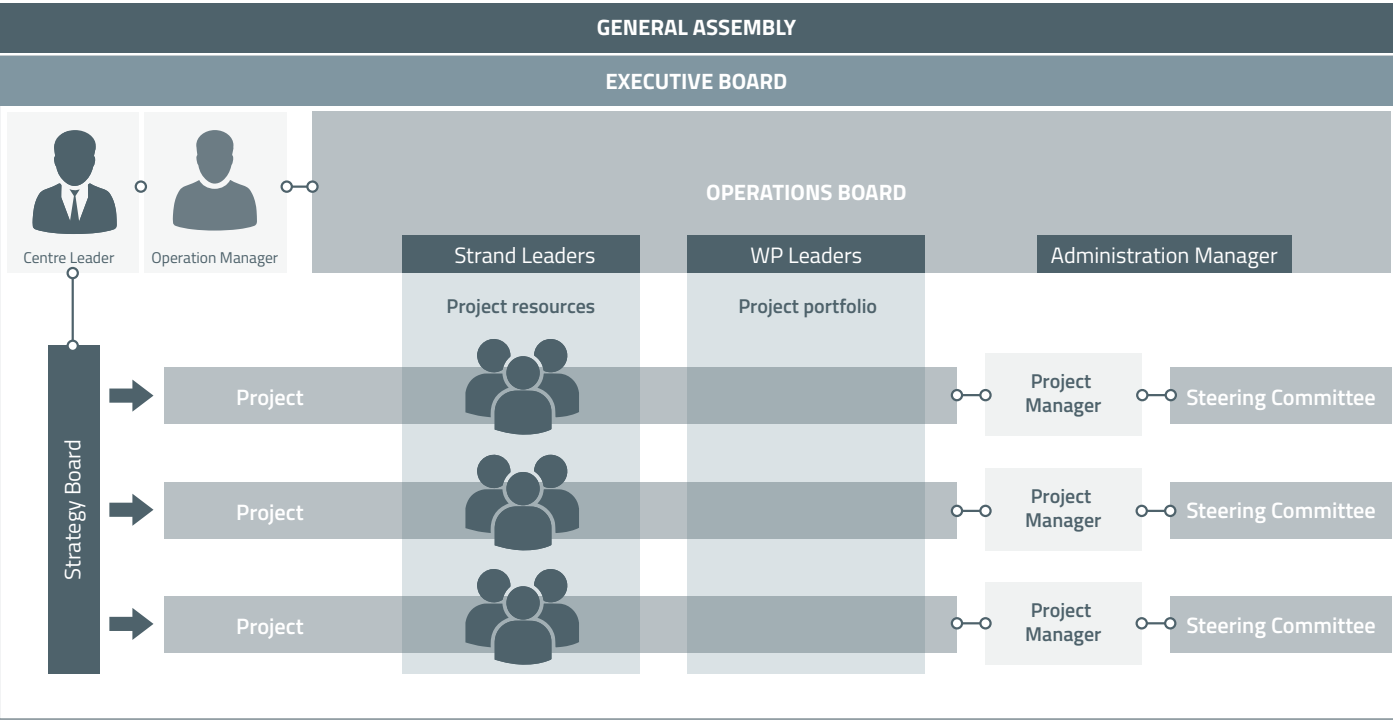
The strategy board and its observers: Einar Broch Johnsen, David Cameron, Geir Horn, Ingrid Chieh Yu, Ian Horrocks, Knut Sebastian Tungland, Arild Waaler and Eric Monteiro.
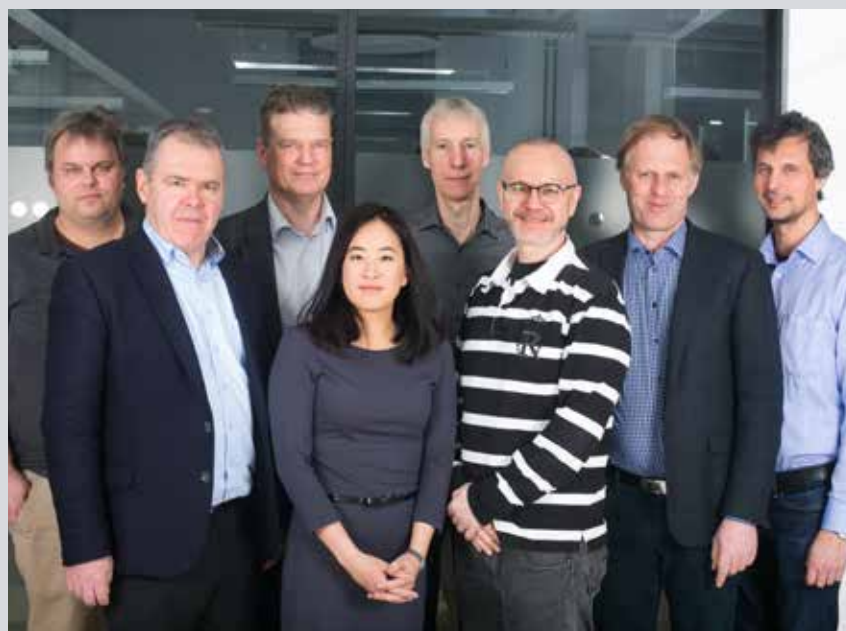
Roles

Role	Name	Affiliation
Chairman of General Assembly	Knut Sebastian Tungland	Statoil ASA
Centre Director	Arild Waaler	University of Oslo
Deputy Centre Director	Einar Broch Johnsen	University of Oslo
Centre Coordinator/Operations	David Cameron	University of Oslo
Scientific Coordinator	Ian Horrocks	University of Oxford
Administration Manager	Lise Reang	University of Oslo
Finance Manager	Geir Ulvestad	University of Oslo
Intellectual Property Manager	Jarl Magnusson	DNV GL
Mentor & Education Coordinator	Ingrid Chieh Yu	University of Oslo
Pilot Strategy Coordinator	Eric Monteiro	NTNU
Faculty Research Strategist	Geir Horn	University of Oslo
WP1 Exploration	Jens Grimsgaard	Statoil ASA
WP2 Operations	David Cameron	University of Oslo
WP3 Cross-Domain Applications	Laura Slaughter	University of Oslo
WP4 Supporting Technologies	Ian Horrocks	University of Oxford
WP5 Strategy and Outreach	Einar Broch Johnsen	University of Oslo
S1 Knowledge Representation	Martin Giese	University of Oslo
S2 Language Technologies	Jan Tore Lønning	University of Oslo
S3 Databases	Boris Motik	University of Oxford
S4 Execution Modelling & Analysis	Ingrid Chieh Yu	University of Oslo
S5 Scalable Computing	Geir Horn	University of Oslo
S6 Work Practices	Thomas Østerlie	NTNU
S7 Data Science	Martin Giese (acting)	University of Oslo

Partners

The SIRIUS consortium has the following partners:

- **Academic partners**
  - University of Oslo, Department of Informatics.
  - Norwegian University of Science and Technology, NTNU, Department of Computer Science.
  - University of Oxford, Department of Computer Science.
  - Simula Research Laboratories.

- **Industrial partners**
  - Computas.
  - DNV GL.
  - Dolphin Interconnect Solutions.
  - EVRY.
  - fluidOperations.
  - IBM.
  - KADME.
  - Numascale.
  - OSISoft.
  - SAP.
  - Schlumberger.
  - Statoil.

SAP joined the centre from the beginning of 2018.

Glossary

ABS   A simulation and analysis tool for complex computer and organizational systems.

EHR   Electronic Health Records

EMA   The Execution, Modelling and Analysis strand in SIRIUS.

GPU   Graphical Processing Unit: a computer processor that originally was specialized in the calculations needed for high-quality computer graphics. These processors are also well suited for neural net machine learning calculations.

HPC   High-performance Computing.

HPO   Human Phenotype Ontology

NVNe   Non-volatile Memory Express: a standard for connecting storage units together.

OBDA   Ontology-based data access.

OMIM   Online Mendelian Inheritance in Man: a database of genes and genetic diseases.

OUS   Oslo Universitetssykehus – Oslo University Hospital.

PCIe   Peripheral Component Interconnect Express: a standard for connecting computing equipment together with high communication speeds.

PDM   Project Data Manager: an employee in an oil company who manages and retrieves data about exploration prospects for end use by technical specialists.

RDF   Resource Description Framework: a standard model for data interchange on the Internet. One of the basic standards in knowledge representation.

SQL   Structured Query Language: a standard language for storing, retrieving and manipulating data in a database.

SPARQL   Query language that allows queries to data represented as RDF in a triple-store database. It is a language for making semantic or ontology-based queries.

Publications

Agibetov, Asan; Jimenez-Ruiz, Ernesto; Ondrésik, Marta; Solimando, Alessandro; Guerrini, Giovanna; Catalano, Chiara E; Oliveira, Joaquim M.; Patanè, Giuseppe; Reis, Rui L.; Spagnuolo, Michela. Supporting Shared Hypothesis Testing in the Biomedical Domain. Journal of Biomedical Semantics 2017

Baader, F.; Horrocks, I.; Lutz, C.; Sattler, U. An Introduction to Description Logic, Cambridge University Press. 2017

Benedikt, Michael; Konstantinidis, George; Mecca, Giansalvatore; Motik, Boris; Papotti, Paolo; Santoro, Donatello; Tsamoura, Efthymia. Benchmarking the Chase. 36th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems. 2017

Bijo, Shiji; Johnsen, Einar Broch; Pun, Ka I; Tapia Tarifa, Silvia Lizeth. A formal model of parallel execution on multicore architectures with multilevel caches. Lecture Notes in Computer Science 2017; Volume 10487 LNCS. s.58-77

Dalla Preda, Mila; Gabbrielli, Maurizio; Giallorenzo, Saverio; Lanese, Ivan; Mauro, Jacopo. Dynamic Choreographies: Theory and Implementation. Logical Methods in Computer Science 2017

de Boer, Frank; Serbanescu, Vlad; Din, Crystal Chang; Hähnle, Reiner; Henrio, Ludovic; Rochas, Justine; Johnsen, Einar Broch; Sirjani, Marjan; Khamespanah, Ehsan; Fernandez-Reyes, Kiko; Yang, Albert Mingkun. A Survey of Active Object Languages. ACM Computing Surveys 2017; Volume 50(5)

de Paoli, Flavio; Schulte, Stephan; Johnsen, Einar Broch. Service-Oriented and Cloud Computing - 6th IFIP WG 2.14 European Conference, ESOCC 2017, Oslo, Norway, September 27-29, 2017, Proceedings. Springer 2017 (ISBN 978-3-319-67261-8) 240 s. Lecture Notes in Computer Science (10465)

Din, Crystal Chang; Hahnle, Reiner; Johnsen, Einar Broch; Pun, Ka I; Tapia Tarifa, Silvia Lizeth. Locally abstract, globally concrete semantics of concurrent programming languages. Lecture Notes in Computer Science 2017; Volume 10501 LNAI. s.22-43

Domenech, Jesus; Genaim, Samir; Johnsen, Einar Broch; Schlatte, Rudolf. EASYINTERFACE: A toolkit for rapid development of GUIs for research prototype tools. Lecture Notes in Computer Science 2017; Volume 10202 LNCS. s.379-383

Dragoni, Mauro; Poveda-Villalon, Maria; Jimenez-Ruiz, Ernesto. OWL: Experiences and Directions - Reasoner Evaluation. Springer 2017 (ISBN 978-3-319-54626-1)

Dragoni, Mauro; Poveda-Villalon, Maria; Jimenez-Ruiz, Ernesto. Preface. Lecture Notes in Computer Science 2017; Volume 10161 LNCS. s.I-IX

Forssell, Jon Henrik; Kharlamov, Evgeny; Thorstensen, Evgenij. Towards Characterising Data Exchange Solutions in Open and Closed Words. CEUR Workshop Proceedings 2017; Volume 1912.

Forssell, Jon Henrik; Lupp, Daniel Paul; Skjæveland, Martin G; Thorstensen, Evgenij. Reasonable Macros for Ontology Construction and Maintenance. CEUR Workshop Proceedings 2017; Volume 1879.

Gabmeyer, Sebastian; Johnsen, Einar Broch. Preface. Lecture Notes in Computer Science 2017; Volume 10375 LNCS. s.VII

Gkolfi, Anastasia; Johnsen, Einar Broch; Kristensen, Lars Michael; Yu, Ingrid Chieh. Resource Management of Cloud-Aware Programs using Coloured Petri Nets. I: Proceedings of the 29th Nordic Workshop on Programming Theory. 2017 ISBN 978-952-12-3608-2. s.31-33

Gkolfi, Anastasia; Johnsen, Einar Broch; Yu, Ingrid Chieh; Steffen, Martin; Din, Crystal Chang. Translating Active Objects into Colored Petri Nets for Communication Analysis. Lecture Notes in Computer Science 2017 (10522) s.84-99

Harrow, Ian; Jimenez-Ruiz, Ernesto; Splendiani, Andrea; Romacker, Martin; Woollard, Peter; Markel, Scott; Alam-Faruque, Yasmin; Koch, Martin; Malone, James; Waaler, Arild Torolv Søetorp. Matching Disease and Phenotype Ontologies in the Ontology Alignment Evaluation Initiative. Journal of Biomedical Semantics 2017; Volume 8(1)

Hestvik, Magnus Røed; Mauro, Jacopo; Yu, Ingrid Chieh. CaSPL-gen: a Context-aware Software Product Line benchmark generator. NIK: Norsk Informatikkonferanse 2017

Hohle, Petter; Øvrelid, Lilja; Velldal, Erik. Optimizing a PoS Tagset for Norwegian Dependency Parsing. 21st Nordic Vonference on Computational Linguistics, NoDaLiDa. 2017

Johnsen, Einar Broch; Gabmeyer, Sebastian. Tests and Proofs - 11th International Conference, TAP 2017, Held as Part of STAF 2017, Marburg, Germany, July 19-20, 2017,

Proceedings. Springer 2017 (ISBN 978-3-319-61466-3) 165 s. Lecture Notes in Computer Science (10375)

Johnsen, Einar Broch; Pun, Ka I; Tapia Tarifa, Silvia Lizeth. A formal model of cloud-deployed software and its application to workflow processing. I: SoftCOM 2017. IEEE Communications Society 2017, ISBN 978-953-290-078-1.

Johnsen, Einar Broch; Steffen, Martin; Stumpf, Johanna Beate. A Calculus of Virtually Timed Ambients. Lecture Notes in Computer Science 2017 (10644)

Kaminski, Mark; Cuenca Grau, Bernardo; Kostylev, Egor V.; Motik, Boris; Horrocks, Ian. Foundation of Declarative Data Analysis Using Limit Datalog Programs. 26th Int. Joint Conference on Artificial Intelligence.

Karlsen, Leif Harald; Giese, Martin. An Efficient Representation of General Qualitative Spatial Information using Bintrees. Leibniz International Proceedings in Informatics 2017; Volume 86.

Karlsen, Leif Harald; Giese, Martin. An Efficient Representation of Qualitative Spatial Information using Bintrees. Department of Informatics, University of Oslo 2017 (ISBN 978-82-7368-425-7) 14 s. Conference proceedings (460)

Karlsen, Leif Harald; Giese, Martin. Qualitatively Correct Bintrees With Roles. Department of Informatics, University of Oslo 2017 (ISBN 978-82-7368-432-5) 14 s. Conference proceedings (467)

Kharlamov, Evgeny; Hovland, Dag; Skjæveland, Martin G; Bilidas, Dimitris; Jimenez-Ruiz, Ernesto; Xiao, Guohui; Soylu, Ahmet; Lanti, Davide; Rezk, Martin; Zheleznyakov, Dmitriy; Giese, Martin; Lie, Hallstein; Ioannidis, Yannis; Kotidis, Yannis; Koubarakis, Manolis; Waaler, Arild Torolv Søetorp. Ontology Based Data Access in Statoil. Journal of Web Semantics 2017; Volume 44. s.3-36

Kharlamov, Evgeny; Mailis, Theofilos; Mehdi, Gulnar; Neuenstadt, Christian; Özçep, .zgür Lütfü; Roshchin, Mikhail; Solomakhina, Nina; Soylu, Ahmet; Svingos, Christoforos; Brandt, Sebastian; Giese, Martin; Ioannidis, Yannis; Lamparter, Steffen; Möller, Ralf; Kotidis, Yannis; Waaler, Arild. Semantic access to streaming and static data at Siemens. Journal of Web Semantics 2017; Volume 44. s.54-74

Klungre, Vidar Norstein. A Faceted Search Index for Graph Queries. Universitetet i Oslo 2017

Klungre, Vidar Norstein; Giese, Martin. A Faceted Search Index for OptiqueVQS. CEUR Workshop Proceedings 2017

Laneve, Cosimo; Lienhardt, Michael; Pun, Ka I; Roman-Diez, Guillermo. Time analysis of actor programs. I: Proceedings of the 29th Nordic Workshop on Programming Theory. 2017 ISBN 978-952-12-3608-2. s.13-15

Lembo, Domenico; Rosati, Riccardo; Santarelli, Valerio; Savo, Domenico Fabio; Thorstensen, Evgenij. Mapping Repair in Ontology-based Data Access Evolving Systems. I: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence. 2017 ISBN 978-0-9992411-0-3.

Lin, Jia-Chun; Mauro, Jacopo; Røst, Thomas Brox; Yu, Ingrid Chieh. A model-Based Scalability Optimization Methodology for Cloud Applications. Proceedings of the IEEE 2017

Liu, Tong; Di Cosmo, Roberto; Gabbrielli, Maurizio; Mauro, Jacopo. NightSplitter: A scheduling tool to optimize (sub) group activities. Lecture Notes in Computer Science 2017; Volume 10416 LNCS. s.370-386

Mauro, Jacopo; Nieke, Michael; Seidl, Christoph; Yu, Ingrid Chieh. Anomaly Detection and Explanation in Context-Aware Software Product Lines. Proceedings of the ACM/IEEE Joint Conference on Digital Libraries 2017

Nikolaou, C; Kostylev, E.V.; Konstantinidis, G; Kaminski, M.; Cuenca Grau, B.; Horrocks, I. The bag semantics of ontology-based data access. 26th Int. Joint Conference on Artificial Intelligence. 2017.

Parmiggiani, Elena. This is not a fish: on the scale and politics of infrastructure design studies. Computer Supported Cooperative Work (CSCW) 2017; Volume 26.(1-2) s.205-243

Parmiggiani, Elena; Hibberd, Ralph; Venters, Will. Politics of repair in medicines supply networks: Harnessing fear of falsification. IFIP WG8.6 working conference: "Re-Imagining Diffusion of Information Technology and Systems: Opportunities and Risks"; 2017

Parmiggiani, Elena; Monteiro, Eric; Østerlie, Thomas. Synthetic Situations and Algorithmic Phenomena. 4S Conference; 2017

Pereira, Sunny; Cross, Valerie; Jimenez-Ruiz, Ernesto. On Partitioning for Ontology Alignment. CEUR Workshop Proceedings 2017; Volume 1963.

Pinkel, Christoph; Binnig, Carsten; Jiménez-Ruiz, Ernesto; Kharlamov, Evgeny; Nikolov, Andriy; Schwarte, Andreas; Heupel, Christian; Kraska, Tim. IncMap: A Journey towards Ontology-based Data Integration. The 17th Conference on Database Systems for Business, Technology, and Web. 2017.

Pinkel, Christoph; Binnig, Carsten; Jiménez-Ruiz, Ernesto; Kharlamov, Evgeny; May, Wolfgang; Nikolov, Andriy; Skjæveland, Martin G.; Solimando, Alessandro; Taheriyan, Mohsen; Heupel, Christian; Horrocks, Ian. RODI: Benchmarking Relational-to-Ontology Mapping Generation Quality. Semantic Web Journal. 2017

Ruiz-Calleja, Adolfo; Asensio-Pérez, Juan Ignacio; Vega-Gorgojo, Guillermo; Gómez-Sánchez, Eduardo; Bote-Lorenzo, Miguel Luis; Alario-Hoyos, Carlos. Enriching the Web of Data with educational information using We-Share. International Review of Research in Open and Distance Learning, 247-265. 2017

Skjæveland, Martin G; Forssell, Jon Henrik; Kluwer, Johan Wilhelm; Lupp, Daniel Paul; Thorstensen, Evgenij; Waaler, Arild.

Reasonable Ontology Templates: APIs for OWL. CEUR Workshop Proceedings 2017; Volume 1963.

Solimando, Alessandro; Jimenez-Ruiz, Ernesto; Guerrini, Giovanna. Minimizing Conservativity Violations in Ontology Alignments: Algorithms and Evaluation. Knowledge and Information Systems 2017; Volume 51.(3)

Soylu, Ahmet; Giese, Martin; Jimenez-Ruiz, Ernesto; Kharlamov, Evgeny; Zheleznyakov, Dmitriy; Horrocks, Ian. Ontology-based end-user visual query formulation: Why, what, who, how, and which?. Universal Access in the Information Society 2017; Volume 16.(2) s.435-467

Soylu, Ahmet; Giese, Martin; Schlatte, Rudolf; Jimenez-Ruiz, Ernesto; Kharlamov, Evgeny; Özçep, Özgür; Neuenstadt, Christian; Brandt, Sebastian. Querying industrial stream-temporal data: An ontology-based visual approach. Journal of Ambient Intelligence and Smart Environments 2017; Volume 9(1) s.77-95

Soylu, Ahmet; Kharlamov, Evgeny; Zheleznyakov, Dmitriy; Jimenez-Ruiz, Ernesto; Giese, Martin; Skjæveland, Martin G; Hovland, Dag; Schlatte, Rudolf; Brandt, Sebastian; Lie, Hallstein; Horrocks, Ian. OptiqueVQS: a Visual Query System over Ontologies for Industry. Semantic Web 2017 s.1-28

Tonon, Alberto; Cudré-Mauroux; Blarer, Albert; Lenders, Vincent; Motik, Boris. Arma Tweet: Detecting Events by Semantic Tweet Analysis. In Proc. of the 14th Extended Semantic Web Conference. 2017

Vega-Gorgojo, Guillermo. Clover Quiz: a Mobile Trivia Game Based on DBpedia Data. CEUR Workshop Proceedings 2017; Volume 1963.

Vega-Gorgojo, Guillermo; Giese, Martin. Exploring semantic datasets with RDF Surveyor. CEUR Workshop Proceedings 2017; Volume 1963.

Zheleznyakov, Dmitriy; Kharlamov, Evgeny; Horrocks, Ian. Trust-Sensitive Evolution of DL-Lite Knowledge Bases. AAAI, 1266-1273. 2017

Presentations

Cameron, David. Scalable data access is necessary for successful digitalization. France-Norway Energy Technology Partnership Day; 2017

Cameron, David. Scalable data access: Lower-cost higher-impact environmental compliance. BYTE Project Final Conference; 2017

Cameron, David.; Kluwer, Johan Wilhelm; Berli, Bjørn. Masterclass on Digital Engineering. Subsea Valley Conference; 2017

Cameron, David. Digitalization: How can we mix the "new oil" and the old oil? The role of IT research. V November Conference; 2017

Cameron, David. Digitalization in the Natural Resources Industry: A Challenge and an Opportunity. V November Conference; 2017

Cameron, David. Getting the Dinosaurs to Dance: Innovation culture meets academia and the oil and gas industry. Cutting Edge Festival. Workshop on Transforming Norwegian Industry; 2017

Cameron, David. Successful big data projects. Fra Big Data til Smart Data; 2017

Cameron, David; Waaler, Arild; Hovland, Dag; Skjæveland, Martin G. Practical Knowledge Representation for Data Access to Subsurface Data: The Achievements and Potential of the Optique Platform. AAPG/SEG International Conference and Exhibition; 2017

Giese, Martin. Scalable Access to Big Data. Swedish e-science Research Centre. 2017

Horrocks, Ian; Motik, Boris, Hovland, Dag. Semantic Technology, and in particular on Optique. University of Oxford. 2017

Johnsen, Einar Broch. Cloud computing: Predicting behaviour using models. BigInsight Seminar; 2017

Johnsen, Einar Broch. Model-Based Analysis of Software Systems, with Applications to Planning Processes. Planning workshop; 2017

Johnsen, Einar Broch. Planning and Scheduling at SIRIUS. SAP/SIRIUS introduction; 2017

Annual Accounts

Costs

All figures in 1000 NOK	2015	2016	2017
Personnel and indirect costs	539	5188	10087
Purchase of research services	-	600	2113
Equipment	-	31	122
Other operational costs	62	9505	10424
Total Sum	**601**	**15324**	**22746**

Funding

All figures in 1000 NOK	2015	2016	2017
Research Council	-	4168	8180
University of Oslo	601	2141	5593
Public partners	-	144	813
Private partners	-	8510	7910
Total Sum	**601**	**15324**	**22746**

PhD Profiles

Name: Alessandro Ronca

Strand: Databases

PhD project: Stream Reasoning

Scientific interests, keywords:
Knowledge Representation, Automated Reasoning

What triggers me scientifically: Studying fundamental aspects of computation which can enable new technologies.

Name: Sigurd Kittilsen

Strand: Execution Modelling and Analysis

PhD project: A Formal Framework for Modelling and Analysis of Context-Dependent Adapting Systems

Scientific interests: Modelling, Complexity theory, Logic, Abstraction, Game engines (e.g Chess engines)

What triggers me scientifically: Learning new (scientific) skills. Understanding and solving hard problems as well as creating problems/ puzzles and solutions.

Name: Lars Tveito

Strand: Execution Modelling and Analysis

PhD project: Rapid Verification of Concurrent and Distributed Specifications

Scientific interests: Formal modelling, concurrency, distributed systems, logic, functional programming.

What triggers me scientifically: Having ideas that others seem to have missed.

Name: Summaya Mumtaz

Strand: Data Science

PhD Project: Similarity Analysis for Petroleum Exploration

Scientific Interests: Classification, Machine Learning , Trend Prediction, Uncertainty Estimation

What triggers me scientifically: Predictive analytics is the use of statistics , machine learning and modelling to analyse current and historical facts to make predictions about future events. The power to predict who will click, buy or do fraud is simply fascinating. This is the area of interest for both industry and academics. This drives me to dig deeper in this field and conduct research in this area.

Name: Vidar Norstein Klungre

Strand: Knowledge Representation

PhD project: Improving the usability of ontology-based visual query formulation tools

Scientific interests: Knowledge Representation, Query Interfaces

What triggers me scientifically: Knowing that I am most likely the first in the world to explore the ideas I work on!

Name: Johanna Beate Stumpf

Strand: Execution Modelling and Analysis

PhD project: A calculus for resource management

Scientific interests, keywords: Formal Modelling, Process Calculi, Cloud Computing

What triggers me scientifically: I like solving problems.

Name: Mina Haghshenas

Strand: Work Practices

PhD project: Digitalizing the requirement pipeline in EPC project

Scientific Interest, keywords: Standardization, Digitalization, Change Management

What triggers me scientifically: Contributing in my field of interest by investigating and trying to solve empirical problems

Name: Leif Harald Karlsen

Strand: Knowledge Representation

PhD project: Qualitatively Correct Bintrees: An Efficient Representation of Qualitative Spatial Information

Scientific interests: Knowledge representation, logic, automated reasoning, databases, spatial data

What triggers me scientifically: Attempting to make elegant yet practical solutions to difficult problems
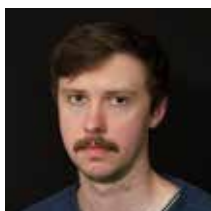
Name: Andreas Thune

Strand: Scalable Computing

PhD project: High-performance numerics and software for parallel reservoir simulation

Scientific interests, keywords: HPC, parallel programming, reservoir simulations, numerical methods and software for PDEs.

What triggers me scientifically: Complex and large-scale problems that require supercomputers to solve.

Name: Daniel P. Lupp

Strand: Knowledge Representation

PhD project: Mappings Remapped: New Perspectives on Data Transformation in Ontology-Based Query Answering

Scientific interests, keywords: data transformation, mappings, rule-based systems, ontology templates, ontology design/maintenance, open/closed-world reasoning

1. What triggers me scientifically: Using languages as an analogy, I am interested in what is lost-in-translation. Various methods exist in translating languages, yet none are 100% capable of conveying precisely the same sentiment in both languages. In a formal structure setting such as data/ontologies, what leeway do we have for adjusting mappings to avoid "misunderstandings"?

Or 2: Triggers, shortened and maybe provocative: Rather than "adding artificial intelligence" (whatever that means) to systems, I am fascinated by how one can remove/avoid natural stupidity inherent to systems and users.

Name: Temitope Ajileye

Strand: Scalable Computing

PhD project: Distributed Reasoning with Parallel Datalog Materialization

Scientific interests: Automated reasoning, knowledge representation and reasoning, graph theory, concurrency

What triggers me scientifically: Making complex phenomena simple

Name: Anastasia Gkolfi (Natasha)

Strand: Execution Modelling and Analysis

PhD project: Model-based analysis of resource-aware programs

Scientific interests: Formal methods, software and hardware verification, modelling techniques, model checking, model-based testing, Petri nets, program analysis by abstract interpretation, reachability problems, algorithms, combinatorics (counting), probabilities, logics for specification and verification, security

What triggers me scientifically: Problem-driven theoretical research which needs to combine different techniques

Name: Shiji Bijo

Strand: Execution Modelling and Analysis

PhD project: Formalization of Data Movement in Multicore Architectures

Scientific interests, keywords: Formal models, parallel architectures, parallel systems

What triggers me scientifically: Intellectual curiosity

Name: Farhad Nooralahzadeh

Strand: Natural Language

PhD project: Domain adapted language technology (information extraction) for oil and gas

Scientific interest: Natural language processing, machine learning, distributional semantics, knowledge extraction

What triggers me scientifically: Researching and exploring the role of intelligent systems in understanding and serving society and humans in an applicable and personalised fashion

Postdoc and Researcher Profiles

Name: Evgenij Thorstensen

Strand: Knowledge Representation

What I do: I work on improving automated reasoning over data in relational databases. A lot of different problems, such as transforming and restructuring relational data, can be viewed as reasoning problems. Likewise, various systems for data integration across data sources can also be seen as systems of logical rules, amenable to the same approach. This research draws on results from logic and logic programming as well as from database theory and applications.

Scientific interests: Database theory; automated reasoning; graph theory; optimization problems; probabilistic inference.

What triggers me scientifically: Understanding the mathematical properties behind clever solutions. I like figuring out why the things that work do so, and discovering new mathematics that can then be applied to create better software.

Name: Jacopo Mauro

Strand: Execution Modelling and Analysis

What I do: I am working in the European project HyVar developing new solutions for the reconfiguration of context and evolvable software systems. By combining optimization, simulation, and formal methods I am contributing in the development of new ways to improve complex system and how they can be deployed on the cloud.

Scientific interests: Optimization, Formal Methods, Cloud Computing, Service-Oriented Computing, Concurrent Languages

What triggers me scientifically: The beauty of programming, formalizing and understanding of complex ideas or system, possibly dominate they complexity or only grasp and be amazed at their difficulty.

Name: Ernesto Jimenez Ruiz

Strand: Knowledge Representation and Data Science

What I do: My research has covered several areas, including bio-medical information processing and integration, ontology reuse, ontology versioning and evolution, ontology alignment, and text mining. I have also designed novel algorithmic techniques and developed java-based tools to serve as a proof of concept of the conducted theoretical research. Currently I am assessing the opportunities and suitability of applying Semantic Web technology to Data Analysis tasks, namely data transformation, data understanding and data cleaning. I would also like to focus on the application of semantic embeddings, making use of RDF data, ontology and rules.

Scientific interests: Semantic Web, ontology alignment, Alignment repair, Evaluation, Semantic Web for Data Science

What triggers me scientifically: Probably the uncertainty. There is not always a clear path from the problem to the solution. Furthermore, I really like my work trying to address new challenges, engaging with the community and collaborating with fellow researchers to solve broader problems.

Name: Crystal Chang Din

Strand: Execution Modelling and Analysis

What I do: I develop software verification systems and use them to prove if a program complies with its requirements. The programs I investigate are concurrent and distributed. I am interested in finding out a modular way for reusing the proof systems such that the effort in verifying communication correctness can be minimised.

Scientific interests: Formal methods, software verification, deductive verification, object orientation, concurrent and distributed systems

What triggers me scientifically: Functional correctness of software is important. How to enhance software quality in the development phase as early as possible and as efficient as possible is my research focus.

Name: Martin G. Skjæveland

Strand: Knowledge Representation

What I do: I am a postdoctor in the Logic and Intelligent Data research group at UiO. The projects I am involved in concern in one way or another ontologies - how to build them, how to use them, how to maintain them. I manage an innovation project on "Ontology Templates", our recent "invention" which we believe will revolutionize the way ontologies are used in information management and be a driver for increased adaption of semantic technologies in large-scale industrial applications.

Scientific interests: Knowledge Representation, Semantic technologies, Ontologies, Ontology engineering, Ontology-based data access

What triggers me scientifically: Elegant and practical solutions with impact: Solutions that solve a difficult problem, that are natural and easy to use, and that make a difference for someone, preferably many.

Name: Elena Parmiggiani

Strand: Work Practices

What I do: Together with my research group in Work Practices at NTNU, I study the interplay between technical and social aspects involved in digitalization efforts in public and private organizations on the empirical level. In so doing, my aim is to contribute to SIRIUS by generating researchable problems at the boundary between academic and industrial aims. I adopt qualitative methods to generate data, including observations, interviews and document analysis. Currently, I am involved with three activities in SIRIUS related to data provenance in oil and gas exploration, hybrid reasoning in interpretation of exploration data, and enhancing data-driven workflows in service outsourcing organizations. In addition, together with the University of Oulu, I am part of a project to study the development of a pan-European digital platform for environmental data management and sharing.

Scientific interests: Digitalization, work practices, oil and gas, environmental monitoring, qualitative research methods

What triggers me scientifically: I am attracted by research problems at the intersection between the academic and the industrial world, especially related to sociotechnical issues of system integration and standardization. I am also fascinated by politically relevant settings, as mirrored by my engagement with the two empirical domains of oil and gas and environmental research.

Name: Violet Ka I Pun

Strand: Execution Modelling and Analysis

What I do: I use formal methods to analyse behaviour and worst execution time of concurrent programs. In addition, I use formal analyses to detect possible data leakage in concurrent programs. I am also interested in formalizing models for multicore architectures with multilevel caches and to analyzing the relationship between these architectures and the performance of software programs. Recently, she has been researching model-based business process planning with tool-supported and automated analyses in terms of formal methods.

Scientific interests: Concurrency programs, type and effect systems, worst execution time, taint analysis, cache coherent multicore architectures

What triggers me scientifically: Curiosity is the major element to drive me to tackle scientific research problems. Also, the process of researching the unknown subject is always fascinating and motivates my scientific thinking.

Name: Dag Hovland

Strand: Knowledge Representation

What I do: I work on several SIRIUS projects, both with health (Bigmed) and with partner companies. My tasks often focus on the area in the intersection between user needs and research. On the one hand is the work of poking into the festering sores that IT problems can become, and connecting this with research questions. On the other hand, modifying, setting up, and customizing research solutions and software into real world situations.

Scientific interests: Ontologies, Relational databases, Query languages, Ontology-based data access, Federation and Equality

What triggers me scientifically: Elegant theories that solve problems people actually have. Improving algorithms for useful programs.

Name: Jiaoyan Chen

Strand: Knowledge Representation and Data Science

What I do: I'm now mainly working on the inter-disciplinary direction of knowledge reasoning and machine learning. The research topic includes (i) how ontology knowledge bases can help explain machine learning approaches, e..g., deep feature transfer; (ii) how knowledge representation and reasoning can boost machine learning in its application in predictive analytics, e.g., air quality forecasting; (iii) how to combine rule-and knowledge-based inductive reasoning and differentable learning for more interpretable, transferable, cost sensitive machine learning and; (iv) how machine learning can be applied in expressive ontology construction and ontology based data access.

Scientific Interests: Semantic Web, Ontology, Knowledge Base, Predictive Analytics, Machine Learning

What triggers me scientifically: Perception, reasoning and knowledge are all necessary to implement artificial intelligence. Description logic is a good solution for knowledge representation and reasoning.

Name: Silvia Lizeth Tapia Tarifa

Strand: Execution Modelling and Analysis

What I do: I am working in the area of formal methods, where I apply and develop techniques for the pre-diction of behaviors in complex systems. Now a days, I am working with formal Modelling and analysis of multicore architectures and memory layout to better understand data movement in parallel processing, trace-based semantics for concurrent and possible non-terminating programs to enable compositional reasoning about the behavior of distributed systems, and formal Modelling concepts for resource usage in distributed systems to analyze resource management and service level agreement.

Name: Jens Otten

Strand: Knowledge Representation

What I do: My research is focused on automating logical reasoning. More specifically, I am developing and implementing tableau- and connection-based proof search calculi for classical and non-classical logics. I have developed the leanCoP series of automated theorem provers, extremely compact yet highly efficient provers for classical and several non-classical logics. More recently I have developed nanoCoP, an efficient connection-based theorem prover that works on a more natural non-clausal input.

I am optimizing and tuning these theorem provers in order to use them to develop and access data in ontology-based databases.

Scientific interests: Automated theorem proving, classical logic, non-classical logics, leanCoP/nanoCoP ontology-based data access

What triggers me scientifically: My objective is to integrate theorem-proving technology into industrial applications. More specifically, I like to develop and deploy the most elegant and efficient theorem provers to automate and optimize ontology-based data access.

Name: Rudolf Schlatte

Strand: Execution Modelling and Analysis

What I do: I am creating models of cyber-physical systems for various projects and case studies. I maintain and develop the Modelling language ABS and its simulator implementation. As SIRIUS lab manager I coordinate researchers' access to machines and software that they need for their work.

Scientific interests: Actor-based languages, distributed systems, executable Modelling, simulation and analysis of operational data.

What triggers me scientifically: Solving problems for my co-workers; getting large amounts of data to analyze and use for input of simulations; tool building and education efforts for partners and students; watching the results of my work getting used by others.

Name: Jia-Chun (Kelly) Lin

Strand: Execution Modelling and Analysis

What I do: I am a postdoctoral researcher at the Department of Informatics, University of Oslo. I work on the HyVar project, especially in the part of Modelling the HyVar toolchain and designing an optimal scalability mechanism for the toolchain. Together with Einar, Ingrid, and Ming-Chang Lee, we developed an ABS-YARN framework and an SSP framework to model and simulate Hadoop YARN and Spark Streaming, respectively. These two frameworks are highly configurable for users to specify different parameters and to predict possible execution performance of applications. In addition, I am also working on a distributed recommendation system, privacy exposure for IoT smart homes, and automatic LSTM-RNN customization for predicting fine-grained traffic speed in large-scale transportation networks.

Scientific interests: Parallel and distributed systems, cloud computing, big data analysis, machine learning, and data mining.

What triggers me scientifically: Doing research is sometimes painful, but I enjoy the process of solving difficult but interesting problems. When I conquer them, the happiness is more than anything.

SIRIUS

ANNUAL REPORT 2017