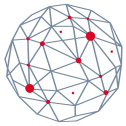
SIRIUS ANNUAL REPORT 2018

Centre for Research-based Innovation
The Research Council of Norway

UiO : University of Oslo

Contents

Summary 5
The Fascination of Software 6
Maturity through Developing Young Talents 8
Spanning Boundaries: Beacon Projects 10
Vision and Objectives 12
Highlights of 2018 13
SIRIUS and the UN Sustainability Goals 16
How SIRIUS is Organized 18
SIRIUS' Partners 20
SIRIUS' Research Programs 24
Analysis of Complex Systems 26
Ontology Engineering 30
Semantic Integration 32
Data Science 35
Profile: Basil Ell 38
Extracting Information from Tables to Populate Knowledge Bases 40
Scalable Computing 42
Industrial Digital Transformation 46
Exploration Beacons 48
Operations Beacons 60
Cross-Domain Beacons 72
International Activity and Dissemination 74
Recruitment, Education and Equal Opportunity 77
Governance 82
Glossary 84
Publications in 2018 86
Presentations in 2018 90
Annual Accounts 90



SIRIUS
Centre for Scalable Data Access in the Oil and Gas Domain

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 three full years of work. 2018 saw the first defences of doctoral projects from centre researchers. We will see more of these in 2019. The beacon projects that were defined in 2017 have been used in 2018 as a tool for writing project proposals and building industry-relevant collaborative projects. This work has resulted in SIRIUS participating in the READI joint industry project. This type of project is a model for our other beacons. This year, we have also identified six research programmes that support the beacon project with research-based tools and methods.

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 an operator (Equinor), service companies (Schlumberger and DNV GL) and IT companies (Computas, Evry, Dolphin Interconnect Solutions, IBM, Kadme, Numascale, OSISoft and SAP). In 2018 TechnipFMC joined as partner, and at the beginning of 2019 we will be joined by Aibel and Aker Solutions. 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 the oil and gas industry. 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 2018.

The Fascination of Software

The Autumn General Assembly of SIRIUS was held in Oslo at the end of October 2018. IBM was host for the meeting, which was held at their new customer centre at Grønland. The Chairman of SIRIUS, Knut Sebastian Tungland, Senior Advisor in Equinor, held the following speech.

Thanks to IBM for hosting this event.

I will not take much time, and I will not focus much on the inner workings of SIRIUS: we will focus on that the rest of the day. I would like us to reflect on how lucky we are.

I am a technology optimist. I believe technology has improved life for humanity and will continue to.

I recently read a book by Hans Rosling, called "Factfulness". Even when the world seems to be getting worse, it's becoming a better place for most of humanity, partly because of technology.

That does not mean we do not have challenges. Climate, overpopulation, and not least how we behave toward each other, in the estrangement in societies, groups within countries that don't respect each other, or discuss only through social media, and the growing tension between nations and geographies.

Still, I believe technology is part of the solution for meeting and conquering these challenges.

For me, the biggest fascination is software. I think myself lucky to live in its infancy. I read an article about Grace Hopper, one of those women who do not get the recognition they deserve. In the fifties she created the first compiler: a fundamental paradigm for how to exploit computers. The building in which SIRIUS resides is named after a Norwegian software pioneer, Ole-Johan Dahl. This is not long ago. Software is evolving, and we have only seen its beginning.

We are discovering new algorithms and new ways of expressing logic and relationships all the time. Sometimes I think this is how David Livingstone must have felt: there is

equinor

Equinor (formerly Statoil) has a leading role in SIRIUS. As the major operating company on the Norwegian Continental Shelf, they set the agenda for digital transformation in the sector. Equinor contributes to the exploration and operation beacon projects, with a leading role in the Digital Field and Reservoir Management, Digital Field Development and Integrated Digital Planning beacons.

Equinor works with SIRIUS to engage technology providers and academics who work together to innovate and solve challenges in the energy industry. SIRIUS provides an open forum that allows transfer of technology and skills into the company.

Photograph: Equinor-Johan Sverdrup

so much land and so little time to discover and understand it all. I have a hard time prioritizing what to focus on. The number of articles published each day on different software related topics and the number of software communities are overwhelming. I discover new and interesting groups every week. Additional to this is the growing ecosystem of startups, incubators and venture capitalists. Finally, software development is a growing field in large companies like mine. In Equinor, last time I counted, we had over 200 software projects in the research organization alone. This is not counting projects in the IT organization or development done for the operating units.

I also find the relationship, and maybe artificial separation, between data and software fascinating. Cloud, Data Science, Machine Learning and Artificial intelligence have become everyday words for people with no understanding of or focus on the underlying technology. For better or worse, they have great expectations, maybe more than the experts. This is only the beginning. As Marc Andreesen said, "Software is eating the world". We get to live in this world.

What we are doing is hard work. It is intellectually challenging, it takes time, one has to relate to people. We develop in communities. Sometimes people will not listen, sometimes you will not understand. We will sometimes fail. There are some many roads to take in this landscape, and sometimes we will take the wrong way. But it is fascinating.

We get to be part of an innovation center where we can address real challenges with technology. Do we make the most of it? Can we do it better? That is what we are here for. But, also, let us enjoy the journey, being part in this exploration.

Knut Sebastian Tungland

Knut Sebastian Tungland is a Senior Advisor in Equinor. 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 Equinor.

Maturity through Developing Young Talents

Three years into SIRIUS' journey, we are maturing as a research centre and are better equipped than ever to face the coming phases. We have a team of highly inspired and motivated people who excel in their fields and we have built a common platform and intersecting areas of interest that are relevant and interesting for both the research community and industry.

SIRIUS is built upon knowledge that comes from research. We are a centre precisely because our partners want to further develop this knowledge together with us and bring new ideas, products and results into their companies. To achieve this, we must continuously strive to stay at the research frontiers and we must constantly challenge established perspectives in research communities and within the industry.

SIRIUS is a complex undertaking. We consistently endeavour to reinforce a shared vision between participants with different interests and expectations. This is what it means to be a centre for research-based innovation. We must constantly realign with the centre's stakeholders as research advances and the business needs of our partners evolve.

To manage this complexity of collaboration we need a centre structure that is both flexible and stable. It must be flexible to accommodate new ideas and projects, while being stable enough to maintain consistency and enable clear communication of our vision and results.

Since the centre's kick-off in May 2016, a clearer structure for SIRIUS has gradually emerged. In 2017, the concept of a business beacon was coined as a way of organising strategic application areas and partner collaboration. In 2018, the concept of research programs has been introduced, renewing and replacing the research strands that we had when the centre first started. These research programs focus on the technology components and research-based methods that we need to deliver results for the business beacons. These technology components and methods are also innovative products in their own right.

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 ontology engineering and semantic integration. After finishing a civ.ing. degree in 1989, Waaler graduated as dr. philos. in Philosophy in 1995 within Philosophical logic. He then had 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 Department. of Informatics at University of Oslo in 2006, when he started to build up the collaboration that eventually turned into SIRIUS.

> **We believe** that this will bolster our capacity to meet new challenges and generate new opportunities as the centre now enters a more mature operational stage.

As we enter 2019 and face a mid-term evaluation, we are taking the opportunity to hand responsibility for the centre's success to a new generation of younger researchers. A new team of young leaders for the research programs is being mobilised, and Ingrid Chieh Yu now serves as SIRIUS` deputy director. Our ambition is to further strengthen our internal cooperation and "centre spirit", reinforce open, interdisciplinary communication and foster creative and cohesive teams. The research programs establish new networks and channels for efficient information flow. They also provide a tighter focus on innovation and delivery that complements the research content of the programs. We believe that this will bolster our capacity to meet new challenges and generate new opportunities as the centre now enters a more mature operational stage.

The new structure of the centre is described on page 18, while the content and progress of each research strand is given starting on page 24.

UiO : University of Oslo

SIRIUS is hosted by the University of Oslo's Department of Informatics. Our work is also tightly coordinated with the Faculty of Mathematics and Natural Sciences. The University provides funding for the centre and provides facilities on the 8th Floor of Ole Johan Dahl's building in Oslo. The Faculty of Natural Mathematics and Natural Sciences has four thematic initiatives: (i) life sciences, (ii) earth and space, (iii) energy and materials and (iv) digitalization and computational science. SIRIUS is a core element in the university's digitalization initiative and also contributes in the other three thematic areas.

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 Det Norske Veritas as a Senior Researcher in DNV Research and Innovation, and returned to the University as Associate Professor in 2011. She is the the Mentoring and Education Coordinator of SIRIUS and from 2019 will be the Deputy Director of SIRIUS. She was one of the Principal investigators of CUMULUS, a NFR funded research project and the site leader of the EU H2020 HyVAR project, which addressed continuous context dependent software evolution of telematics and on-board systems in cars. She is leading the activities in designing a scalable cloud infrastructure for monitoring and individually customizing software upgrades for remote devices.

Spanning Boundaries: Beacon Projects

SIRIUS' 2017 Annual Report introduced the idea of beacon projects as a way of organizing the innovation work in the centre. Our idea was to identify broad but distinct areas where scalable data access could benefit industry and the wider society. Our scoping work in 2016 and 2017 showed that there were many ideas for how and where SIRIUS could work. We sorted and prioritized these ideas to find those that both had the best links to our research and skills and solved our partners' industrial problems. The beacons initiated in 2017 have all matured in 2018. In some cases, we have chosen to change the name of the beacon so that it gives a better description of what we are doing.

From 2019 SIRIUS has eight beacon projects:

1. **Geological Assistant** (formerly Digital Support for Exploration Processes). We are working on providing a geological assistant that will assist a geologist with workflows and decisions. See page 48 for more details. We have built an interdisciplinary team of computer scientists and geoscientist to work on this. This team participated in a valuable field trip to the Pyrenees, as described on page 50.

2. **Subsurface Data Access & Analytics** (formerly Access to Exploration Data). SIRIUS is building on the Optique platform for ontology-based data access to demonstrate how national data repositories like DISKOS and data sets like Equinor's Volve data set can be developed into platforms for exploration, research and innovation. We are also working on analytics of subsurface data, in the first instance with an innovation project with IBM and Equinor on image recognition. See page 52 for more details.

3. **Digital Field & Reservoir Management.** In 2017 this beacon consisted solely of a collaborative project that focused on improving the performance of open reservoir simulators on high-performance computers. This work has continued in 2018. In addition, we have also seen that data science and digital twins are of interest in field management and petroleum technology. We have therefore worked with Equinor, Petrobras and the Federal University of Rio Grande do Sul (Brazil) to define an ambitious program that will define research-based best practices for implementing digital twins for field management. See page 55 for more details.

Our beacons apply computer science to pressing business problems

4. **Integrated Digital Planning.** We are continuing our work with Equinor on the application of formal simulation tools for planning of projects, commissioning, maintenance and logistics. This beacon has been strengthened by the involvement of two new partners in SIRIUS: SAP and TechnipFMC. See page 60 for more details.

5. **Digital Twins.** The digital twin concept reached the top of Gartner Group's well-known hype curve in 2018. "Everybody" is building a digital twin of their process, plant or product. Building sustainable and usable digital twins requires high-quality data science and knowledge representation. The alternative is an increased siloing of data. SIRIUS is working to implement a research program that will support the implementation of scalable digital twins. See page 64 for more information.

6. **Digital Field Development** (formerly Digital Thread). Much time and effort are devoted to tracking of a design and its requirements through the lifetime of a product, component or system. Preparatory work in 2017 has resulted in the commissioning of a large Joint Industry Project called READI, that is looking at demonstrating how requirements can be digitalized to reduce the cost and improve the performance of engineering, procurement and construction. See page 68 for details of this JIP.

7. **Personalized Medicine.** SIRIUS' work in the BigMed lighthouse project continued in 2018. We are using this work to launch further projects that take SIRIUS methods and tools and apply them in a medical context. See page 72 for more.

8. **Environmental Applications** (formerly Arctic Earth Observation). We are continuing our work with the CIRFA centre in Tromsø to incorporate scalable data access and cloud computing in their research and pilot systems. We are also working with NIVA on the application of semantic technologies and data science in the analysis of environmental modelling data. See page 73 for more.

SIRIUS' beacon projects span the boundaries between computer science academics, information technology vendors and applications in oil & gas, medicine and environmental applications. They are designed to take the results from SIRIUS' research programs and show how they can create innovation and changed business models in our partner companies.

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.

Vision and Objectives

Vision	
To accelerate the development and adoption of innovative data access technology in the oil & 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 a 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. 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. Allow scalable processing and storage of big volumes of data. Process real-time streams of sensor data. Exploit affordable hardware platforms.
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 2018

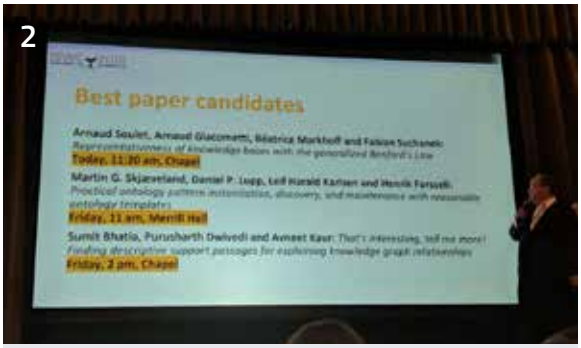
1 Workshop in Porto Alegre

2018 has been a year where SIRIUS has built a fruitful collaboration with a leading Brazilian group that specializes in petroleum informatics. A delegation from SIRIUS spent a week in Porto Alegre, holding a workshop on Digital Geosciences for researchers and master's students at the Federal University of Rio Grande do Sul. We have also had a visit from Professor Mara Abel, the head of this group, and one of her PhD students is spending a year here in Oslo. Read more about this on page 57.

Workshop participants at UFRGS in November 2018.

2 International Semantic Web and Description Logics Conferences

Delegations of SIRIUS researchers attended the International Semantic Web Conference (ISWC) in Monterey, CA, USA and the Description Logics Workshop in Tempe, AX, USA. Both meetings were in October 2018. Our workers presented well-received papers. SIRIUS was a finalist for best research paper at ISWC (Practical ontology pattern instantiation, discovery, and maintenance with reasonable ontology templates by Martin G. Skjæveland, Daniel P. Lupp, Leif Harald Karlsen and Henrik Forssell). We won best student paper at the Description Logics workshop (Generating Ontologies from Templates: A Rule-Based Approach for Capturing Regularity by Christian Kindermann, Daniel P. Lupp, Uli Sattler, and Evgenij Thorstensen).



Announcement of the best paper candidates at ISWC.

Torbjørn Røe Isaksen, Norway's minister for Trade and Industry, visited the stand for SIRIUS and UiO Energy at Subsea Valley. Here together with Kristin Vinje (left), David Cameron (2nd from right) and Vebjørn Bakken (right).

3 Closer Collaboration with Subsea Valley

During 2018, SIRIUS, together with UiO Energy, has built up collaboration with the Subsea Valley NCE. We are active participants in Subsea Valley's digitalization and data sharing initiatives and exhibited at the 2018 conference, held on March 21st, in the Digital Village.

4 Field Trip to Spain: Geological Reasoning in the Wild

In October, SIRIUS researchers together with representatives from Schlumberger organized a three-day workshop for the Geological Assistant project in Aínsa, Spain. This was held together with the University of Oslo's annual field work for master's students in Geosciences. Read more about this on page 50.

5 Presentation at UIIN Conference in London

The University Industry Innovation Network is an organization that brings together innovation practitioners from around the world to discuss how to build better relations between industry and universities. SIRIUS has been active in UIIN in 2018, participating in a workshop on building industrial relationships and presenting a paper at the UIIN Conference in June.

6 First Mentoring Program Concluded

The first round of the SIRIUS mentoring program was completed in April 2018 with a closing seminar. This program brought together SIRIUS researchers and mentors from industry to build career awareness and personal development. The program was successful and will be repeated in 2019-20. See page 77 for more.

A full-packed hall at the UIIN Conference in London.

Jacopo Mauro presenting at the Mentoring Program. Jacopo left SIRIUS in mid 2018 to become an Associate Professor at the University of Southern Denmark, Odense.

Ainsa, Spain

Eric Monteiro, Mara Abel and Adnan Latif in the panel discussion at the November Conference.

7 Digital Twins Workshop

SIRIUS held a one-day workshop on Digital Twins in Oslo on 20th March 2018. This brought together participants from Equinor, Petrobras, Schlumberger, Halliburton, DNV GL, Computas, OSISoft, Aker Solutions, Aibel, IBM and SAP. Participants shared their ideas of what a digital twin is and should be. The results of this workshop were published at the Scandinavian Simulation Conference in September. Read more about this on page 64.

8 Briefing on Digitalization for Petroleum Industry

Digitalization continued to be a much-discussed issue in the oil & gas industry. DNV GL and SIRIUS held a briefing for oil & gas executives on 25th May 2018. This brought together management from operator and engineering companies to hear about SIRIUS and share thoughts about the common challenges facing the industry and how to address them. The keynote addresses were given by the chairman of Digital Norway, Walter Qvam and senior management in DNV GL.

9 November Conference in Rio de Janeiro

SIRIUS was once again very active in the annual November Conference on Norwegian-Brazilian Energy Research, held on November 11-12 in Rio de Janeiro. David Cameron was chairman of the program committee and SIRIUS-related workers Mara Abel, Adnan Latif and Eric Monteiro held presentations in the Digitalization and Automation Session.

10 Visiting Fellows at Alan Turing Institute

SIRIUS' collaboration with Alan Turing Institute, the United Kingdom's national research centre for Data Science deepened. Two SIRIUS researchers, Ernesto Jimenez-Ruiz and Evgenij Thorstensen were visiting fellows at the ATI in 2018. You can read more about this on page 75.

11 PhD Presentation at OG21 Forum

The OG21 Forum was held in Oslo on 29th November 2018. This annual event presents the results of the oil and gas research and innovation programs in Norway, as defined and sponsored by the OG21 programme. It brings together the research and technology leaders of the entire Norwegian oil and gas industry. This year, each research university in Norway was invited to propose a PhD candidate to do a short presentation of their work. The University of Oslo's candidate was Daniel Bakkelund from SIRIUS. He presented his work on using machine learning to fill in missing data in equipment databases.

Daniel Bakkelund presenting at the OG21 Conference.

SIRIUS and the UN Sustainability Goals

The University of Oslo's strategy is that its research, teaching and innovation supports the United Nations' Sustainability Goals. SIRIUS shares this ambition, even though a superficial analysis could assume that our focus on oil and gas works against this. This section goes through the different goals, stating how SIRIUS' work can help and what we are doing to achieve these goals.

Personalized healthcare enabled by information technology can improve patient safety through better combination of clinical data and patient records.

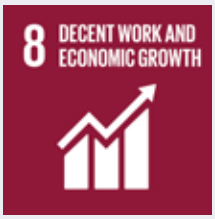
Laborious data retrieval and review activities can be automated and simplified, and health professionals can have easier access to literature and unstructured data. SIRIUS is addressing this goal through the BigMed beacon project for personalized health care, research related to search in medical linked data and friendlier user interfaces to medical information systems. Semantic data integration has been demonstrated as a tool that allows automatic filling out of pre-operative forms. Finally, we are working on applying formal methods for better hospital planning. See Personalized Medicine, on page 72, for more information.

Digitalization is generating needs for general and specialist training in how to use and interpret data. This means that there is a need for scalable data access tools in education. Our specific focus here is engineering and geosciences. Once again, we can automate and simplify data retrieval and review activities and provide easier access to literature and unstructured data.

SIRIUS is contributing to quality education through its Digital Geosciences initiative, in collaboration with UFRGS in Brazil. We are also working in the University of Oslo's research school for informatics undergraduates.

Effective design, construction and deployment of water treatment and sanitation systems is dependent on good data access. The internet of things provides cheaper and better ways of controlling and optimizing water systems. This requires methods for secure and effective implementation. SIRIUS is supporting this goal through work with NIVA on a PhD project on data science applications. Our work on digital twins and requirements are directly applicable to capital-intensive water systems.

Scalable data access can reduce the capital cost of new energy systems and infrastructure. It can lead to better operations, lower energy usage and a lower carbon footprint. Good data access can also lead to more effective planning of maintenance and logistics for complex, remote and offshore facilities. SIRIUS addresses this through its operations beacon projects, that address data challenges in engineering requirements, digital twins and integrated planning. Experience from oil and gas is directly transferrable to renewable energy.



Scalable data access can automate tedious and unproductive interfacing and data transfer operations in supply chains. This allows more effective and rewarding work practices and frees workers to creatively optimize their business. SIRIUS is working to support this in several ways. Ontology-based data access provides a framework for automation of tedious data retrieval and management tasks. Digital requirements management reduces documentation burden in engineering. Our geological assistant empowers geologists with new and useful tools. Better data access allows companies to move employees out of harm's way and improve the safety of facilities.

Scalable data access contributes to the effectiveness and performance of infrastructure projects. It provides a foundation for new products and services in industrial data management and data science. It also enables the implementation of open platforms for data-driven innovation.

SIRIUS is contributing to this through its laboratories for industrial data access and its operations beacons that address the digitalization of requirements, engineering and industrial operations.

A smart city is a complex system of sensors and processes, just like a large industrial plant or an oil platform. Methods developed and applied in industrial IT and automation can be applied fruitfully to develop smart cities. SIRIUS's, work on digital twins and ontology-based data access can be used to build robust and scalable systems for managing smart cities.

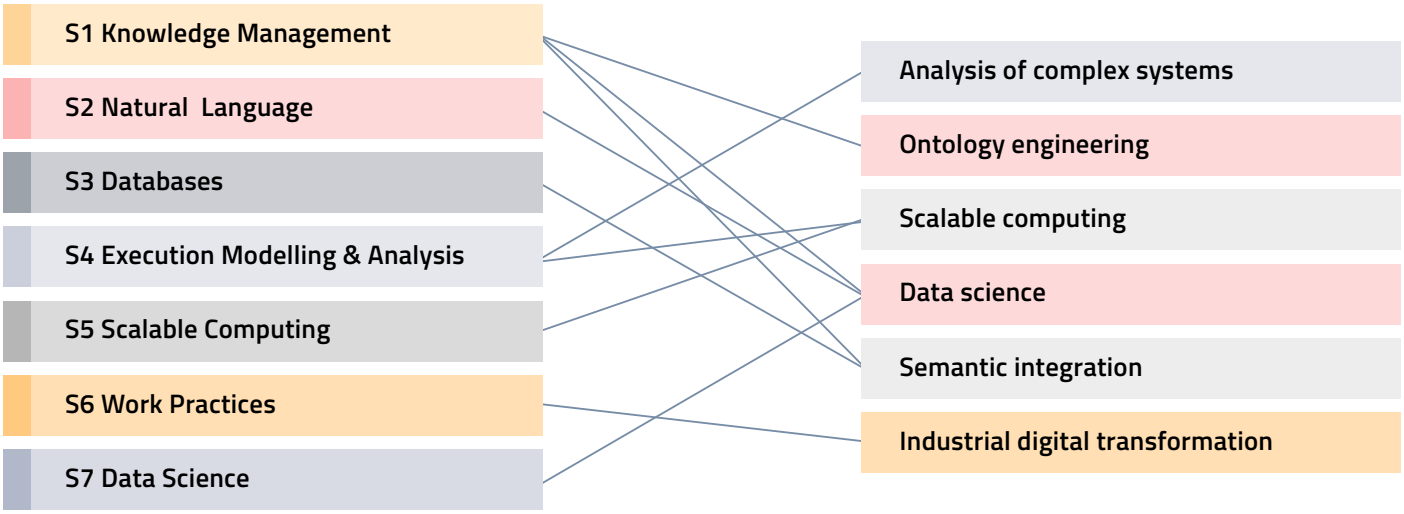
Scalable data access allows systems to be developed that actively reduce energy use and carbon footprint in resource industries. SIRIUS' methods and tools reduce the cost and improve the performance of applications that raise operational performance. SIRIUS is working with the CIRFA centre in Tromsø on applying its methods for semantic data access and cloud computing to manage earth observation and climate data. See Environmental Applications on page 73.

Scalable data access will improve the handling of data for oceanic research and monitoring, providing better information on the status of the marine environment and operations in that environment. SIRIUS' researchers have a background from work practices in submarine environmental monitoring projects. Our digital twin and digital requirements research can be applied in maritime operations. Complex data systems for oceanic research and monitoring can benefit from ontology-based data access and user interfaces.

SIRIUS is internationally oriented. It collaborates in training and educational partnerships across international boundaries. We are building a promising collaboration with universities and industry in Brazil. We are also participating in relevant European private-public partnerships (BDVA and SPIRE) to ensure that we are working together to solve problems.

How SIRIUS is Organized

SIRIUS' projects are organized into either business-related work packages or technically oriented research programs. Each researcher works in a research program and may be involved in one or more beacon projects. Strands, as described in previous reports, have been replaced this year by research programs, as this gives a greater emphasis on deliverable knowledge and tools. The relationship between strands and research programs is shown in the following figure:



This matrix organization enables us to obtain the right balance of fundamental research and focus in innovation and industrial applications. These research programs then provide a foundation of methods and tools that can be prototyped and piloted in the beacon projects.

This structure is shown in the following figure, where the projects are organized into Work Packages.

WP1 Exploration	WP2 Operations	WP3 Cross-domain applications
Geological assistant	*Integrated digital planning*	*Personalized medicine*
Subsurface data access & analytics	*Digital twins*	*Environmental applications*
Digital field & reservoir management	*Digital field development*	

WP4 Research Programmes					
Analysis of complex systems	*Ontology engineering*	*Scalable computing*	*Data science*	*Semantic integration*	*Industrial digital transformation*

WP5 Strategy and Outreach

SIRIUS has five work packages:

- Exploration: beacon projects in exploration and subsurface applications. See page 48.
- Operations: beacon projects in the design, construction, operation, maintenance and decommissioning of complex industrial facilities. See page 60.
- Cross-domain Applications: applications in areas outside and beyond the natural resources industries, notably health and environmental applications. See page 72.
- Research Programs: SIRIUS' computer science results that can be applied in projects in several of the business work packages. See page 24. It is here that we achieve excellence in computer science research.
- Strategy and Outreach: projects that define the direction of SIRIUS' research and innovation, including education, equal opportunity and dissemination.

Each work package has a manager, with responsibility for the portfolio of beacon projects in that work package. Each beacon has a leader, with responsibility for the projects in that beacon. Each research program has a leader, who has responsibility for the technical quality of the program and for the researchers who work in it.

SIRIUS has fourteen 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 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.

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 fellows 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 is host for 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, Simula Research Centre participates with work on high-performance computing and scalable computing.

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, specialized software suppliers and researchers.

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 TechnipFMC joining the centre in August 2018 and in Aker Solutions and Aibel joining in 2019. Negotiations are continuing with several leading operating, service and software companies.

Fluid Operations AG was acquired by Veritas Inc. in 2018. This purchase involved a change in strategy for the Fluid Operations business. They therefore withdrew from the centre from the end of 2018.

No single partner is interested in all the beacon projects, although a large operator like Equinor could have a role in all except Personalized Medicine. This is the case. The following table shows the current involvement of partners in the beacons.

	Aibel	Aker Solutions	Computas	DNV GL	Dolphin	Equinor	Evry	IBM	Kadme	Numascale	OSIsoft	SAP	Schlumberger	TechnipFMC
Geological Assistant						■							■	
Subsurface Data Access & Analytics					■	■	■	■	■	■			■	
Digital Field & Reservoir Management						■					■			
Integrated Digital Planning					■	■				■		■		■
Digital Twins	■	■	■	■	■	■				■	■	■		■
Digital Field Development	■	■	■	■	■	■				■				■
Personalized Medicine			■	■			■	■				■		
Environmental Applications						■								

Spotlight on TechnipFMC

TechnipFMC joined SIRIUS from August 2018. In conjunction with the signing of the agreement, Kristin Vinje, Vice-Dean for Innovation at the Faculty of Mathematics and Natural Sciences, University of Oslo, wrote a blog post in her blog on Titan (https://titan.uio.no/kristinvinje). A translation of this article from Norwegian is given here.

On Tuesday 7th August 2018, the University of Oslo and TechnipFMC signed a cooperation agreement which brings another heavy industrial player into the major research project SIRIUS at the University of Oslo. TechnipFMC is Norway's largest oil service company. In Norway, Ann-Christin Andersen has been given global responsibility for the company's commitment to digital transformation. Now she wants to enter into collaboration with SIRIUS to secure access to the cutting edge of research in handling large amounts of data.

Signing of the agreement at TechnipFMC. From left to right: Knut Johannesen Vidvei (TechnipFMC), Kristin Vinje (University of Oslo), Morten Dæhlen (University of Oslo), Ann-Christin Andersen (TechnipFMC), Elisabeth Nøst (TechnipFMC) and David Cameron (SIRIUS).

TechnipFMC has been a leading player in the oil and gas industry for over 40 years, with origins in Kongsberg Subsea. Among other things, they have participated in many large field development projects in Norway, including Åsgard, Ekofisk, Snorre and Statfjord.

Now they want to cooperate with the University of Oslo on the development of digital solutions for the oil and gas industry. This is about being able to release capacity and interact better in an industry where technological development is very fast. They want to be at the forefront of technology development in order to be able to develop smart, digital solutions that improve the utilization of the enormous amounts of data generated in the oil and gas sector.

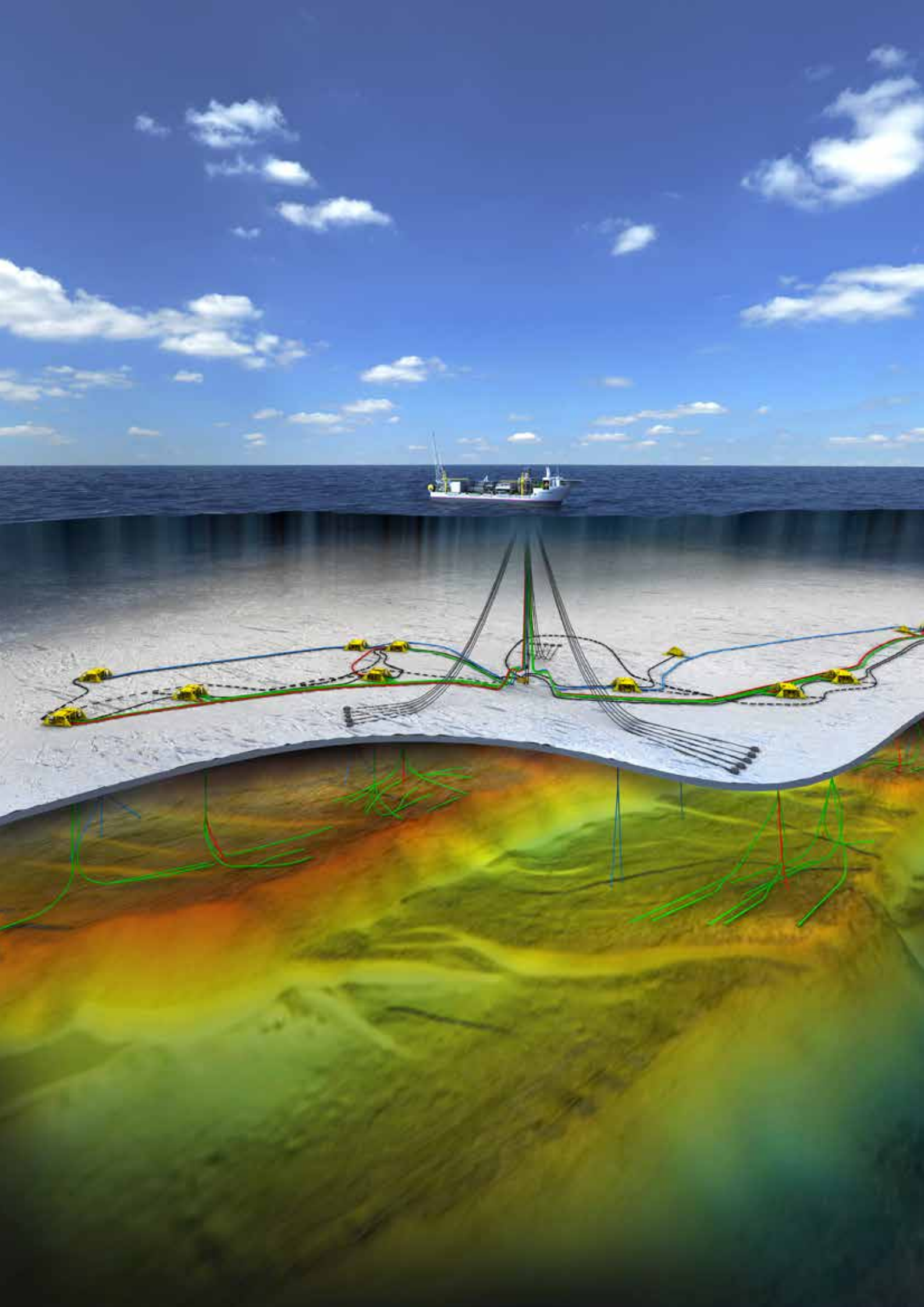
The knowledge generated in SIRIUS is generic and will benefit other fields of work and industry segments, such as health, production and public sector activities. But the research-based innovation must first and foremost contribute to the participants in the SIRIUS project. SIRIUS has attached 15 doctoral students and publishes scientific articles and articles in international research journals.

The close cooperation with the business community is crucial for the success of the SIRIUS centre. Through this agreement with TechnipFMC, we gain significant access to knowledge, software and smart solutions. In a consortium that now consists of 13 industrial partners and several heavy academic environments, in addition to the University of Oslo, we hope to find good solutions faster by pulling in the same direction. We will utilize each other's expertise and experience and take turns on the best commercial solutions.

For us at the University of Oslo, it is about us that we want the knowledge we develop to be used. We want to benefit society. We will contribute to value creation. Through such collaborative projects, we believe that our good research environments will be even better because we learn from the experience that the business community has so that we can illuminate new issues in new ways.

TechnipFMC

TechnipFMC is Norway's largest oil service company, with over 37,000 employees worldwide and around 3,000 employees in Norway. They are a leading supplier of equipment and sustems for subsea production and processing of oil & gas. They are a world-leading engineering and oilfield services company. Its engineering business builds liquefied natural gas, gas treatment, petrochemical and mineral processing plants. SIRIUS is working with TechnipFMC's, centre for digital solutions and various business areas, with focus on the operations work package: integrated digital planning, digital twins and digital field development.



SIRIUS' Research Programs

Ian Horrocks

Ian Horrocks, FRS is professor in the Oxford University Department of Computer Science, a visiting professor in the Department of Informatics at the University of Oslo and Scientific Director of the SIRIUS Centre. His research interests include logic-based knowledge representation and reasoning and semantic technologies, with a particular focus on ontology languages and applications. He was an author of the OIL, DAML+OIL, and OWL ontology language standards, chaired the W3C working group that standardised OWL 2, and developed many of the algorithms, optimisation techniques and reasoning systems that underpin OWL applications. He has participated in numerous national and international research projects, and was Scientific Director of the FP7 Optique project, which deployed semantic technologies in the Oil & Gas and Power Generation industries. He is a Fellow of the Royal Society, a member of Academia Europaea, a Fellow of the European Association for Artificial Intelligence and a Fellow of the British Computer Society. He has published more than 200 papers in major international conferences and journals, winning best paper prizes at KR'98, AAAI-2010, and IJCAI-2017, and a test of time award at ISWC-2013. He is one of the UK's most highly cited computer scientists, with more than 48,000 citations, and an h-index of 91.

Analysis of Complex Systems

Complex systems are usually very difficult to understand and analyse because they are characterised by many interdependent tasks or activities happening at the same time. In the oil and gas domain, we can observe such systems in e.g., diverse and interdependent industrial and geological processes and in supporting technologies such e.g., software applications, big networks of heterogeneous robots and sensors, parallel super computers accessing data. For some of these systems it is useful to capture such behaviours and interactions in order to find errors, redesign, extend, and improve them, while for others a much clearer picture of such behaviours and interactions can facilitate other kinds of analysis. One way to understand and analyse such systems is by using formal methods. Formal methods are mathematical based techniques to abstractly represent or model the behaviour and interaction of systems. This program develops and applies techniques that support the representation and analysis of correct behaviours and interactions of complex systems, together with analysis related to time-dependent systems, resource management, constraint solving, optimisation and prediction.

Ontology Engineering

The digital transformation of the industry depends on rich information models that are intelligible to both computers and humans. Such a model should represent how domain experts view their domain in order to enable them to view and explore the data they require. Constructing, maintaining, and using such models is far from straightforward, however. The aim of this research program is to provide tools and methods to domain experts, information modellers, and ontology experts to improve the efficiency and quality of ontology development, maintenance, and use.

Semantic Integration

Data within the oil and gas domain typically resides in several different sources and can have vastly different forms and access methods. In order to ensure optimal decision making all of this data must be taken into account; an end-user needs to be able to view and understand all data. Accessing the data in their legacy format requires in-depth, low-level knowledge of how the data is stored, which is a considerable challenge for end-users. By integrating all data under a common ontology, users can

Our research programs work together to provide solutions for oil & gas data challenges.

view and explore the data in a language they understand. This research program aims at addressing issues that come up during this integration process, in particular by designing and developing scalable infrastructure for the integration of multiple large datasets and large-scale ontologies.

Data Science

Industry-relevant data comes in many forms, from structured sources (e.g., databases) and unstructured sources (e.g., natural language documents intended to be read by humans). Having access to all this data is only as useful as the methods one has for evaluating and using this data to make decisions. This research program employs natural language processing, machine learning, and statistics in order to extract as much information as possible from both unstructured and structured data sources. It will in particular focus on developing novel approaches for extracting information from data while taking into account its structure and semantics.

Scalable Computing

In order to ensure that industry gains a long-term benefit from the tools and methods developed in SIRIUS one must ensure the scalability of these solutions. This research program focusses on research in high-performance computing coupled with scalable cloud computing to support scalable big-data application processing.

Industrial Digital Transformation

When adopting new digital technologies, the central challenge industrial companies face is to identify and cultivate the organizational pre-conditions necessary for realizing the potential of these new technologies. Consequently, the realization of their potential falls significantly short of expectations. This research program focuses on providing a set of rules and principles to guide the planning and implementation of digital technologies .

Eric Monteiro

Eric Monteiro is professor of Information Systems at the Norwegian University of Science and Technology (NTNU) and adjunct professor at the University of Oslo. He is Pilot Strategy Coordinator in SIRIUS. His research interest focuses on the transformation of work practices accompanying digitalization efforts in organizations. This involves identifying the organizational and technical conditions for realizing the potential of digitalization in practice. He has been project leader for a portfolio of research projects into digitalization of oil and gas in close collaboration with industry. He is part of the management group on BRU21, a large-scale, fully industry-funded national initiative on digitalization and automation in oil and gas. He has a robust academic publication track record (citations 6305/210, h-index 27/8, with Google Scholar/ISI).

Analysis of Complex Systems

The oil & 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 with 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 can be run in communicating but 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 a challenging: many possible settings and parameters can be tuned. Poor choices can result in system failure, high costs and displeased customers.

> **Our expertise** lies in modelling complex parallel and distributed systems, including object-oriented and service-oriented systems, cloud computing and the Internet of Things.

The ADAPT project started in 2018, with Lizeth Tapia Tarifa as leader and FRINATEK funding from the Research Council. 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 work-flows 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 SIRIUS' current toolbox on model-based prediction.

Spotlight on ABS: Model-Based Predictions for Complex Parallel Systems

Our aim is to predict the behaviour of complex systems using the analysis of models. Decisions can then be made based on these analyses. Our expertise lies in modelling complex parallel and distributed systems, including object-oriented and service-oriented systems, cloud computing and the Internet of Things. Analysis techniques for these models range from simulation, which analyses a single run of a system, to deductive verification, which analyses all possible runs of a system. We analyse both functional and non-functional properties such as safety properties, timing properties, resource management and scaling strategies. Our work on resource-restricted parallel systems with timing constraints is currently being applied to planning and logistics in the context of SIRIUS.

We bring background from two, successfully completed, European projects: Envisage and HyVar. These have been described in detail in earlier reports. This background is unified through the ABS modelling language and analysis framework. Each of these projects provides important tools and experience that can be used in the SIRIUS experiments and pilots. Today, we also bring solutions developed for planning and logistics in SIRIUS, as well as needs identified in this domain, back into the overall development of ABS.

ABS incorporates experience from SIRIUS' planning beacon.

The ABS Modelling Language (http://www.abs-models.org)

ABS is a language for Abstract Behavioral Specification of distributed and concurrent systems, which supports the modelling of resource restrictions and resource management. It combines implementation-level specifications with verifiability, high-level design with executability, and formal semantics with practical usability. It is a concurrent, object-oriented, modelling language that features functional data-types. It is designed to develop executable models with a parallel, object-oriented program flow. ABS targets distributed and concurrent systems by means of concurrent object groups and asynchronous method calls. It supports model variability based on feature models and delta-oriented specifications. Deployment modelling can be based on high-level deployment models. The ABS system supports the modelling of resource-aware and resource-restricted systems and provides a range of techniques for model exploration and analysis, based on formal semantics.

Finally, ABS is an open source research project.
You can find the project on GitHub at
https://github.com/abstools/abstools.

From Project-driven to Community-driven Research

We are currently moving ABS from a project-driven to a community-driven technology. To this end, the first international ABS workshop was held in Oslo, Norway (May 2017), the second international ABS workshop was held in Darmstadt, Germany (May 2018), and the third international ABS workshop will be held in Amsterdam, The Netherlands (May 2019). Researchers and research groups from, e.g., Oslo, Bergen, Darmstadt, Bologna, Madrid, Odense, Braunschweig, Torino, Paris, Djakarta, and Amsterdam are all stakeholders in the development of ABS and associated technology.

Silvia Lizeth Tapia Tarifa

Lizeth is the leader of the Analysis of Complex Systems research program. She is working in the area of formal methods, where she applies and develops techniques for the prediction of behaviors in complex systems. Now she is working with formal modeling 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 modeling concepts for resource usage in distributed systems to analyze resource management and service level agreement.

Scientific interests: Formal models, programming language semantics, complex systems, analysis, predictions

What triggers me scientifically:
Curiosity, interesting problems and new knowledge.

Lars Tveito

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.

Sigurd Kittilsen

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.

Johanna Beate Stumpf

Project: A calculus for resource management

Scientific interest: Formal Modelling, Process Calculi, Cloud Computing

What triggers me scientifically: I like solving problems.

Jia-Chun (Kelly) Lin

Jia-Chun (Kelly) Lin was a postdoctoral researcher at the University of Oslo and is now Assistant Professor at NTNU Gjøvik. She worked on 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, she 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, She is also working on 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 sometimes is painful, but I enjoy the process of solving difficult but interesting problems. When I conquer them, the happiness is more than anything.

Ontology Engineering

The digital transformation of the industry depends on rich information models that are intelligible to both computers and humans. These models should ideally represent the relationships in a manner to which domain experts are accustomed. This way users may explore and extract implicit information from data through the help of reasoning without the need for understanding the technical details of how and where the data is stored.

> **The aim** of the ontology engineering research program is to develop tools and methods that improve the efficiency and quality of ontology development, maintenance and use in industry.

However, the construction, maintenance, and use of such a model, called an ontology, are far from straightforward. Creating and maintaining a high-quality ontology requires close collaboration between domain experts, information modellers, and ontology experts to ensure that the model works as intended. Furthermore, an ontology quickly becomes a very complex artefact in order to express and make use of all the desired information artefacts, which in turn makes maintaining the ontology a real issue.

The aim of the ontology engineering research program is to develop tools and methods that improve the efficiency and quality of ontology development, maintenance and use in industry. These tools and methods are tailored to different users' expertise and requirements, facilitating a separation of concerns, where each user group can focus on what they know best. Domain experts and programmers no longer need to have an in-depth understanding of logic and semantic technologies, whereas ontology experts and information modellers no longer need to become experts in the domain of interest. This is achieved by:

- lowering the barrier for domain experts to understand, build, and use ontologies without the support of ontology experts.
- providing programmers and information modellers with powerful interfaces for interacting with the ontology and integrating the ontology in existing software platforms.
- equipping ontology experts with powerful tools to oversee the development of the ontology

Reasonable Ontology Templates (OTTR)

Reasonable Ontology Templates (OTTR) is a language and framework for representing and instantiating recurring patterns within ontologies. It facilitates a modular design approach following the don't-repeat-yourself principle.

Martin G. Skjæveland

is a postdoctoral researcher at the University of Oslo and leader of the SIRIUS research program on Ontology Engineering and Reasonable Ontology Templates. Martin's research interests are ontology engineering, ontology-based information systems and ontology-based data access with their applications to real-world challenges. Martin has industrial experience from working as a specialist in information management and ontology-based methods at DNV GL, and as work package leader for the Statoil use case in the Optique EU FP7 project. His work is published/ presented in/at leading semantic web journals and conferences.

This ensures that any changes made to a pattern are automatically transferred to any of its instances within an ontology.

The feasibility and industrial scalability of this approach was demonstrated in collaboration with Aibel in a paper nominated for best research paper at the International Semantic Web Conference 2018. It showcases the use of templates on a large-scale industrial ontology. It also addresses the various open-source tools geared towards different user groups, as well as algorithms for sophisticated maintenance of ontologies.

OptiqueVQS (Visual Query System)

The OptiqueVQS is a visual query builder that (i) supports users in constructing queries over an ontology; (ii) evaluates the query over a SPARQL endpoint; (iii) displays the query results. It was originally developed in the EU project Optique as an integrated component of the Optique platform. The vision in the Ontology Engineering research program is to deliver a tool suite that enables end users to efficiently interact with and explore large repositories of semantic data and large-scale ontologies.

OptiqueVQS has now been reimplemented as a stand-alone open-source application. Furthermore, a backend component for adaptive value suggestions has been designed and implemented. There has also been research published last year on ranking of user interface suggestions based on past queries.

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 twelve 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.

Leif Harald Karlsen

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

Vidar Norstein Klungre

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!

Semantic Integration

The Semantic Integration research program designs and develops scalable infrastructure that supports semantic integration using large ontologies (with many thousands of classes) and massive data sets (many billions of tuples). It will demonstrate the efficacy of these tools through deployment in the beacon projects. Specifically, we work with ontology reasoners capable of supporting the development of large-scale ontologies and semantic data stores which answer realistic ontology-based queries over massive data sets.

Digitization of oil and gas depends on integrating data from different sources. These sources have different forms and access methods. For example, some data owners may make data available in a raw format, or some may make it available only through their custom Application Programming Interfaces (APIs). The end user of this data wants a uniform view of the data, without the need to understand the underlying, often low-level methods needed to retrieve the data.

Most of the data in the oil and gas industry resides in a traditional database management system (DBMS). This data is usually accessed using queries written in the SQL language. However, it is challenging for non-experts to use SQL, especially when they want to access and use heterogeneous data. We do not want to expose this data directly to the end user. Instead we want to allow the user to ask for the data using their description of reality. This is the vision behind semantic integration. SIRIUS is working on two possible ways of providing semantic integration: query rewriting and materialization.

Alessandro Ronca

Project: Stream Reasoning

Scientific interests: Knowledge Representation, Automated Reasoning

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

Query rewriting: The most efficient way of accessing data depends on its representation format. For example, data in DBMS is in tabular format. This means that the query language, SQL, for that is such that it expects tabular form data input and gives tabular format data output. However, one can design other data representation methods, and just define a mapping of one method to another. For instance, RDF graphs can be stored in a DBMS by serialization of the graph edge-wise and storing it as a table. This means that the corresponding graph querying language designed for RDF, SPARQL, must be mapped to SQL. This process is called query rewriting or mapping.

The advantage of query rewriting is that it gives flexibility to use any data storage and representation format at the

DEPARTMENT OF COMPUTER SCIENCE

UNIVERSITY OF OXFORD

The Information Systems group at University of Oxford's Computer Science Department participates closely Data Science, Ontology Engineering, Scalable Computing and Semantic Integration research programs. SIRIUS finances DPhil candidates at the department, and Oxford finances more as an in-kind contribution. SIRIUS also co-finances post-doctoral researchers. Prof. Ian Horrocks, FRS, is the Scientific coordinator of SIRIUS.

backend, and any other format for querying, as long as the new querying method (language) can be formally mapped to the data storage format.

This approach was used successfully in the Optique EU project, where queries were written in a semantic form and then mapped to SQL databases owned by Statoil (now Equinor) and Siemens. The tool used for query rewriting was ONTOP, which is a query rewriting engine for the OWL 2 Query Language. The ONTOP project is hosted by the Free University of Bolzano, and SIRIUS is continuing to contribute to the development of this tool. Our plans are to extend ONTOP with mechanisms that support aggregation queries and queries for analytics.

> **In SIRIUS** we have been investigating how to extend RDFox to support streaming data, and how to distribute RDFox reasoning over a cluster.

Materialization: An alternative to query rewriting is materialization of a query. This involves copying, materia-lizing, the data that is needed for a query into a format that makes the query efficient and allows ontology-based reasoning. This allows us to interpret the backend data and infer additional data about this data by using clever querying methods. RDF and associated ontologies provide a rich resource for doing this. An ontology allows you to define rules of interpreting data about data, which means additional data can be generated, giving us more insight into the existing data. Consider a simple example: a compressor is labelled in one database using a NORSOK-format tag and in another using a serial number. An ontology can define a rule which states that these two names indicate the same equipment, say compressor1 is same as compressor2 and compressor2 is the same as compressor3. Then the materialization process interprets that compressor1 is same as compressor3 also, by transitivity. Thus, query rewriting and materialization together allow flexible access and interpretation about data stored in any format.

Materialization can be done effectively using RDFox, which is a state-of-the-art triple-store or graph database. RDFox is uniquely capable of answering queries over more than 10 billion facts, where the answer also accounts for the knowledge represented in large ontologies. RDFox supports the OWL 2 RL profile for ontologies and the SPARQL query language. Additional features include non-tree shaped rules, arithmetic/aggregation functions, stratified negation as failure and incremental reasoning.

Work on RDFox started in 2014, and it is still under active development. In SIRIUS we have been investigating how extend RDFox to support streaming data, and how to

Dr. Medha Atre

Dr. Medha Atre is leader for the Semantic Integration research program in SIRIUS and Senior Researcher at the University of Oxford. Prior to that, she was an Assistant Professor in the Deptartment of Computer Science at the Indian Institute of Technology, Kanpur. Her area of research is data management and various aspects of it such as storage, query optimization, data representation methods. During her PhD and later she single-handedly developed a system called BitMat for RDF graph processing, and in the process also extended some of the theoretical properties of SPARQL pattern query processing..

Smart and efficient answers to complex queries.

distribute RDFox reasoning over a cluster. The former is important in settings where streaming data interacts with complex domain ontologies; this interaction makes it difficult to use traditional window-based methods of dealing with data streams. The latter is important for dealing with very large data sets as RDFox stores data in memory for efficient reasoning. We are also working to overcome the limits imposed by its main-memory design by completing the development and evaluation of a fully distributed version. And we are extending its capabilities (e.g., with support for streaming data) and improving optimisations (e.g., query planning).

In 2018, we have also designed and implementation of a new ontology reasoner, called Sequoia. This, for the first time, applies consequence-based reasoning to the entire OWL language. Consequence based reasoning was until now only applicable to subsets of OWL. Sequoia already outperforms state-of-the-art reasoners on hard ontologies. In 2019 we will work further to develop algorithms and our prototype implementation into a fully-fledged OWL reasoner that significantly advances the state of the art.

Temitope Ajileye

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

Federico Igne

Project: Topic with the area of description logics, ontologies, query evaluation & automated Reasoning

Scientific Interests: Knowledge representation, Automated reasoning, Datalog, Answer Set Programming, Distributed systems.

What triggers me scientifically: The very concept of learning something new on a daily basis is what motivates me on my research. Having a challenge or a problem to tackle also helps keeping me focused.

Apart from my research area, I'm also deeply interested in any topic related to software freedom, software privacy & security and decentralization & federation of the Internet, both in their more theoretical aspects and applications. I work on a few projects related to these topics in my free time.

Daniel Lupp

Project: "Mappings Remapped: New Perspectives on Data Transformation in Ontology-Based Query Answering"

Scientific interests: data transformation, mappings, rule-based systems, ontology templates, ontology design/maintenance, reasoning.

What triggers me scientifically: Rather than "adding artificial intelligence" (whatever that means) to systems, I am fascinated by how one can remove or avoid the natural stupidity inherent in systems.

Data Science

Accessing data, which is the overall theme of SIRIUS, is only useful if we can use this data to make decisions. For this reason, SIRIUS has partnered with the BigInsight SFI to form the DataScience@UiO Cluster. We have also employed Basil Ell to lead our data science research program. In 2018 we have also linked SIRIUS' research in language technology and data science into a single research program.

Data Science is a broad and poorly defined field. SIRIUS' focus is on domain-adapted data science. Our aim is to contribute in three narrow areas where our mix of skills can advance research and help the oil and gas industry. These are:

- Extracting data from unstructured sources (documents that are designed to be read by people) so that it can be used by machine learning and data mining. These standard data science methods need structured data. Our research looks at how we can convert text, written in a natural language, into structured data. An example of how we do this is given in Extracting Information from Tables to Populate Knowledge Bases on page 40.

Daniel Bakkelund

Project: "Increasing the power of statistical algorithms in machine learning through augmentation with ontology theory".

Scientific interests: Mathematical modelling, statistical modelling and probability, order theory, discrete mathematics, graph- and network theory, especially ordered structures, graph- and network comparison, clustering theory and applying axiomatic-deductive methods for solving real-world problems through proven theories and algorithms.

What triggers me scientifically: Problems that lie in the crossing between applications and theoretical mathematical research, where the existing theoretical machinery for solving the applied problem doesn't exist, and has to be developed.

Farhad Nooralahzadeh

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: To research and explore the role of intelligent systems in understanding and serving society and humanity in an applicable and personalized way.

- Interpreting text that uses oil & gas language. Unfortunately, oil & gas language differs from normal literary Norwegian or English. This means that language processing tools need to be adapted to the special vocabulary of oil & gas documents.

- Using all the structure in structured data in our data science. Data from business, engineering and operational have a rich structure. Things are linked together. Invoices are linked to deliveries, bills of materials and work orders. Flow measurements in an oil platform are linked together by the flow of hydro-carbons through the systems. This structure is represented by graphs: networks or trees of connection between data. Machine learning tools, however, deal only with vectors, long lists, of data. This means that they cannot take account of the information that is available in the linkages between data.

Peyman Rasouli

Project: A Formal Interpretable Framework for Opening Black Box Decision Systems

Scientific Interests: Machine Learning, Data Mining, Automated Reasoning, Formal Methods, Optimization

What triggers me scientifically: Making black box decision systems explainable, transparent, and accountable personalized way.

In 2018 the language technology group in SIRIUS demonstrated a system for extracting relations from scientific text. This tool was built with convolutional Neural Networks using semi-supervised learning and structured domain knowledge. The tool was a top performer in the SemEval international shared task. This was a research competition on relation extraction from scientific texts. Our system ranked third out of 28 participants. Farhad Nooralahzadeh is working further with this tool to develop algorithms and a prototype implementation that deals with the complexity of realistic exploration reports.

This work depends on domain-specific word embeddings. The semantic properties of technical terms need to be captured and put in a form where they can be used by extraction tools. SIRIUS released such a tool in May 2018. Its results were validated by geoscientists from the Department of Geosciences in Oslo. It was also validated with Schlumberger's proprietary glossary of oil and gas terms and the GeoSciML (**http://www.geosciml.org**) standard.

Jiaoyan Chen

is a post-doctoral researcher at Oxford and is 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 such as 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 differentiable learning for more interpretable, transferable, cost sensitive machine learning; (iv) how machine learning can be applied in expressive ontology construction and ontology based data access.

Aasta Hansteen topside vessel
transfer December 2017.
Transfer from Dockwise White Marlin to
Dockwise Swift and Dockwise Teal.

Photograph: Espen Rønnevik / Woldcam / Equinor

Profile: Basil Ell

SIRIUS is doing work in the intersection between semantic technologies, statistical methods and language processing. We see that these distinct fields need to cross-fertilize to produce scalable data access solutions in and beyond the oil & gas industry. For this reason, we have recruited Basil Ell, a German researcher whose interests lie right at this intersection. He is working half-time in Oslo and half in Bielefeld, Germany. David Cameron from SIRIUS interviewed Basil in Oslo in January 2019.

Can you tell us a little about Bielefeld?

The University of Bielefeld has around 25000 students and has a strong focus on social and political sciences. It will soon also have a medical faculty. It was founded in 1969 and is built around a single building housing many faculties, with the idea of building interdisciplinarity. The university has outgrown this building, so the CITEC group is in a new building, a kilometre from the main campus. I work in a group of fifteen researchers, headed by Professor Philipp Cimiano.

The Bielefield University main campus

Basil Ell

Basil Ell studied Computer Science (B.Sc., M.Sc.) at Mannheim University of Applied Sciences, Pohang University of Science and Technology (South Korea) and the Indian Institute of Technology (IIT) Madras (India). In 2009 he joined the Knowledge Management group at the Institute of Applied Informatics and Formal Description Methods (AIFB) at Karlsruhe Institute of Technology (KIT), where he obtained a PhD with the thesis "User Interfaces to the Web of Data based on Natural Language Generation" in 2015, under the supervision of Professor Rudi Studer. There he was mainly working on Natural Language Generation from SPARQL queries and RDF graphs and on semantically-enhanced Virtual Research Environments for the History of Education. He joined the Semantic Computing group at the Cluster of Excellence Cognitive Interaction Technologies (CITEC), University of Bielefeld in May 2016 as a postdoctoral researcher and is currently interested in the Semantic Web and Natural Language Processing in general and in particular in knowledge base population, semantic parsing, and information extraction from text (relation extraction), and tables (table understanding).

The CITEC building in Bielefeld

How did you come to work in SIRIUS?

I came to work in Oslo due to a meeting between Martin Giese in SIRIUS and Professor Cimiano, where SIRIUS' need for a researcher with both semantic and language background was discussed. This led to an invitation to give a guest lecture and then the offer of a job, starting in September 2018.

Can you tell us a little about your research interests?

I am interested in the gap between humans and machines. Language is difficult for machines to understand: it is ambiguous and vague. There are many ways to say things. Machines need a well-defined way of representing knowledge, an organized way of describing the things expressed in texts and how they fit together. The first thing I am interested in is how to get information out of texts and into a knowledge representation. Once we have this, we can then also do Natural Language Generation: taking information in a knowledge representation and producing text.

We, thus far, have looked at how to take text, for example in Wikipedia, and use it to fill up a knowledge base. We want to do this automatically, so we need a system that can be trained, rather than being built manually. This is helped if we can use existing, publicly available structured data (DBpedia or Wikidata) to guide the extraction so that we can align this structured data with text.

What sort of projects can you see being done in SIRIUS?

An immediate area of application is SIRIUS' work on requirements. Manually converting existing requirements from text to a digital form is difficult and time-consuming. However, requirements are usually highly structured, so it is likely that we can parse requirements as text statements automatically to generate digital requirements. This work requires substantial computing power, so there is scope for designing clever algorithms and use high-performance computers.

Similar problems can be found with electronic health records, drilling reports, operational reports and inspection logs.

Within SIRIUS, I will be trying to strengthen the Natural Language Processing (NLP) work and bring it closer to the Knowledge Representation activities. This is actually the way we are addressing Data Science: NLP is a sub-problem of data science. Deep learning has been a popular and successful tool for NLP, but it seems to hit a performance barrier because it does not exploit known structure in the data. Our vision is to create machine learning algorithms that exploit the structure in data and its semantics.

Extracting Information from Tables to Populate Knowledge Bases

Tables are an interesting source of data for information extraction. Unlike natural language, where a large variety of forms exist to express a certain statement, tables can have a structure that makes information extraction easier, for example, when a row contains information about one entity and each column represents a relation or attribute and this pattern is repeated for every row.

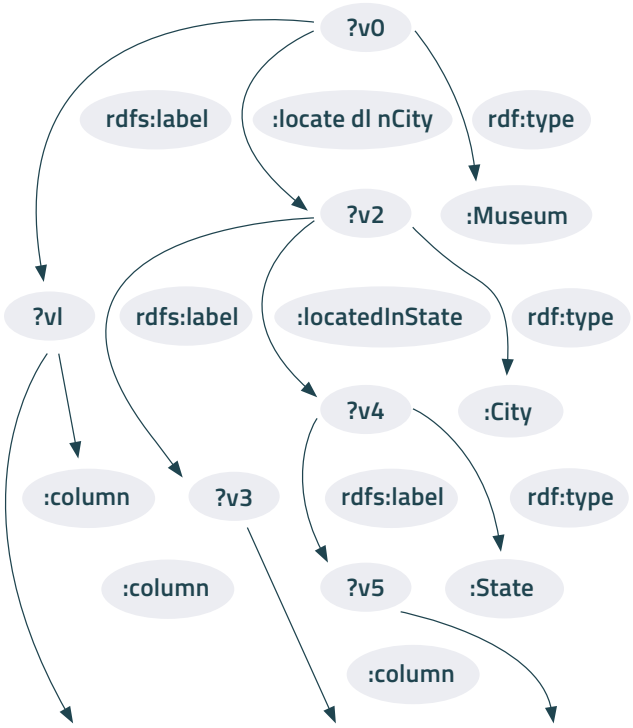
When humans try to understand the meaning of a table, they make use of their background knowledge and try to identify in the table the entities, relations between entities, and attributes of entities that they know.

Consider the table below.

Metropolitan Museum of Art	New York	New York
National Museum of Wildlife	Jackson Hole	Wyoming
Cincinnati Art Museum	Cincinnati	Ohio

Let's assume we have the following background knowledge (limited for the sake of the example). We know i) there are the categories museum, city, and state; ii) Metropolitan Museum of Art is a museum; iii) a museum is located in a city; iv) a city is located in a state; v) there is a city with the name New York and a city with the name Jackson Hole; vi) there is a state with the name New York and a state with the name Wyoming; and vii) the city New York is located in the state New York and the City Jackson Hole is located in the state Wyoming. Given the table and this background knowledge, we can try to understand the structure of the table, e. g., in each row in the first column we have the name of a museum, in the second column we have the name of the city the museum (that is mentioned in the first column) is located in, and in the third column we have the name of the state the city (that is mentioned in the second column) is located in.

Once we have understood the pattern given the knowledge that we already have, we can extend our knowledge by applying that pattern to each row. For the example above, we might learn that i) there is an entity with the name "Cincinnati Art Museum", ii) that this entity is a museum, iii) that Cincinnati is a city, iv) the museum is located in Cincinnati, v) Ohio is a state, and vi) that Cincinnati is located in Ohio. Thus, we can extend our knowledge graph with 6 facts. For real tables, one would try to derive a pattern from multiple rows. Note that if a table does not contain entities that are contained in the background knowledge or none of the relations expressed in the table are contained in the background knowledge, then the table cannot be understood.



Metropolitan Museum of Art	New York	New York
National Museum of Wildlife	Jackson Hole	Wyoming
Cincinnati Art Museum	Cincinnati	Ohio

Basil Ell and his colleagues are developing an approach for automated table understanding that we sketch here. Given a knowledge graph as background knowledge and a table, we create a hypothesis graph for each row as follows: for each cell, each entity that the value in the cell might refer to is added to the (initially empty) hypothesis graph. For each two entities mentioned in a row we add all relations between those entities that we are aware of according to our knowledge graph to the hypothesis graph. Given the set of hypothesis graphs we created for a table, we perform frequent subgraph mining to identify the commonalities of the hypothesis graphs. The result for our table might be the RDF graph pattern that is shown in the figure to the left, where the meaning of columns is described. For example, it expressed that for each row there is a thing that is a museum and that has a name that appears in the first column.

When applying the pattern to a row, we can remove hypotheses that do not follow the pattern. For the example above, "New York" might refer to the city or the state. If we find that the other entities in the column refer to cities, we can remove the hypothesis that links to the state. Thus, the pattern helps in disambiguating entities.

Once we obtain an understanding of a table, beyond the use case of knowledge base population, further use cases can be to automatically generate mappings between data sources in an ontology-based data access (OBDA) scenario, or tables can be indexed by the relations they express so that table retrieval can be improved.

Experiments will be done with a large corpus of tables extracted from the Web: the WDC Web Table Corpus 2015 (**http://webdatacommons.org/webtables/**). This contains 233 million tables, among them 90,266,223 relational tables. As background knowledge, we will use DBpedia (400 million facts) and Wikidata (54 million facts). Given the large number of facts, the large number of tables and given that frequent subgraph mining is a computationally hard problem, a high-performance computing infrastructure is needed. Both the table corpus as well as the knowledge graphs are publicly available. However, the results of these experiments may motivate to apply the approach on tables that are not publicly available with a combination of public knowledge graphs as well as proprietary knowledge graphs.

Scalable Computing

The Scalable Computing research program builds knowledge in high performance computing (HPC) and couples this with scalable cloud computing to support scalable big-data application processing. Specifically, we look at solutions for scalable and reconfigurable hardware, software design for parallel numerical simulations and autonomic cross-cloud application deployment and reconfiguration.

Main achievements

To ensure that industry gains a long-term benefit from the tools and methods developed in SIRIUS one must ensure the scalability of these solutions. This research program focusses on research in high-performance computing coupled with scalable cloud computing to support scalable big-data application processing. New hardware from Dolphin has been developed and will be integrated with the flexible PCI express computing infrastructure established at SIMULA Research Laboratory. The NUMA shared memory computer at the University of Oslo has been moved from the University's central computing facilities to the Department of Informatics, and work has been progressing well on the reservoir simulation software. A call for a PhD position in cross-cloud computing is currently open.

NUMASCALE
BIGGER DATA ANALYTICS

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 so many CPU cores.

High-performance test beds

Based on input and requirements from the SIRIUS centre partners, Dolphin has started research and development activities that will lead to further improvements in functionality, flexibility, scalability and performance. A new PCIe Gen3 switch prototype has been developed to enhance scalability. Internal tests indicate support for up to 60 nodes, deployment for high performance real application testing and tuning to the eX3 infrastructure installed at SIMULA Research Laboratory is planned to begin in 2019. This will enable additional SIRIUS partner laboratory activities. In the meantime, Dolphin has contributed its current PCIe adapter card and PCIe switch offering basic infrastructure for flexible sharing of compute and storage resources for big-data high-performance applications. This gives us an 8 node PCIe Gen3 x8 cluster that is currently available for the SIRIUS partner laboratory activities at SIMULA Research Laboratory.

The NUMA machine available for SIRIUS experimentation has been moved to the Department of Informatics in 2018. This results from the machine being reserved for SIRUS and the University of Oslo's central IT system operators desired the space and allocated personnel's capacity to support other infrastructure more generally needed at the University. Hosting the computer at the Department of Informatics gives unique access to the machine for the SIRIUS Laboratory managed by the local system operators, and although we have suffered downtime from the move this will be a better and more

Geir Horn

Geir Horn is Head of European ICT Research at the Faculty of Mathematics and Natural Sciences, and represents the Faculty on the SIRIUS strategic board and leads the SIRIUS Scalable Computing program. He holds a Cand. Scient. degree in cybernetics and a PhD in computer science on mathematical learning in combinatorial optimisation problems, both from the University of Oslo. He started his career at the Centre for Industrial Research in Oslo working on embedded software and fieldbus sensor systems, leading him on to distributed and parallel computing. Geir's current research interest is on stochastic combinatorial optimisation 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 technofixes, and Internet of Things.

Dolphin Interconnect Solutions

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.

Shiji Bijo

Project: Formalization of Data Movement in Multicore Architectures

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

What triggers me scientifically: Intellectual curiosity.

HPC on realistic oil & gas problems.

accessible solution for the SIRIUS Laboratory in the long run. It is anticipated that the NUMA machine is fully operational again from the end for the first quarter 2019.

The central system operators at the University of Oslo maintain a cloud infrastructure for research and innovation together with other major universities in Norway. Researchers from the SIRIUS Laboratory have started using this cloud infrastructure for small and medium sized computing jobs. In addition, the cloud infrastructure has been heavily used for testing the cross-cloud platform developed in the Horizon 2020 project MELODIC, which is linked to the Scalable Computing effort of the SIRIUS Laboratory.

Generally, experience from using the cloud has been very positive, and as soon as the NUMA machine is operational, it is planned to establish a cloud interface also for this and thereby offering the capability to use this machine both as a high-performance computer, and as a private cloud for experimentation in the SIRIUS Laboratory. We will test a real cross-cloud combining this private cloud with the universities' research cloud and other public and commercial cloud offerings, such as the cloud platform offered by the SIRIUS partner IBM.

Applications

The work on improving the reservoir simulations has been progressing well over 2018 with close interaction with the Equinor staff to discuss about further development of the ongoing collaboration. It is described on page 55.

An emerging collaboration within SIRIUS is an application within semantic natural language processing to derive labels from variables in queries formulated in SPARQL, a semantic query language for databases, and Resource Description Framework graphs and process these to understand the intended meaning. This will require significant computing resources, and the first results are expected as soon as the NUMA machine again becomes available.

An interesting future direction to investigate for both application areas may be to look at combinations of high-performance computing infrastructures with commercial cloud scalability. This architecture may allow temporary scaling on rented cloud resources, at the penalty of some additional execution time, but at reduced cost compared with dedicated and private hardware infrastructures. This will be initiated with the new PhD student starting to work on Scalable Computing in 2019.

Anastasia Gkolfi

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

Industrial Digital Transformation

Digital technologies offer the potential to transform organisational practices, roles and tasks. Often, however, the realization of this potential falls significantly short of expectations. The central challenge industrial companies face is to identify and cultivate the organisational pre-conditions necessary for realizing the potential of digital technologies. With pervasive digitalisation of contemporary organisations there is, as such, a need to operationalize existing scientific knowledge on implementation, adoption and use of digital technologies into actionable methods and recommendations that take organisational and work practice aspects into consideration for planning and implementing large-scale digital industrial transformation.

In 2018, we therefore changed the title of our scientific program from Work Practices to Digital Industrial Transformation. This change reflects the evolution of SIRIUS. With maturation of the research within centre, we have transitioned from having a supportive role towards developing our own a research agenda.

The **Digital Transitions Framework (DTF)** is a pivotal piece of our research program. This consists of a set of 'how-to' rules and principles, a design theory, which offers advice for planning and implementing digital technologies in large and complex organizational settings. These how-to principles are backed by theoretical, 'because of' justifications that are derived from existing scientific knowledge on the implementation, adoption, and use of digital technologies. The first version of the DTF is based on a systematic review of the literature on digital ecosystems. It is a set of justifications and how-to rules that a company can use to set up and transition to digital ecosystems.

We will continue developing the DTF by a) conducting empirical investigations to further build the theoretical set of 'because of' justifications, and b) empirically validating the validity of theory's design principles and rules. PhD student Mina Haghshenas' engagement with the READI Joint Industry Project (JIP), an industry-wide project on digitalizing requirements in the engineering, procurement and construction industry, is an example of the former (see page 68.). She works embedded with the JIP's core personnel at DNV GL and is developing empirical knowledge on the issues of setting up and cultivating a digital

NTNU
Norwegian University of Science and Technology

The Norwegian University of Science and Technology, NTNU, participates as a research partner through its Department of Computer Science (IDI) and the research group led by Eric Monteiro. This group combines computer science and social sciences to investigate how digital technologies are adopted in organizations.

Thomas Østerlie

Thomas Østerlie leads the SIRIUS research program on Industrial Digital Transformation. He is a research scientist at the Department of Computer Science. He holds a PhD in information systems from NTNU from 2009 and has 10 years of experience as software professional. He specializes in doing empirical research in close collaboration with industry. Østerlie has worked extensively with the oil and gas industry since 2009 and has previously led strategic initiatives on innovation in organizations at NTNU Social Research.

Elena Parmiggiani

Elena has just been appointed associate professor at NTNU. She studies the interplay between technical and social aspects involved in digitalization efforts. Her aim is to contribute to SIRIUS by generating researchable problems at the boundary between academic and industrial aims. She uses qualitative methods to generate data, including observations, interviews, and document analysis. Currently, she is 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, with the University of Oulu she is 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: 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.

ecosystem from the ground up. The results of her PhD project will be continuously fed into DFT's theoretical set of 'because of' justifications with attendant how-to rules and principles.

It was initially important to be specific on how we contributed to setting up and initiating research within SIRIUS. Work practices highlighted our role in identifying researchable problems grounded in the industrial partners' practical needs; both to avoid falling into the well-known trap of a researcher addressing highly interesting research challenges that amount to industrial non-problems, and to provide prototype feedback and evaluations to further refine and improve upon the applied aspects of the centre's technical research.

We have and will continue to pursue this agenda over the coming years. In 2018, we did so predominantly in the sub-surface domain, by working in close collaboration with Kadme on the DISKOS interface and by further developing the multi-scenario reasoning concept through ongoing work in the inter-disciplinary Geological Assistant project (see page 50).

As reflected in the renaming our research program, we have set out on a clearer path towards advancing the industrial partners' ongoing digital transformation activities. By drawing upon and further developing the Digital Transformation Framework, we offer practical and actionable advice grounded in state-of-the-art scientific knowledge on implementation, adoption and use of digital technologies that we believe most industrial companies will benefit from.

Petter G. Almklov

Petter G. Almklov joins the SIRIUS centre in a 20% position from 2019. He is Adjunct Professor at the Department of Computer Science at NTNU and holds a PhD (Dr. Polit) in organizational anthropology from NTNU after a Siv Ing degree in geological engineering. In addition to his position at IDI he is employed as senior researcher at NTNU Social Research. Almklov did his PhD on interdisciplinary collaboration in a subsurface department in a petroleum company. A particular focus in this work and subsequent research is on the relationship between work practice and representational digital artefacts (models, logs, maps, seismic data), studying how interpretations and decisions come about, particularly in operational situations. After his PhD much of his research was conducted at NTNU's Centre of Integrated Operations, particularly in connection real time reservoir optimization and subsurface modelling. His work has been published in leading journals in the fields of Information Systems and within Science and Technology studies. He has also extensive experience from high tech contexts outside the petroleum domain such as space operations and transport.

Exploration Beacons

Progress in the Exploration Beacons

The Exploration work-package manages a portfolio of experiment, prototype and pilot projects that support business processes in the subsurface part of the oil and gas business: exploration, geosciences, reservoir modelling and wells.

Scalable data access can increase the speed of decisions in field assessments by 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 subsurface 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.

An exploration workshop in 2016 and scoping workshops with Schlumberger and Kadme in 2017 identified a portfolio of project ideas for the Exploration work package. Along with the progress on the projects initiated in 2017, two new projects identified in these workshops were started in 2018. These are "PetData: A tool suite for exploration data wrangling" and the "Exploration Image Annotation" project.

The PetData project was initiated with Equinor, Kadme and Birkbeck, University of London in early 2018. This literature review and feasibility assessment looked at extending OBDA to the areas that were not addressed in the Optique project and make the OBDA system better suited for geoscientists and project data managers. A detailed project proposal was prepared and submitted to the Research Council, with Equinor and Lundin as industrial partners. The application was not successful, but we plan to continue to work on this activity together with Equinor and Lundin.

Jens Grimsgaard

Jens Grimsgaard is leader of the Exploration work-package in SIRIUS. He is a geophysicist based at Equinor'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, StatoilHydro, Statoil and now Equinor 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, contributions to the Scalable Computing program through hardware-compatible numerics, parallel programming and high-performance computing.

KADME

Knowledge and data management expertise

KADME (Knowledge And Data Management Expertise) team is composed of people that have many years of experience improving access to data within the oil and gas industry.

Part of the team has been involved with DISKOS since its inception in 1993. The tagline "Scalable Data access in the Oil and Gas domain" perfectly represents the driver behind our Whereoil technology, which is an Insight Engine for E&P data sources designed for Big Data and analytics. Whereoil connects to several information sources, both structured and unstructured, and offers a robust API for machine learning and E&P workflow integration.

To unlock the full potential of their 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 them 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. 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.

An Innovation project on annotation of exploration images was started, with Equinor and IBM as partners. It has an initial timeline of one year. A detailed literature review and feasibility assessment of the problem, concept and methodology was done during the summer and implantation of an annotation tool is underway. This project is expected to finish in mid-2019 and will deliver an open-source tool that supports semantically-driven tagging of exploration images for use in machine learning.

We have been working with Equinor and Schlumberger on the Geological Assistant project, which is described on page 50 below. A kick-off workshop held in February was followed by a detailed project proposal submitted for Research Council funding in September. This application was unfortunately unsuccessful. A geological trip was arranged for Informatics and Geoscience researchers (along with the students from the Department of Geoscience at the University of Oslo) to understand geological reasoning in practice. In 2019, this project is split into a short-term innovation and long-term research project. The short-term innovation project aims to produce a demonstration of a Geological reasoner to Equinor and a refined version to be demonstrated at the Schlumberger SIS Global Forum.

We have known for a long time that there are synergies between problems, methodologies and solutions in oil & gas domain and the health sector. A new project is under consideration to extend the methodology used in the Exploration Image Annotation project to the cardiovascular and angiography images from the Health sector.

A Platform for Subsurface Innovation

SIRIUS is working on a vision of providing a platform for innovation in the sub-surface. A recent book by Andrew McAfee and Erik Brynjolfsson, "Machine Platform Crowd"

EVRY

With a ten-thousand strong team, based in offices across Europe and Asia, Evry provides 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.

Evry is one of the operators of the DISKOS system and wishes to contribute to making it a platform for the oil and gas industry. Evry is also working on the upgrading of Joint Qualification System, the common procurement pre-qualification database for Norway.

Evry is also working with SIRIUS on exploring applications in health care.

traced the role of platforms as enablers for innovation by crowds of workers. We believe that there is a need to open up subsurface data to researchers and innovators so that they can try out their ideas on real data. We also believe that national data repositories, like DISKOS, have the potential to provide such a platform. For this to be done, however, we need to improve the access to the data and allow it to be linked with data in other databases. We also need to improve the access to unstructured text information in these databases.

This vision is shared by the SIRIUS partners Evry and Kadme. By using insights and technologies from several of the SIRIUS research programs 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. But the possibility of exploiting minor finds, better utilization of infrastructure, smarter production strategies are all 'buried' within the databank, waiting to be excavated if you have the right tools. The purpose and benefits speak for themselves.

SIRIUS has built a prototype of a platform for teaching and research at the University of Oslo. This used the Optique platform to link data in several internal data sources and data in DISKOS. This proved the feasibility of the approach and is a launching point for further work. In 2019 this prototype will be extended into a system that will allow joint work with our colleagues at the Federal University of Rio Grande do Sul. We will also be working with the open data set provided by Equinor from the Volve field. This data is a valuable test of the feasibility of our ideas and methods.

SIRIUS is also supporting and participating in other initiatives for sharing and interoperability of subsurface data. Our research provides a framework that can be used to implement an effective, robust and scalable framework for data access. The Subsea Valley National Centre of Expertise is organising a working party on data sharing, in which both Equinor and SIRIUS are active. We are interacting, together with our partners, with other initiatives, such as DataLink and the Open Subsurface Data Universe.

Geological Assistant

The Geological Assistant is a SIRIUS innovation project between university researchers, Equinor, and Schlumberger. The project's goal is to develop a tool-supported method for exploration geologists to better assess and evaluate exploration prospects by applying established techniques from knowledge representation and formal methods from software verification. The project includes researchers from the University of Oslo and NTNU with expertise ranging from implementation and use of digital technologies, knowledge representation, formal methods, and naturally, geology.

Geological Field Work in the Pyrenees

In October, SIRIUS researchers together with representatives from Schlumberger organized a three-day workshop for the Geological Assistant project in Aínsa, Spain.

Aínsa is a Spanish village on the south side of the central Pyrenees. This area is rich in different rock outcrops for studying different depositional environments, such as fluvial, deltaic and marine, and structural geological features on both regional and local scales.

We arranged the workshop in such a remote area to coincide with the annual master's level field course

Crystal Chang Din

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 minimized.

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

What triggers me scientifically: Functional correctness of software is important. I work on enhancing software quality in the development phase as early as possible and as efficiently as possible.

organized by the University of Oslo's geology department. This is a field course on sedimentology and sequence stratigraphy. By being in Aínsa at the same time allowed us to draw upon this group's geological expertise, as well as providing us with the opportunity to observe geologists at work in the field.

We spent the workshop's first day of the workshop in the field together with professors and students participating in the field course. Our objective was to study the work practices of geologists and learn how they work in the field; how ideas are communicated and shared, and how a geologist reasons about geology. Through observation and active participation in the field activities the project team arrived at an understanding on how geologists through drawings, use of concepts, body movements, and by iterating between regional, basin and local scales exchange their ideas and convey their interpretations.

The following day was spent in the hotel's conference room working on distilling and shaping our observations and experiences from the previous day into a design. We first defined an end-user workflow, which in turn formed the basis for design of an end-user experience. Linking the design with formal logic we have a framework where geologists can create a conceptual geological model that combines observations from the field with pre-existing knowledge and the geologist's interpretation. In addition to

ensuring that the resulting narrative is in agreement with formalised geological knowledge the system will help the user make a better model by challenging their assumptions and identify corresponding knowledge and data.

To further expand and challenge our understanding of the requirements for the Geological Assistant the third and last day was again a day in the field. The geology professors kindly showed us more and new geological cases. While teaching us geology they explained and answered our

Jens Otten

My research is focused on automating logical reasoning. More specifically, 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 efficient provers for classical and several non-classical logics. More recently I have developed nanoCoP, an efficient connection-based theorem prover.

I am adapting these theorem provers to use them to develop and access data in ontology-based databases

Scientific interests: Automated theorem proving, logic, 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.

questions about their work process when reasoning about geology.

SIRIUS is grateful to Ivar Midtkandal, Alvar Braathen, Miquel Pyatos More and the students of GEO4216 at the University of Oslo for letting us join their field trip.

Schlumberger

Schlumberger is the world's leading provider of technology 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, Schlumberger 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 innovative technology that adds value for our customers. We consistently invest more in R&E 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. It is a multidimensional environment that unites planning and operations. Bringing together advances in technical disciplines such as artificial intelligence, data analytics, and automation—underpinned by decades of unrivalled domain knowledge—the result is an E&P experience like no other.

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

The workshop participants: Jens Otten, Charlotte Fraser (Schlumberger), Hallgrim Ludvigsen, (Schlumberger), Irina Pene, Thomas Østerlie (NTNU), Ingrid Chieh Yu, Dag Hovland, Crystal Chang Din, Per Kulseth Dahl, Michael Heeremans, Kristoffer Torseth (NTNU).

Irina Pene

Irina Pene works as a researcher in SIRIUS and has a Master's degree in Geology with more than 15 years of experience as a petroleum geoscientist, mostly from exploration onshore in Romania and offshore in Norway. She has worked with both oil and service companies such as OMV-Petrom, Halliburton, Repsol Norge, where she has developed skills within seismic interpretation and prospect evaluation, as well as working with geological databases and software support. Her role in SIRIUS is to provide projects with the domain knowledge needed to make progress.

Subsurface Data Access and Analytics

Subsurface Data Access: Data Wrangling

Digital transformation of sub-surface work processes is about overcoming the bottleneck of data access and increasing the quality of interpretations by means of better use of data. The data access bottleneck means that up to 70% of exploration experts' time is spent on finding, accessing, integrating and cleaning data before analysis can even start.

Geoscientists find that *it is hard to get an overview of all available data related to an area of interest.* This data is spread over different applications and many internal and external data sources. No unified view is, as a rule, available up front, though a Project Data Managers (PDM) helps. *It is difficult to extract data from databases.* Should complex

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 its Research division, in particular the Yorktown Research Centre in the United States and the Brazil Research Centres in Rio de Janeiro and São Paulo.

IBM Research Brazil focuses on research related to oil and gas. This aligns very well with SIRIUS' aims. For example, the University of Oslo, Equinor and IBM are working on a project to develop tools to annotate geoscience images. Images must be annotated to train machine learning tools to allow geoscience images to be more precisely classified. Retrieving and processing information from large databases of geological documents is challenging, especially as lots of images are embedded in the documents. 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.

queries have to be written, a Central Data Manager (CDM) is needed. *It is challenging to extract data and information based on geological and petrophysical attributes*, as it is not possible to run these queries simultaneously on multiple data sources. *It is challenging to integrate datasets before analysis can start.* This is often tedious manual work that the geoscientists must do themselves. It is difficult to extract data and knowledge from the text documents as

Adnan Latif

Adnan Latif is project manager of the subsurface beacon projects in SIRIUS. He has background in Geoscience, Business Administration and Project Management. He holds a masters degree in Geophysics form the University of Oslo, and MBA degree in marketing and finance and PMP from Project Management Institute. Before joining SIRIUS, he has worked in the oil & gas industry as a Senior Geoscientist and has a broad experience within oil & gas exploration and project management. His main role in SIRIUS is to lead the project portfolio of subsurface projects.

there are very few tools that can deal with the contents of unstructured documents and reports.

Geoscientists are well aware of the limitations of the workflow. As a result, valuable analysis does not include all essential data and conclusions may be inaccurate or at worse, erroneous, due to incomplete data foundation.

SIRIUS is developing methods, tools and competence in *sub-surface data wrangling*. A sub-surface data wrangler has competence in both geoscience and digital technologies. These skills are crucial for integrating the workflows of geoscientists and data managers. A data wrangler can also

The SIRIUS Geoscientists (from left to right): Michael Heeremans, Irina Pene and Adnan Latif.

efficiently exploit opportunities brought by new IT technologies, such as identifying relevant data sources, developing complex ad-hoc queries over federated databases, and retrieving information from reports stored as text documents. In these ways, a data wrangler can bring data much closer to the project teams and give geoscientists a radically better possibility of extracting data and information with the exact specification (in terms of complex geological and petrophysical attributes) they need for their subsurface evaluation.

For the data wrangler to be less dependent on the data managers than geoscientists are today, we need to capture the specialized knowledge of data managers and build this into data wrangling tools. A successful attempt in this direction was Optique, a 14M Euro EU project that finished in 2016. Optique showed that geoscience knowledge could be captured in a knowledge graph and reusable mappings, built by data managers, could connect this knowledge graph to data in databases. Optique then demonstrated that complex queries over several federated data sources (including EPDS, NPD FactPages, Open Works installations, GeoChemDB, CoreDB and DDR) could be easily written and efficiently executed. Since this process was fully automated, tasks that normally would take several days could now be performed in minutes.

Optique showed the potential to transform the way data is gathered and analysed by streamlining the workflow and making it more user-friendly. However, Optique had limitations. It could only work with relational databases, had no built-in support for quantitative analytics, could not access unstructured data, and had limited tool support for constructing and maintaining the necessary knowledge graph and mappings. SIRIUS is working to address these shortcomings and aims to significantly broaden the applicability of the approach for use in exploration projects. Furthermore, as pointed out in the recent KonKraft report, DISKOS is particularly an important data source for exploration.

This beacon responds to the KonKraft recommendations on improving the user-friendliness of DISKOS by using DISKOS to motivate data wrangling challenges and evaluate results.

Michael Heeremans

Michael Heeremans is a senior engineer at the Department of Geosciences, University of Oslo. He has a PhD in structural geology and tectonics. His main interests are focused around facilitating good work practices for students and researchers in the field of research data management and common digital skills. He is actively participating in the Software and Data Carpentry communities, that focus on teaching basic lab skills for research computing and universal data literacy. He is also involved in the SIRIUS project as a geoscience domain advisor for the exploration work package. He has long experience in seismic interpretation and is well acquainted with oil and gas industry work flows.

Subsurface Analytics

Faster access to relevant data is of interest only if the data is can be used to create insight and guide decisions. Emerging sciences like Data Science can enable end-users to generate valuable insights for the subsurface evaluation and can provide significant aid to exploration decisions.

The subsurface analytics part of this beacon is focussing on developing methodologies and tools for data analytics, comprising topics like data analysis, machine learning, natural language processing and visualisation. Applications currently in focus are:

- Prediction of reservoir parameters using analogues, where the focus is to investigate techniques to identify and quantify uncertainties, thus predicting more accurate parameters to be used in exploration modelling. The primary objective is to extend Machine learning models that can incorporate Oil & Gas domain information and handles prediction uncertainties to a reliable extent.
- Domain-adapted information extraction for the oil & gas domain, where the focus is to construct a system which will extract and make searchable factual information from large quantities of unstructured, natural language text. In future, we aim to extend NLP/ML capabilities to extract structured (numerical) information from a large amount of unstructured data (documents) to generate valuable insights, which is beyond human capacity.
- The engineering aspects of building data access and analytics pipeline.

Digital Field and Reservoir Management

Oil & gas fields are becoming more and more complex. Production facilities and reservoirs interact with each other. The quality and quantity of instrumentation is increasing, with down hole monitoring and multiphase metering. Companies are looking at implementing real-time reservoir control through building digital twins of the entire field. SIRIUS' beacon project in digital field and reservoir project applies our skills in semantic technology, data science and scalable computing to improving how computers support modern field management and petroleum technology.

Better HPC support for reservoir simulation

SIRIUS is working with Equinor on how the computational performance of reservoir simulators can be improved. The multi-phase flow in oil & gas reservoirs is very complex and is modelled by a large system of nonlinear partial differential equations. High resolution is needed in reservoir simulations to resolve these flow phenomena. This leads to large amounts of computation that can only be properly handled by parallel computing platforms. Moreover, optimization and statistical analyses require repeated calculations. We need efficient high-performance computing.

Most reservoir simulators are proprietary systems. This has hindered swift development of HPC techniques. The OPM initiative (https://opm-project.org), however, represents a new community effort to promote openness in reservoir simulation research. OPM makes a number of real-world data sets openly available and coordinates collaborative development of open-source software for simulating porous media processes, like reservoirs. An extensive software framework has already been developed in OPM.

Summaya Mumtaz

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 analyze 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 a research in this area.

simula

Simula Research Laboratory performs high-quality research within information and communication technology. Simula is comprised of several companies dedicated to research, education and innovation. Simula's HPC team, based at Fornebu, provides the scalable computing software, hardware and numerical computing skills in SIRIUS.

However, the current OPM reservoir simulator is not yet capable of achieving the full potential of modern parallel computing platforms. This is due to both imperfect components in the numerical scheme and mismatches between parts of the implementation and modern hardware architectures. A SIRIUS PhD project (Andreas Thune, supervised by Xing Cai (Simula, Alf Birger Rustad (Equinor) and Tor Skeie (Simula)) aims to enhance the HPC capability of OPM's reservoir simulator. In 2018, Andeas rigorously tested and profiled the current reservoir simulator in the Open Porous Media (OPM) framework using an open data set. The numerical and parallel performance of the simulator, as well as the underlying code details were carefully studied. One essential numerical component in the numerical strategy was improved, leading to a considerable speedup of the workflow of Equinor's reservoir engineers. The result is a substantial improvement in numerical behaviour and thus a considerable reduction in the computing time. These results were presented at the OPM annual meeting. This progress opens possibilities for Andreas' ongoing research work, which will very likely lead to his first journal submission in the start of 2019.

Andreas Thune

Project: "High-performance numerics and software for parallel reservoir simulation"

Scientific interests: 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.

This work is internationally focused and Andreas visited The National Supercomputer Centre in Guangzhou, China, for one month in March-April 2018.

As part of the Scalable Computing research program, this project provides 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.

Xing Cai

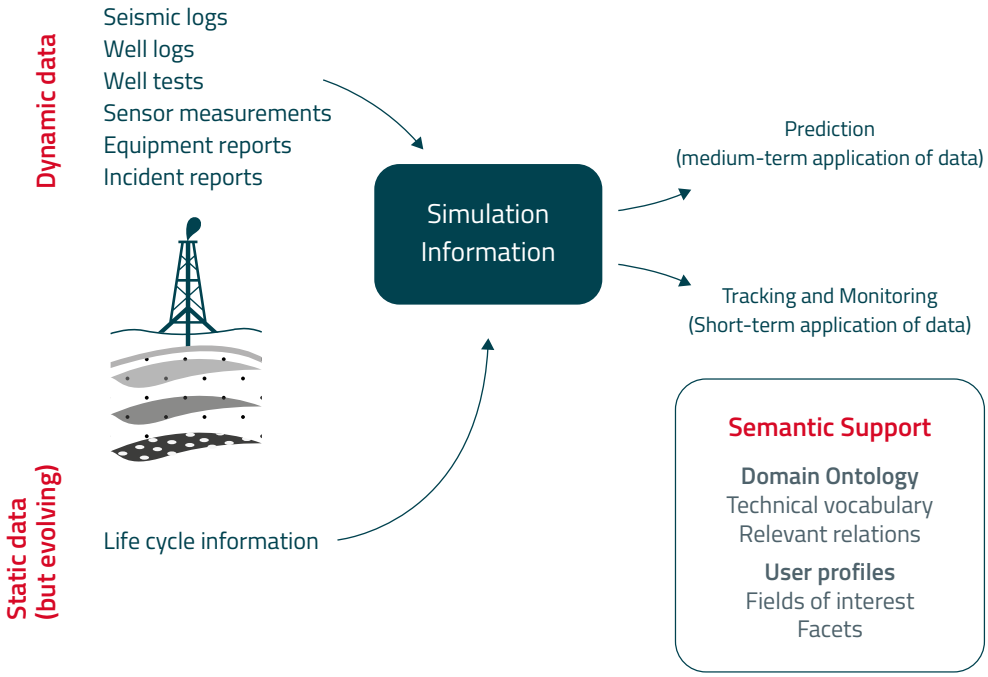
Xing Cai is a professor in scientific computing at University of Oslo and a chief research scientist at Simula Research Laboratory. He currently leads the IKT-PLUSS project "Meeting Exascale Computing with Source-to-Source Compilers" and co-leads of the FRINATEK project "Productivity and Energy Efficiency through Abstraction-based Parallel Programming". He is also one of the two scientific leaders for the new national eX3-infrastructure: "Experimental Infrastructure for Exploration of Exascale Computing". In addition, he led the finished FRINATEK project "User-friendly programming of GPU-enhanced clusters via automated code translation and optimisation". His main research interests centre around developing numerical algorithms and software that can effectively use cutting-edge parallel platforms. In the context of SIRIUS, he is coordinating Simula's research contributions to the strand of Scalable Computing through hardware-compatible numerics, parallel programming and high-performance computing.

Digital Integrated Field Management

SIRIUS has been working with the field management and petroleum technology groups in Equinor and Petrobras to identify IT challenges in digitalization of field development. Complex greenfield developments, like Libra and Johan Sverdrup, and brown-field projects, like Snorre, face difficult field management problems. Improvements in production and recovery require combination of petrotechnical analysis and data science. These analyses build upon a set of data and models that together provide a digital twin of the whole field: reservoir, wells, flowlines and production facilities. The components of this twin are commercially available but are locked into commercial and technical silos. Retrieval of data is time-consuming and requires database skills. Integration solutions that work for a single

application cannot be rolled out to other facilities. Differences in data and application landscape make it difficult to collaborate between disciplines for a single field and within disciplines for different fields.

The digital twin idea looks simple in principle, as shown in the figure. We measure, simulate and predict. Challenges lie in the diversity of the data used, uncertainty of the data and models and in the wide variety of commercial, proprietary and tailor-made solutions in the workflow. Digital twins are difficult to build, scale and maintain. There is a danger that current ambitions for digital twins will result in too-complex systems that fail to solve business problems. SIRIUS has proposed a project that addresses these challenges by advancing the research front in IT so that usable, maintainable and scalable Field Management Digital Twins can be used by industry in 2024. This research is interdisciplinary and focuses on petroleum-relevant IT disciplines: data semantics, analytics and machine learning, database design, data access, integration of data, applications & APIs and management of streaming and real-time data. Both the Norwegian and Brazilian corporate partners in this project have already initiated ambitious plans for whole-field digital twins. The purpose of this project is to further build the petroleum-relevant IT knowledge that is needed to overcome the challenges in making digital twins useful and sustainable.



This project has been prepared and proposed in collaboration with the Federal University of Rio Grande do Sul. More of this collaboration is given in the following section.

Collaboration with the Federal University of Rio Grande do Sul

In 2018, SIRIUS built up a promising collaboration with a leading research group in petroleum informatics in Brazil. This group is led by Professor Mara Abel and is located at the Department of Informatics of the Federal University of Rio Grande do Sul (UFRGS). This university is located in Porto Alegre, a city with 4.5 million inhabitants, located in southern Brazil.

SIRIUS came into contact with Mara Abel and UFRGS through her participation in the PPDM standards organization. Her academic background is interdisciplinary, with a master's degree in Geology, followed by a PhD in Computer Science. With this background, she has built a leading group in the application of computer science, and semantic technologies in particular, to problems in the petroleum industry. This profile matches SIRIUS' vision and goals very well, and we have found that the approaches used by her group are complementary to the ideas and methods developed in SIRIUS.

As a result of this, we applied for funding from DIKU, the Norwegian agency for Internationalization in Education, and CAPES, its Brazilian equivalent, for building an educational exchange program around digital geoscience. This application was successful, and has been used to finance visits from UFRGS to SIRIUS and a delegation from SIRIUS to UFRGS. This work has also been incorporated into a wider collaboration between the University of Oslo and UFRGS, formalized by a collaboration agreement signed in February 2019.

Vinicius Medeiros Graciolli

Vinicius Medeiros Graciolli is from Porto Alegre, Brazil and is a PhD student from UFRGS (Universidade Federal do Rio Grande do Sul), where he works with a research group which specializes in artificial intelligence applied to problems in the oil industry. Most of their work consists of machine learning and conceptual modelling. He is staying at the University of Oslo for a year to work on ontology modularization: ways to slice a large ontology spanning many knowledge domains into more digestible chunks which can be presented to users more specialized into certain aspects of that ontology.

UFRGS has a well-established collaboration with Petrobras, the Brazilian national oil company and with IBM Research oil & gas centre in Rio de Janeiro and São Paulo. This means that we are exploring the potential for building common projects with Petrobras and SIRIUS partners such as Equinor and IBM.

UFRGS hosted a six-month visit by Farhad Nooralahzadeh from September 2018 to Februrary 2019. Farhad is working on a PhD in natural language adaptation to oil & gas. During this visit he was able to work with researchers in Petrobras and IBM on problems related to his thesis.

In November 2018, a delegation from Oslo, consisting of Arild Waaler, David Cameron and Adnan Latif visited Porto Alegre to hold the initial Digital Geosciences workshop. In 2019 we will send two master's students to UFRGS and also receive a further visit by a UFRGS PhD student. We have the foundations here for a fruitful long-term collaboration.

Participants at the Digital Geosciences workshop in November 2018.

Operations Beacons

The operations beacons cover the parts of the oil and gas supply chain that do not involve subsurface data and exploration. They concentrate 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.

During 2018 we started work on each of the three beacons identified in 2017. These were:

- **Integrated digital planning.** Applying formal methods of analysis and simulation to planning problems in commissioning, supply-chain logistics and maintenance. This work is done in collaboration with Equinor. SAP and TechnipFMC are also now participating in this work.
- **Digital Twins.** Finding better ways of building and delivering systems that use operational data for decision support. One of the buzzwords of 2017 and 2018 was digital twins. SIRIUS partners OSIsoft, Computas and DNV GL worked 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. Our ambition is to contribute to or coordinate EU and Norwegian projects in application of digital twins.
- **Digital Field Development.** Implementing and standardizing better ways of managing requirements throughout the lifecycle of the facility. This work was realized in DNV GL's DREAM project and the READI Joint Industry project. This work involves DNV GL, Equinor, TechnipFMC and Aibel.

Our ambitions remain 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.

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.

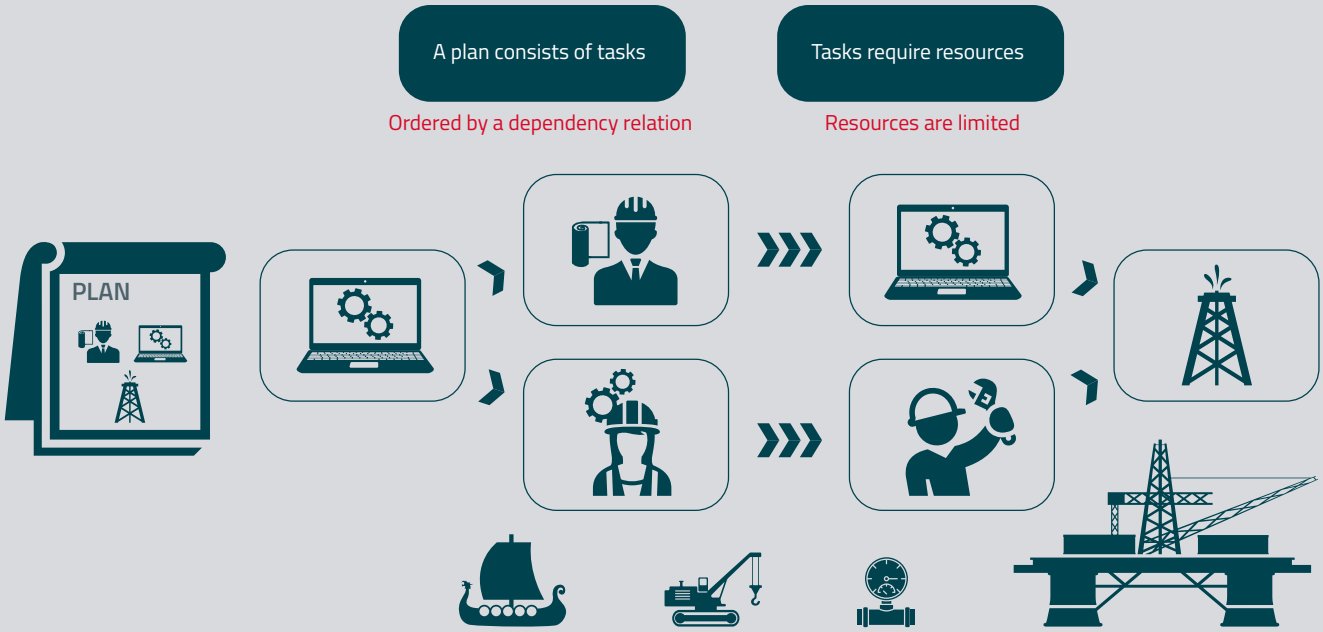
SAP is at the centre of today's business and technology revolution, where the ability to innovate is the core driver for purpose-driven and successful businesses large and small. Our engagement in building an SAP laboratory at the University of Oslo aims at supporting future professionals to create innovative 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.

Integrated Digital Planning

SIRIUS' planning beacon 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 Equinor 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 on 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.

Abstractly, a plan can be seen as a collection of tasks that

Figure 1: Plans, tasks and resources



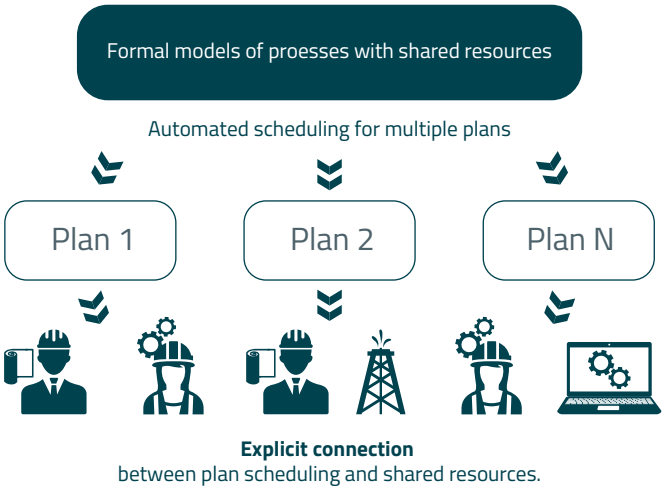
are ordered by a dependency relation. To perform the plan, resources need to be allocated to different tasks. However, tasks from different plans compete for resources, which may delay the scheduling of plans. One idea we are pursuing, is to expose the resource requirements of the different tasks earlier in the planning process, such that a plan's impact on the shared pool of available resources is explicit, and explore these requirements to improve visibility and facilitate scheduling.

Our 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

Einar Broch Johnsen

Einar Broch Johnsen is professor at the University of Oslo. His main research interests address the design and analysis of parallel and distributed systems, including resource-restricted and context-dependent behaviour. 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 was 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). Johnsen is or has been the supervisor of 14 PhD students and has published more than 100 peer-reviewed research papers.

Figure 2: Exposed requirements for the shared pool of resources improves visibility.



Equinor's vessels and cargo transport. These new tools allow customizable visualization of output from the ABS simulator.

We are working with industrial data from different parts of a complex supply chain and integrate 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 below. In this use case, ABS is used to define a general framework for modelling transport plans by means of abstractions for, e.g., vessels, containers, bulk cargo, route segments, and delivery deadlines.

The model is populated by specific data to represent a 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 ABS simulation. The

Figure 3: Raw operational vessel data.

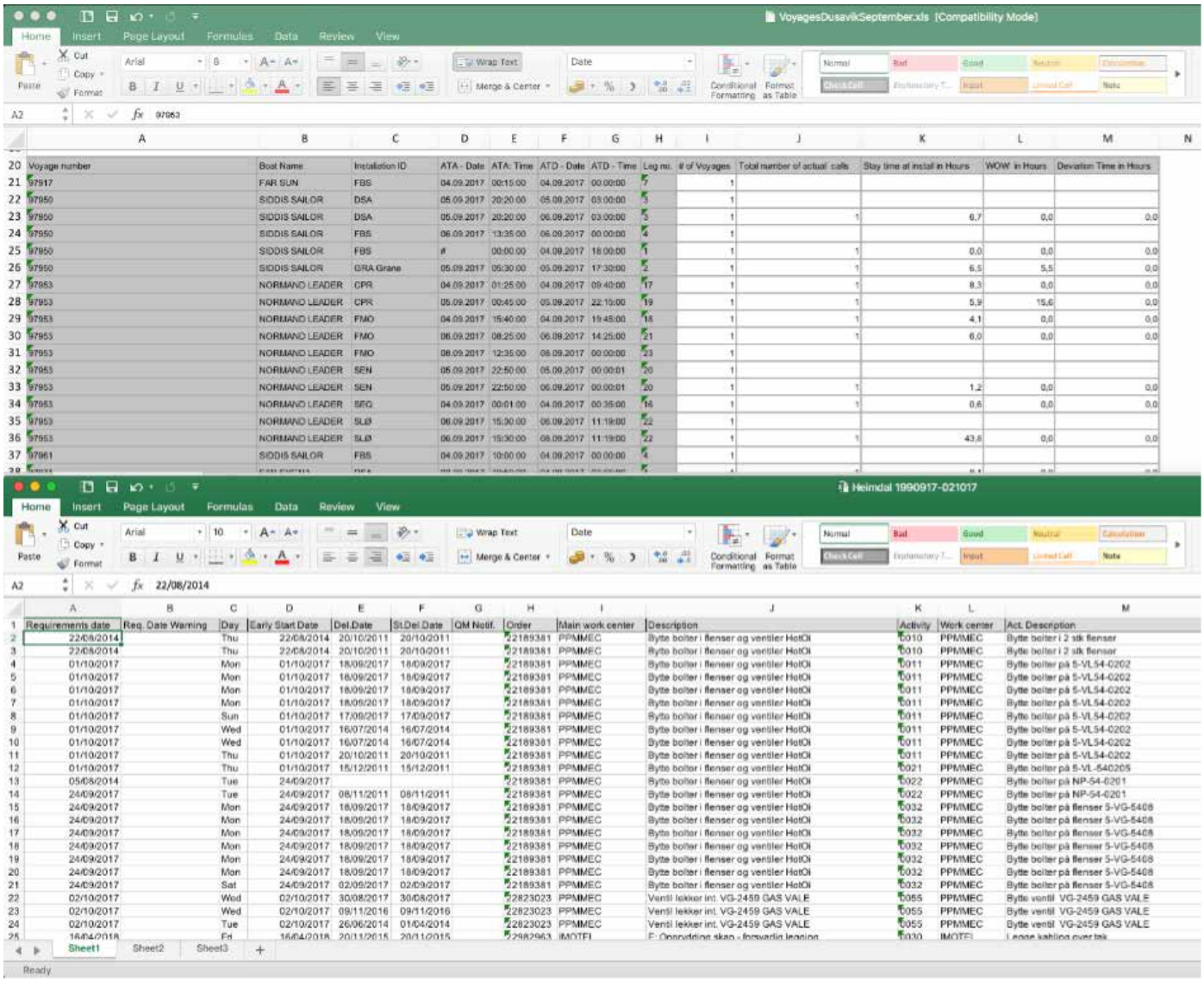


Figure 4: Visualization of time series data depicting vessel movements.

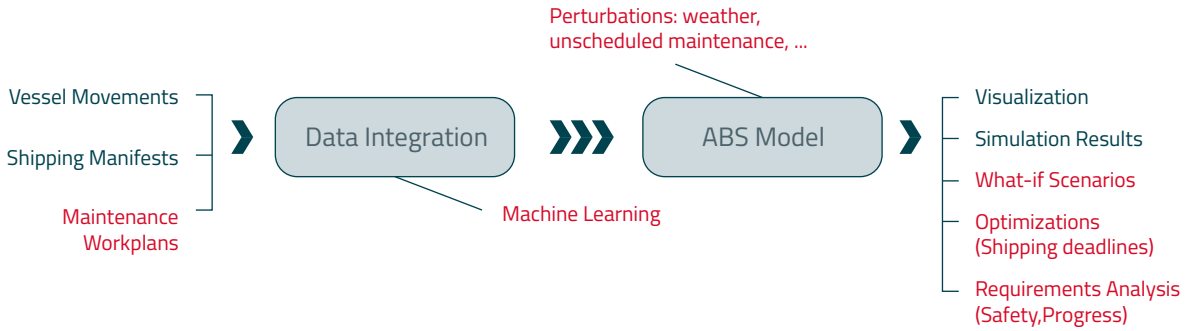
modeller specifies a time window, and data for this time interval is automatically extracted from the SQL database, converted into ABS data structures, and fed into the model as the concrete plan is simulated. This allows the ABS model of the concrete plan for the given time window to be simulated with up-to-date data. The planner is presented with a graphical view of the simulated plan, as depicted in the figure above. Our prototype allows the user to browse container lists for containers loaded to an offshore installation, or loaded on a vessel, and from there track the container's content back to work orders. 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 over the simulation data.

We will combine these simulations with stronger 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 by the following diagram, which shows the current work (in black) and some possible directions that can be explored (in red) to extend the current case study; these future directions provide concrete scientific challenges as well as added value in the planning processes.

Rudolf Schlatte

Rudolf Schlatte is a senior researcher at the SIRIUS centre, Department of Informatics, University of Oslo, and the laboratory manager of SIRIUS. His main research areas are semantics and implementation of distributed actor-based languages, and modelling and simulation of cyber-physical processes, and analysis of operational data. He has worked for various EU projects, including FP6 CREDO, FP7 HATS, FP7 Optique and FP7 Envisage, and was a fellow at the United Nations University, International Institute for Software Technology in Macao China in 2007-2010. He is the maintainer of the ABS language simulation engine, which is used internally in selected pilot projects in SIRIUS as well as by European academic partners. He has co-authored more than 30 scientific publications in international conferences and journals. He is the Norwegian representative of the ICT COST Action IC1405 on Reversible Computing.

Figure 5: Possible extensions of the current work.

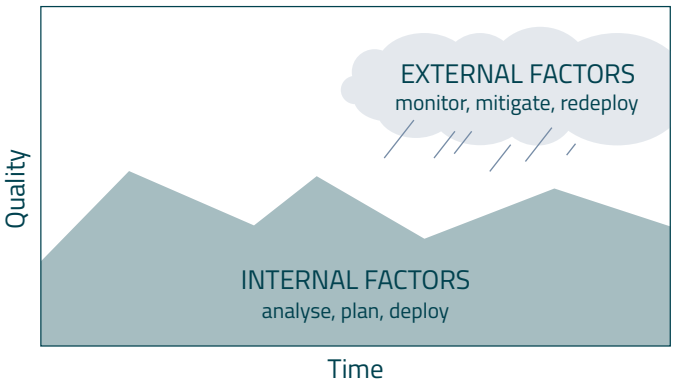


Finally, it is important to remember that we can only plan up to a certain level of accuracy, and that external factors will always perturb the plans (see the figure at the right). For this reason, it is essential to have planning tools that expose changes in the underlying, assumed data, to facilitate fault mitigation and replanning on the fly. We are currently working with TechnipFMC and SAP to broaden the scope of planning beacon.

Digital Twins

SIRIUS held a one-day workshop on Digital Twins in Oslo on 20th March 2018. This brought together participants from Equinor, Petrobras, Schlumberger, Halliburton, DNV GL, Computas, OSISoft, Aker Solutions, Aibel, IBM and SAP. Participants shared their ideas of what a digital twin is and should be. The results of this workshop were published at the Scandinavian Simulation Conference in September. They are also summarised here.

The digital twin remained an industry buzzword in 2018. Gartner group identified digital twins as one of the top ten strategic technology trends for both 2017 and 2018 and then placed the idea at the very top of their hype curve for 2018. What do we mean by a digital twin? The idea may have its origin in NASA, with roots back into the Apollo program. The US Government defines a twin as "An integrated multi-physics, multi-scale, probabilistic simulation of an as-built system ... that uses the best available models, sensor information, and input data to mirror and predict activities/performance over the life of its corresponding physical twin". Put simply, a twin is a digital replica of a physical object or system.



Evgeny Kharlamov

is a Research Scientist at Bosch Centre for Artificial Intelligence and an associate professor in SIRIUS. Evgeny's work has led to 100+ publications including top tier international conferences and journals. His work on industrial applications of semantic technologies has several times won or been nominated to best paper awards at prestigious conferences. He has worked in the Optique project and in a joint Siemens-EPSRC funded OSSA project on semantic analytics. Evgeny's research interests span different aspects of Artificial Intelligence and Semantic Technologies with applications in Industry 4.0, (Big) Data Integration, light-weight ontology reasoning, ontology-to-DB mapping bootstrapping, end-user-oriented data access interfaces, and ontology evolution and privacy.

Digital twins build on the idea that digitalization can use 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 suppliers. 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?

Brandon Perry

Brandon Perry is a Senior Software Research Engineer in the office of the CTO of OSIsoft LLC, He holds a B.S. in Mechanical Engineering and a B.A. in Hispanic Studies, both from Carnegie Mellon University. He has ten years of experience as an engineer in the technology sector and has consulted globally to industrial customers in the area of operational data modelling and analysis. He developed a product that integrated operational data with a leading geospatial information system. His research interests are complex time sequence data, industrial asset and data modelling, self-serve information systems, semantic data access, distributed systems and human-computer interaction.

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 complexity?

How can digital twins be made sustainable, maintainable and useful despite these challenges? We believe that a solution will need collaboration between computer scientists, control and simulation engineers, data scientists and end-user technical specialists. Here we will present the computer science and data science parts of solution. This research program combines the sub-disciplines of knowledge representation, natural language technologies, formal methods, scalable computing and data science. This knowledge of technologies must be informed by the deep domain knowledge that is embedded in the digital twins' simulation models and is owned by a facility's engineers, operators and managers.

computas

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 whom more than 90% hold an MSc or higher. 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**
- **Artificial Intelligence, Big Data and analytics services**
- **Digital business transformation and user experience services**
- **Architecture and project management advisory services**

In 2018 Computas sponsored an industrial PhD candidate, Rustam Mehmandarov, and participated in the digital twin beacon. The initiatives are well aligned with the company's business strategies and will further strengthen our commitment to SIRIUS.

We must use a semantic backbone that supports the integration of several different digital twin applications around a shared understanding of a facility's design. This backbone will need to find a pragmatic balance between comprehensiveness and maintainability. It will also need to build on standards in a way that prevents reinvention of the wheel and allows modular construction of semantic models. SIRIUS' recent advances in the construction of ontologies using templates promise to allow this.

Simulation providers can use these ontologies to exchange model configurations with engineering databases and the calculated results with the monitoring and optimization layers of the digital twins. Point-to-point connections through tag cross-reference lists can be replaced with declarative mappings – where data in the simulation results or configuration is mapped to items in the semantic backbone.

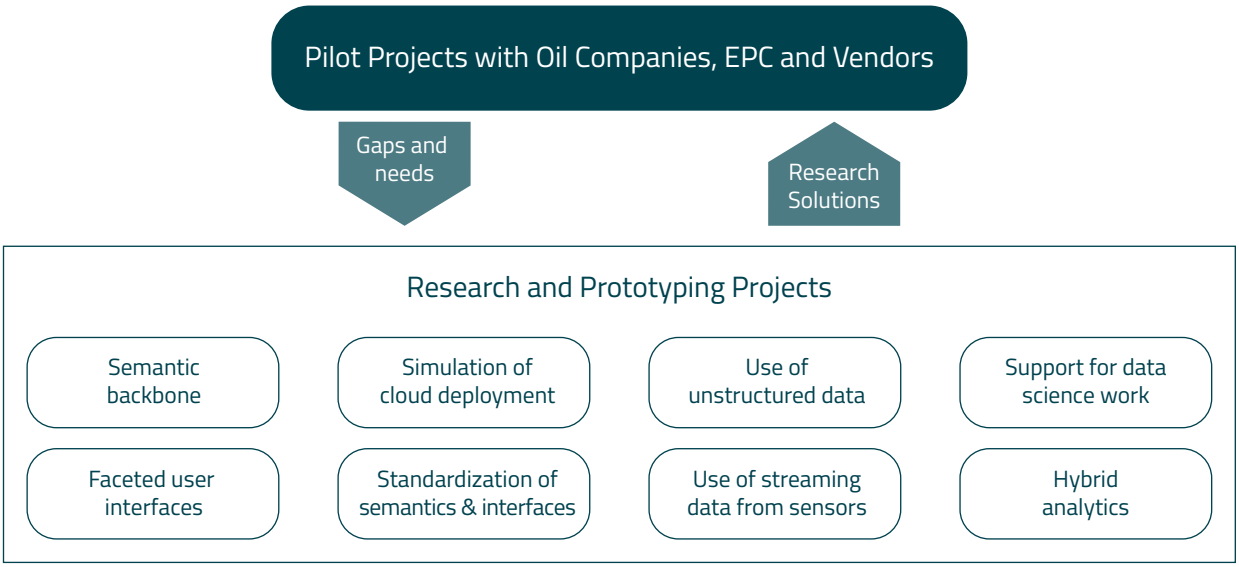
Good semantic models can also address the usability problems. Mapping data to concepts that are used by the end-user allows automatic generation of graphical interfaces that meet a specific user's needs. SIRIUS' OptiqueVQS framework is just such a tool.

A successful digital twin needs correlation of data from designs, measurements and simulations with text from logs and documents. There are now many commercial and academic tools for parsing and processing text. The IBM Watson framework is one such commercial offering. However, language algorithms that are trained on general data sets do not perform well when confronted with oil and gas terminology. SIRIUS centre is working on this domain adaption challenge, with promising results and good performance in solving standard challenge problems.

Challenges of maintenance and computational overload can also be addressed through using formal methods to design and monitor the deployment of a digital twin. As we noted above, a digital twin is a collection of interacting computational components, deployed across one or more cloud platforms and including edge devices. The behaviour of this system is difficult to predict, especially at design time. However, SIRIUS' ABS simulation tool for the computer systems themselves can be used to test different deployment plans and resolve challenges. The same model can also be used as a monitoring tool for the deployed system – a digital twin of the digital twin.

Finally, there remains the challenge of uncertainty, validation and data science. The digital twin is built on models. To quote George Box, all these models are wrong, but some are useful. A digital twin will contain many models. Some will be based on physical principles: structural, geometrical and process simulations. Others will be purely empirical, based on machine learning. These models must be validated against observed facility behaviour and aligned so that they mirror observed normal behaviour. Aligning models to observed data is difficult and remains an art. Finding out whether a discrepancy is due to an error in data, a wrong parameter, poor model structure or an actual malfunction in the facility requires a good understanding of the facility and well-developed judgment. This is true
whether the models a rigorous physical model or a machine-learnt empirical representation.

A research agenda for digital twins in the oil & gas industry.



A maintainable digital twin will contain structured tools that allow validation and tuning of all the models in the system. We believe that hybrid analytics – the combination of data science with physical and engineering simulations – is a valuable and fruitful area of research. Machine learning can benefit from being constrained by the laws of physics, while the laws of physics contain parameters that are uncertain or expensive to measure. Good statistical practice is needed in the engineering communities and engineering knowledge is needed among data scientists.

We are working on two projects related to data science for digital twins. The first of these is related to data access. Data science projects in industry are currently not scalable. Each new implementation needs to start from scratch, finding data, checking it and making it available. Our proposed semantic backbone will allow data science solutions to be rapidly transferred to similar sites in an organization.

The second area of research is related to the use of sensor data in data science. When a data scientist talks about streaming data, they usually mean a sequence of discrete event records – like tweets or sales transactions. The stream of data from analogue sensors is subtly different. The underlying signal from a sensor is continuous. The process of digitization itself introduces uncertainty and error in the calculation. Filtering and data compression provide further sources for error. Common data science frameworks expect data in vectors at common times. Production of this from a time series data base requires interpolation. All these details increase the cost and decrease the usefulness of data science work. A well-defined semantics and query tool for time-series data from sensors could solve many of these challenges.

Companies in the oil and gas sector are installing digital twins now, using commercially available platforms and siloed applications. We have an opportunity to engage with the observed problems of our colleagues in operations and maintenance through pilots.

Each pilot has a narrow enough focus to be doable. The companies we are collaborating with have linked these installations to a well-defined business case. Current pilots are ambitious: if successful they will bring previously unachieved levels of interdisciplinarity, effectiveness and access to data in design, operation and maintenance. At the same time, the pilots are focused on one specific business problem. By working with existing pilots and proposing new pilots we plan to establish a virtuous cycle, where shortcomings in today's technology and methods can be filled with research-driven innovations.

Digital Field Development

Field development is the process through which a discovered oil field is built so that it can produce oil and gas. This is a complex process, which goes through a so-called Capital Value Process, where investment decisions are made at decision gates. The model used by Equinor is shown in the following figure.

Francisco Martin-Recuerda

Francisco Martin-Recuerda is Principal Engineer at DNV GL AS. He has extensive experience in knowledge representation, semantic technologies, data representation and integration and data governance. At DNV GL, Francisco gives technical support for the design, implementation and deployment of semantic-based solutions for improving efficiency of engineering and business processes. He has been collaborating with DNV GL Data Management Competence Centre, Aker Solutions, Aibel, Siemens, Lundin and the READI Joint Industry Project. Prior to DNV GL, Francisco has been working for more than 13 years as researcher, consultant and project manager in several large research projects funded by EU including Knowledge Web, SEKT, DIP and SEALS. The aim of these research projects was to improve the state-of-the-art on semantic technologies and ontology engineering to facilitate data and process integration. Francisco holds a PhD in Artificial Intelligence (2016) from the Universidad Politecnica de Madrid (Spain), an MSc in Artificial Intelligence (2003) from the University of Edinburgh (UK) and MEng in Computer Sciences (1995) from the Universidad de Granada (Spain).

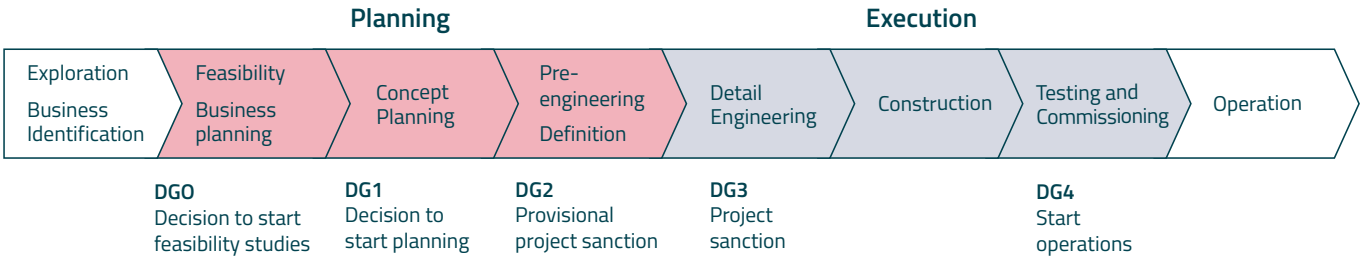
Similar models are used by other companies. For example, BP calls the three planning steps Appraise, Select and Define. As a field is moved from DG0 to DG4, large amounts of information is generated about the field and its facilities. This information is used as the basis of the investment decisions at each decision gate. It forms the design description of the facilities and provides the audit trail for the development. This data grows in size, detail and complexity, until it becomes the as-built data for an oil platform and its subsea equipment.

This information is created by the operating company and its suppliers. At present this is done by handing over documents that are then worked on by the contractor and are handed back to the operator. This process is laborious and costly. For this reason, the industry, represented by the KONKRAFT working party, is working to digitalize these processes and improve the effectiveness of data creation and sharing between operators and contractors.

SIRIUS is working on projects that look at the creation, use and management of three types field development information:

- Requirements.
- Design information.
- Project plans and schedules.

These three types of information interact, and should not be seen in isolation. A requirement must be met by a design, and this design must be built and commissioned. We believe that semantic technologies are a way of standardizing and sharing this information so that hand-overs are quick, efficient and error-free.



In 2018 we have been working on requirements, together with DNV GL's DREAM project and the KonKraft READI project. These projects are described in more detail below. Good management of design data in engineering is needed for successful use of digital twins in operations. A semantic framework built during design should be the framework used for operations. This means that the operator of the plant has easy access to the design data they need to make decisions.

In 2019 SIRIUS will be looking with application of ontology engineering and semantic integration to design information and project plans and schedules.

DREAM and READI: Cooperation to Manage Digital Requirements

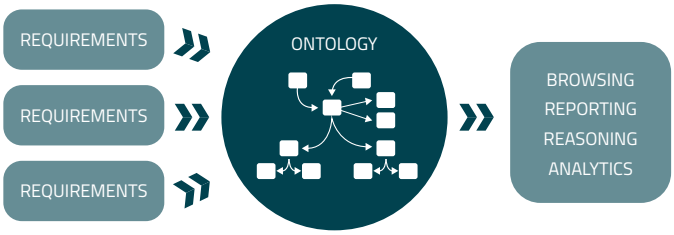
DNV GL, SIRIUS and the Norwegian oil & gas industry are working together on a joint industry project that will pilot a new and digital way of managing the requirements in engineering, procurement and construction projects. Tore Hartvigsen, Senior Principal Engineer at DNV GL presents this project.

The oil & gas industry is built around a set of specifications, standards and recommended practices, and many of these have been developed and are maintained by DNV GL. Digital transformation has implications for how these standards and published, shared and used. Currently DNV GL's portfolio contains about 800 different requirement documents related to our focused business sectors. As a first step in digitalization, these documents have been converted from text into XML. This means that requirements are translated from unstructured text to structured text. This improves document and requirement management processes, but does allow us to exploit the full benefits of digital transformation.

DNV GL spends about 5% of its revenue on research, innovation and business development. In 2018 the DREAM (Digital REquirements and Assurance Management) project was sponsored by an internal research programme. This project investigated how concepts, methodology and tools developed by a joint SIRIUS and NORSOK Z-TI initiative in 2017 could be applied for full digitalization and management of DNV GL's portfolio of requirements. Close cooperation with SIRIUS was initiated. Methods, tools and

demonstrators were further developed for testing based on real use cases in the maritime, oil & gas and energy sectors. We worked with domain specialists from each sector to select, review and test selected sets of requirements and represent them in a new format that allows computers to perform automated reasoning. In this way, we can check automatically if requirements are met, how they related to data about a project and whether they are consistent. This digital format also allows storage and sharing of requirements in data bases. This initial work was promising and has resulted in the DREAM tool: a framework for representing requirements in a fully digital form. The principle behind this tool is shown in the figure below. Relevant requirements, asset data and definitions are all implemented in a knowledge model – an ontology. This representation, stored in a suitable database, allows computer programs to browse, evaluated, validated and analyse requirements with a precision that is impossible with paper documents.

The challenge of digitalizing requirements reaches far outside DNV GL. The Norwegian Oil & Gas industry's KonKraft initiative seeks ways to reduce cost and improve effectiveness in operation. A strategy report in 2017 identified reduction of paper and duplication as a vital aspect of digitalization. For this reason, KonKraft set up a Joint Industry Project (JIP) called READI in mid 2018. READI is an acronym for Requirements Asset Digital Information. The knowledge, experience and products available from the NORSOK Z-TI pre-project and the DREAM project made DNV GL a natural choice to coordinate and manage this project. This joint project involves other SIRIUS partners, Equinor, Technip-FMC, Computas, Aker Solutions and Aibel, as part of a wider group of engineering and operating companies.



Semantic technologies are used to integrate requirements, dictionary data and asset data to enable reasoning and analytics to be used on data related to oil & gas installations.

DNV GL enables organizations to advance the safety and sustainability of their business delivering services to maritime, oil & gas, energy and other private and public sectors. Our specifications, standards and recommended practices, combined with services delivered by a highly skilled and trained organization, have founded the basis for our business since DNV was established in 1864.

DNV GL's primary involvement in SIRIUS is in the area of ontology engineering and its application of field development and digital twins. We also collaborate in the personalized healthcare beacon through the BigMed project.

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. Realising 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.

Rustam Mehmandarov

Project: "Digital Compliance Management for Operations (DCMops)".

Scientific interests: Semantic Technologies, Knowledge Representation, Cloud Architecture.

What triggers me scientifically: Being able to apply theoretical knowledge in the industrial context.

The READI project works on the digitalization of NORSOK standards. The NORSOK technical information standards intended to ensure common practices and rules for information exchange and ownership throughout the life cycle of oil & gas assets. They stated principles for digital information to be accepted as a project deliverable, specified ownership for the project information generated in the various project phases and defined the responsibility to secure quality and accessibility to such information to allow concurrent engineering processes involving several organizations.

Phase 1 of READI is updating and converting five NORSOK standards for management of technical information. These standards were written in the mid 1990s, with the ambition of reducing capital cost and improving collaboration. They have, for more than two decades, led to a shared understanding of the information that needs to be organized and circulated between the parties in a field development project. However, each operating company has developed different practices and detailed requirements specifications. This results in practices where the contractors and manufacturers have to use different requirements and associated work processes from project to project. The industry suffers from superfluous and inconsistent requirements and a very high cost for production and management of the documents and information needed.

READI builds on previous work done in a JIP, run by EPIM, called STI. This project established a common industry understanding of the necessary documents to be delivered for all types of equipment. It delivered its result as planned at the end of 2018. One of the five NORSOK standards to be revised by READI, Z-018-Supplier's documentation of equipment, will continue of the work done by STI. For another standard, NORSOK Z-001 Documentation for Operation, READI builds on a DNV GL recommended practice DNVGL-RP-0101, Technical documentation for subsea projects, which STI had also adapted for vendor documentation requirements. READI will digitalize the requirements from this recommended practice. The project will combine the requirement converted into a digital representation with asset data from existing subsea installations. This allows us to investigate and demonstrate possibilities and effects of automatic requirement verification. The tools for this are tailor-made tools in READI and are all developed with significant contribution from SIRIUS. This involves the use and further development of templates for mapping requirements, dictionary and asset data formats from their sources to the READI knowledge model. This is done by using the OTTR templates (see the SIRIUS 2017 annual report for a description of OTTR).

It is demanding and labour intensive to convert a requirement from text to a digital format. A human is needed to do this translation. Natural Language Processing (NLP) can allow us to convert of text-based requirements to digital requirements automatically. Text requirements often contain patterns that can be used in NLP. This work builds on the tools under development in the Data Science research programme.

Digitalization of requirements will not save money unless it is used as a way of digitalizing work practices. For this reason, the READY JIP is working with Mina Haghshenas, a PhD student from NTNU and SIRIUS' program for Industrial Digital Transformation. She participates in the daily work tasks in the JIP and uses the project as a research case in her study of the organizational aspects and effects of running a large Joint Industry Project like READI.

Johan W. Klüwer

Johan W. Klüwer is a Principal Specialist, with 12 years' experience, in DNV GL. With a background in applied logic, Johan specializes in development of enterprise ontologies, with emphasis on enabling automated reasoning for compliance with industry standards. In recent years he has made substantial contributions to Aibel's MMD ontology and to Aker Solutions' information model for the PUSH project, and advised Siemens Corporate Technology as well as various DNV GL departments. He has worked in Oil & Gas, Health and life sciences, Maritime, Defense, and Public sector projects. In the Optique project, Johan contributed to dissemination activities and ontology development. His current focus is in developing a methodology for digital requirements for the READI joint industry project. He has contributed to modelling best practice and upper ontologies, mainly in making ISO 15926 work with OWL DL ontologies, and with ontology templates applying the OTTR framework to ontology engineering. He was part of initiating the SIRIUS Colloquium on industrial ontologies, and has lately engaged with the international Industrial Ontology Foundry, in an effort to disseminate the Norwegian community's advances in ontology-based methods to a wider audience.

Mina Haghshenas

Project: "Digitalizing the requirement pipeline in an EPC project"

Scientific Interests: Standardization, Digitalization, Change Management

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

Tore Hartvigsen

Tore Hartvigsen has worked with oil & gas, defence, manufacturing, health and public clients in projects that include software and hardware development, product acquisition, implementation, support, evaluation, risk management, cyber security assessments. His interests include information risk management, information quality management, cyber security, software development and semantic technologies. He was project manager for the DREAM project and is responsible for methodology, tools, training and support in the READI project.

Cross-Domain Beacons

Personalized Medicine

Personalized medicine is the practice of making medical decisions, tailoring medical interventions, and fitting the management of disease to a patient's individual characteristics. Genetics plays a major role in this process and is used to help predict a patient's response to treatment or risk of disease.

Characterizing the patient's physical or biochemical characteristics as well as any responses to interventions is essential for tailoring treatment and diagnosing the patient.

Laura Slaughter

Laura Slaughter received her PhD in 2002 in Information Science from the University of Maryland. This was followed by post-doctoral 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 health-care 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.

Dag Hovland

Dag Hovland works on several SIRIUS projects, both with health (BigMed) and with partner companies. His 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.

Any observable trait, such as a behaviour, developmental delay or physical anomalies, is what is called a phenotype. Phenotypes are a manifestation of genetic variation and are essential data to collect that form the basis for personalized medicine.

SIRIUS researchers have been working on the BIGMED personalized medicine project (2016-2020) to create tools for identifying and specifying clinical phenotypes. Dag Hovland and Laura Slaughter worked on an ontology-driven form-based tool to be used by paediatricians to select phenotypes observed in newborns with suspected rare genetic disorder. The tool uses the Human Phenotype Ontology (HPO) and provides some guidance to selection of phenotypes to be used in the diagnostic process.

SIRIUS researchers met with BIGMED clinicians and lab technicians to outline functionality of the phenotype tool which will be completed in 2019. Functionality includes using the ontology-driven system to track phenotype iterations throughout the diagnostic process, adding triggers that indicate a need for reanalysis, the addition of other input methods than the form check-boxes including free text.

Ernesto Jimenez-Ruiz has contributed to the work on clinical phenotypes through work to find alignments between disease and phenotype ontologies. As part of the Pistoia Alliance Ontologies Mapping Project to find or create better tools for mapping ontologies, he has examined pairwise alignments between the Human Phenotype Ontology (HPO) to Mamma-

lian Phenotype Ontology (MP), the Human Disease Ontology (DOID), and Orphanet Rare Disease Ontology (ORDO). This work can be applied to BIGMED, in the case of newborn rare disease diagnosis, the workflow includes examination of phenotype to possible diagnosis, because the source ontologies are dynamic and change often, automation of the mappings is essential to the maintenance of tools and services developed as part of the project.

Laura Slaughter attended a workshop on automatic identification of clinical phenotypes from EHR texts in Tromsø in June, 2018. Activities include scoping using an open source dataset of electronic health records (MIMIC III dataset) in order to explore extraction of phenotypes automatically, including free text as specified by the clinicians in the BIGMED project.

In 2018, Violet Ka I Pun has started investigating the workflow a of prenatal visit in a hospital in Bergen. An initial executable model of the workflow, as well as a preliminary visualisation of the model, has been developed. In addition, Pun has initiated collaboration with the Bergen Regional Health Authority, which agreed to provide a use case for the project. In addition, she is establishing a collaboration regarding scheduling and planning the use of medical equipment for surgical theatres in Bergen. This collaboration also involves the University of Lübeck in Germany, which can provide the knowledge of the interconnection of medical devices.

Name: Violet Ka I Pun

Violet Ka I Pun is Associate Professor at the University College of Western Norway and a researcher in SIRIUS. Maryland. She uses formal methods to analyse behaviour and worst execution time of concurrent programs. In addition, she uses formal analyses to detect possible data leakage in concurrent programs. She is also interested in formalising models for multicore architectures with multilevel caches, and to analyse 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 my main motivation to tackle scientific research problems. Also, the process of researching the unknown subject is always fascinating, and motivates my scientific thinking.

Erik Bryhn Myklebust.

Project: Machine learning on graphs with applications in environmental risk assessment.

Scientific interests: Machine learning, climate and environment sciences.

What triggers me scientifically: Finding solutions that can increase understanding of the environmental processes.

Environmental Applications

SIRIUS has had cross-domain activities in the Arctic Earth Observation in 2016 and 2017 through a collaboration with the CIRFA Centre for Research-based Innovation in Trømso. Equinor is partner in both CIRFA and SIRIUS, and was instrumental in setting up this collaboration. In 2018 SIRIUS and CIRFA worked on submitting a Marie Curie proposal, that was unsuccessful, but that formed the basis for CIRFA participating in a successful Horizon 2020 proposal later in the year. This project is coordinated by Manolis Kourabakis from National and Kapodistrian University of Athens, who is a long-term collaborator with SIRIUS. In addition, CIRFA has recruited a PhD fellow, who will be jointly supervised by Einar Broch Johnsen from SIRIUS, will work on efficient cloud deployment of the CIRFA pilot applications.

Scalable data access is a challenge in many environmental applications. Observations and measurements take many different forms and are spread across many different databases. This situation is the same as that faced by engineers in the oil & gas industry. For this reason, we are collaborating with NIVA, The Norwegian Institute for Water Research, to explore how semantic integration and ontologies can improve data access and the quality of decisions made in environmental toxicology. A NIVA stipend holder works with SIRIUS and is supervised by Professor Martin Giese.

In 2019 SIRIUS will be holding a workshop on scalable data access in the environmental sciences, where we will follow up our work with CIRFA and NIVA and other organizations such as met.no and NORSAR.

International Activity and Dissemination

Oil & Gas Digitalization: The International Perspective

The oil & gas industry is an international business and SIRIUS' work is of interest and relevance far beyond the Norwegian Continental Shelf. It is essential that work done in Norway influences and is influenced by the work done by companies and Universities in North and South America and Asia. We have already set up collaborations in Brazil, and in 2019 have appointed Keith Lewis in a 20% position as International Advisor and Beacon Manager. Here, Keith presents his perspectives on digitalization in the global oil & gas industry.

In 2018 there was noticeable change in the programs of international oil & gas conferences. Finally, it seemed that the oil & gas industry had discovered the benefits of the digital age. Digitalization took the centre stage at conferences and the industry was proud and confident to speak about their digital aspirations.

A number of new players spoke at industry events. Technology giants such as Google and Amazon presented the benefits of their digital services. Many traditional oil and gas service providers had a new digital pitch. Oil and gas companies presented how digital technology and process improvements would transform their business.

All shared the hope of a digital future that could revolutionize an industry that had been suffering from low oil prices and high costs. From an industry perspective, this was a real boost and a welcome change from the past few years where the main topic of conversation was cost reduction to remain competitive in a low oil price environment.

Beyond the buzz of the international conferences, the industry is actively developing digital strategies, road maps, and plans to deliver pilot projects for all phases of the exploration and production value chain. This activity is global, with knowledge and information being spread between international locations by companies with global reach and through international research collaborations.

Governments also see the growth opportunities that the digital age can deliver to their economies. As a result, they are stimulating international research collaboration, not only as a vehicle to build competence but also as a way to create international trade opportunities.

Significant momentum has been built up in 2018. This will carry on into the future, as companies look at the success and learnings from their peers. However, like any other change process, digital transformation is meeting resistance. The present challenge for the industry is to deliver on its digital aspirations.

For many oil & gas professionals the focus on digital transformation has been overwhelming, Key individuals see digital transformation as being merely today's hype. They see challenges for implementation with little incremental benefit. Following an extended period of staff reductions, they struggle to see how they can cope with the additional work load.

In early 2019 we still see polls of oil & gas professionals where 50% of the participants don't know if their company makes use of big data analytics. However, this is an improvement on early 2018. What you also see within the industry is a growing group of change agents who have embraced digital transformation. These change agents will be the driving force that overcomes the resistance to change.

The digital opportunities that add value are similar for many companies across the industry. This means that replication benefits should be high. In fact, this replication is essential. We cannot create a bespoke digital project for every asset or application. Up-scaling of pilot projects must have a critical mass of replication not only within a specific company but also across the industry.

The main challenges for replication are (1) the quality of data, (2) the ability to integrate data (break down internal data silos) and (3) the commercial constraints around data and data platforms. These will be more evident as companies start to scale up their existing pilot projects or start to look at more ambitious opportunities such as full-asset digital twins. Complex digital twins in an offshore environment will require integration of subsurface, wells, sub-sea and production facility data, together with the coupling of physics-based models. To be successful, the present data landscape will need to be challenged and improved industry collaboration is essential.

Furthermore, some countries have restrictions on international data portability. Similarly, joint ventures often have legal restrictions on data sharing. This has a negative impact on cloud computing and cloud data storage and reduces the effectiveness of data analytics and international research. We need to engage with governments to loosen these portability restrictions.

Similarly, within joint ventures, partner companies need to take a global perspective and be prepared to review the restrictions within their Joint Venture Agreements. Companies need to evaluate present data restrictions with a digital transformation mind set, and decide precisely what data is essential to restrict and what can be shared with the participating partners, affiliate companies, and contractors/suppliers.

Significant additional effort is needed to deliver short to mid-term transformational gains. This will come at an additional cost and demand on critical staff within an organization.

However, if we cannot deliver a high level of replication and we continue with the present restrictions on data portability and transfer there is a risk that the additional cost will be greater that the digital benefits.

Digital transformation can be made less onerous and expensive through improved collaboration between companies, who work together to provide industry data standards and Joint Venture agreements that benefit all stakeholders. There is also a role for international industry representative bodies to engage with host governments with data portability restrictions and explain in detail the benefits of data portability. SIRIUS provides a forum for technical collaboration that can be used to develop these standards and enable effective data access.

We are living in exciting times, let's work together to deliver the oil and gas industries digital aspirations.

Brazil

SIRIUS has a key role in SINOS, Norway's common academic engagement with the oil and gas sector in Brazil. We provided the keynote address at the annual November conference in Rio de Janeiro in 2017 and chaired the programme committee for the same conference in 2018. This engagement is funded by the Ministry of Oil and Energy 2018-2020. From these foundations we have built a collaboration with the Federal University of Rio Grande do Sul (UFRGS), Petrobras, Shell and Total. This collaboration supplements the subsurface beacon projects and provides access to a leading group in petroleum informatics. Initial funding has been received from Educational funding organisations: DIKU and CAPES. We submitted one of five proposals received in the autumn 2018 joint FINEP-PETROMAKS call (UiO-UFRGS-Equinor-Petrobras). In November 2018 a delegation from the University of Oslo ran a joint seminar on Digital Geosciences at UFRGS. A PhD. student from UFRGS is visiting SIRIUS from October 2018-September 2019 and another from April 2019-March 2020.

Alan Turing Institute (ATI)

The Alan Turing Institute (ATI) is the United Kingdom's national centre for data science. SIRIUS participates in the ATI through the University of Oxford's partnership in the centre and through direct participation of SIRIUS researchers in the centre. In 2018 two SIRIUS researchers, Ernesto Jimenez-Ruíz and Evgenij Thorstensen have been visiting fellows at ATI.

Ernesto started at the ATI in April 2018, working in the Artificial Intelligence for Data Analytics (AIDA) project. Data analytics is the process of transforming a raw dataset into useful knowledge. By drawing on new advances in artificial intelligence and machine learning, this project is aiming to develop systems that will help to automate the data

Keith Lewis

Keith Lewis is a seasoned exploration and production professional with an International career spanning over 37 years in Brazil, USA, Africa, Europe, Asia and the Middle East. His recent assignments include Leader of Applied Technology for the Petrobras Operated Libra Ultra Deep-Water Project in Brazil.

Previous positions include Project Manager for the Subsea Well Response Project, a joint initiative of 9 major oil & gas companies, working together to enhance the industries capacity to better respond to International sub-sea well control incidents. Vice President for Shell Global Solutions Upstream in The Americas and Venture Director of Shell Nigeria Exploration & Production Company.

Evgenij Thorstensen

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, and discovering new mathematics that can then be applied to create better software.

analytics process. Data analytics comprises many different stages and phases. While some elements of the data analytics process have benefited from considerable development through software or tools, there has been little methodological research into so-called data' wrangling', even though this is often laborious and time-consuming, and accounts for up to 80% of a typical data science project. AIDA aims to fill this gap. Data wrangling includes understanding what data is available, integrating data from multiple sources, identifying missing, messy or anomalous data, and extracting features in order to prepare data for computer modelling.

Ernesto attended the International Semantic Web Conference (ISWC) in Monterrey (California) where he presented a poster and a workshop paper. He also co-organised the Ontology Matching workshop and the Ontology Alignment Evaluation Initiative, both collocated with ISWC. He is also supervising Erik Bryhn Myklebust, the PhD student in the Environmental Applications beacon. Erik plans to visit the Alan Turing Institute from March to May 2019.

European Private-Public Partnerships

The centre is an active participant in the European digitalization agenda through membership in two cPPP organizations: the Big Data Value Association (BDVA) and A.SPIRE (cPPP for the process industries). The Hybrid Autonomous Organization (HAO) structure in SIRIUS makes it necessary to engage in two fora in Europe – one related to IT vendors, the other to IT users. This dual role has been recognised and SIRIUS will be working in 2019 in a role that seeks to bridge the common interests between BDVA and A.SPIRE.

European Projects

SIRIUS' research and innovation program built on three EU projects that concluded in 2016 or 2017: Optique, Envisage and HyVar. A follow up to Optique was proposed in the ICT15 call in 2016, but did not fit the profile of funded projects. We participated in an MSCA call in early 2018, again without success, although this proposal, without us, but with the CIRFA SFI, succeeded in an H2020 call later in 2018. In late 2018 we were invited by SINTEF to participate in the DIGITWIN proposal to the H2020 ICT program on internet of things.

Geir Horn, the leader of SIRIUS' Scalable Computing research program, is also coordinator of a H2020 ICT project called MELODIC (ICT-731664). It is planned that SIRIUS will provide data access use cases that will demonstrate the applicability of the cloud infrastructure developed here.

Ernesto Jimenez Ruiz

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.

Recruitment, Education and Equal Opportunity

Researcher Training in Research Programs

Researcher training at the Ph.D. level is organized by research programs. Each SIRIUS stipend holder works in a specific discipline towards a problem defined by the centre's partners. Stipend holders are offered a one-week residential course on the oil and gas industry. This is provided as an in-kind contribution by Schlumberger. This has been held once, in late 2016 and will be run again in 2019.

DataScience@UiO

In 2017 SIRIUS, together with the BigInsight SFI, applied to the Faculty for Mathematics and Natural Sciences for four innovation stipends under the heading DataScience@UiO. This successful application resulted in two stipend holders in SIRIUS and two in BigInsight. These students participate in a network with regular meetings and support a program assisting other researchers who want to use data science in their research.

Mentoring Program

The SIRIUS mentoring program is a vehicle for increasing industry-academia collaboration. It aims to:

- Offer individual SIRIUS researchers a personal development strategy and to train future research leaders in academia or industry.
- Shape SIRIUS center identity by exchanging expertise, values, skills, perspectives, attitudes, and through networking.
- Increase mutual understanding between SIRIUS companies and academia to foster collaboration, engagement, and build career competence on both sides of the mentoring relationship.
- Promote and harness the full potential of diversity in SIRIUS: in culture, age, gender, and expertise.

The first mentoring program was run for SIRIUS in 2017 and 2018. This program matched ten PhD students and junior researchers from the University of Oslo and NTNU (6 women and 4 men) with mentors from partner companies Equinor, Schlumberger, IBM, Evry, and Numascale. The program focused on building a common center identity and engaging the mentors in two ways. First, it created an arena for both the individuals and the group of mentees, contributing to career development and opportunities beyond SIRIUS. Second, the program formed a shared understanding between the industrial and academic partners, creating a foundation for innovation in SIRIUS.

Some of those who work with DataScience@UiO: From left to right: Sylvia Liu, Arnoldo Frigessi, Ida Scheel, Ingrid Glad, Evgenij Thorstensen and Daniel Bakkelund (Photo: Gunhild M. Haugnes/UiO).

This shared understanding was built through regular meetings between mentor/mentee pairs, seminars and business visits for the mentees. These activities were thoroughly in line with the program's objectives. Overall the program has exceeded expectations.

During the closing seminar in April 2018 the mentor/-mentee pairs presented their experiences over the last year. Both mentors and mentees have stated that the program was both rewarding and useful. The mentees pointed out that they gained personal development and awareness, expanded their network and better understood industry culture. The mentors increased their understanding of the relationship between academia and industry, obtained new perspectives and built new relationships. They feel that the program has strengthened the partnership between the company and SIRIUS.

With this positive feedback and experience we are looking forward to start the next program in the second half of 2019. We will send out invitations to our partners to sign up as mentors for a new group of PhD students and researchers, starting with a Mentor Master Class in late August.

Masters' Program

SIRIUS offered master's projects in 2017-2018, but a focus on master's level education has started from the 2018-2019 academic year. This program is also supported by awards from DIKU (formerly SiU) that support master's students to be interns at Petrobras in Rio de Janeiro and/or work on their project at the Federal University of Rio Grande do Sul. This work is integrated with the Exploration beacon projects.

Gender Perspectives

The centre has good gender balance in the stipend holders and post-doctoral students. The mentor program had a positive effect on developing female researchers, especially through the involvement of senior women in the oil and gas industry as mentors.

Female members at SIRIUS also benefited from the FRONT project, organized by the Faculty of Mathematics and Natural Sciences at UiO, with aims to improve the representation in scientific and leadership positions within the Faculty through trainings in leadership and organisational development.

In 2018, SIRIUS women's club was established. It is an internal initiative at the centre where female researchers have regular informal gatherings to strengthen the female network at SIRIUS.

Undergraduate Researcher Track

An undergraduate researcher track, "Forskerskolen", has proven highly successful to recruit students into research at the Faculty of Medicine. The Department of Informatics has set up a national pilot for a related program in informatics, recruiting bachelor students into a structured program built around research in SIRIUS and the HISP project **(https://www.mn.uio.no/ifi/english/research/networks/hisp/).**

Theses Defended in 2018

Anthony Potter

Anthony Potter, a SIRIUS stipend holder from the University of Oxford, defended his DPhil. thesis in early 2018. His dissertation was entitled "Query Answering in Distributed RDF Databases". The project was supervised by Professor Boris Motik.

To simplify data integration and exchange, modern applications often represent their data using the Resource Description Framework (RDF). As the amount of the available data keeps increasing, many RDF datasets cannot be processed using centralised RDF stores. A common solution is to distribute RDF data in a cluster of shared-nothing servers, and to query the data using a distributed query algorithm. Existing approaches typically use a variant of the data exchange operator to shuffle partial query answers between servers and thus ensure that every query answer is produced. Decisions as to when and where to shuffle the data are usually made statically – that is, when a query is compiled. Potter argues that such approaches

Lise Reang

Lise Reang is Administration Manager in SIRIUS. She has a varied background within administration, coordination and service. She worked nine years for the Kongsberg Group, first as a service co-ordinator for collaboration environments, then as a Site Manager / conference coordinator and finally as management assistant for executive management in Oil & Gas. Lise worked for AS Vinmonopolet for six years where she established, managed and built the internal IT helpdesk. Lise's main tasks in the centre are to facilitate collaboration between colleagues and contribute to a stimulating work environment and good communications. She assists and guides in administrative questions and helps to plan and run different kinds of events.

The Mentoring Program and Centre Management at the final meeting of the Program. From left to right: David Cameron, Peter Eilsø Nielsen (Equinor), Farhad Nooralahzadeh, Jacopo Mauro, Ann Vølstad (AFF), Elisabeth Østrem (AFF), Lise Reang, Arild Waaler, Crystal Chang Din, Lizeth Tapia Tarifa, Frode Myren (IBM), Ingrid Chieh Yu, Åshild Hanne Larsen (Equinor), Ernesto Jimenez-Ruiz, Johanne Beate Stumpf, Siri Bentsen (Equinor), Einar Landre (Equinor), Violet Ka I Pun, Per Eivind Solum (Schlumberger), Anastasia Gkolfi, Derek Lacey (IBM), Elena Parmiggiani, Tom Erling Henriksen (Evry), Øystein Haaland (IBM). Vidar Klungre and Einar Rustad (Numascale) were not present.

can miss opportunities for local computation and thus incur considerable overheads. Moreover, he presents a novel distributed query evaluation algorithm for RDF based on dynamic data exchange, where all computation that can be done locally is guaranteed to be performed on a single server. This approach can successfully process any query even if the memory available at each server is bounded, and we argue that this is critical in distributed systems where intermediate results can easily exceed the capacity of each server. He also presents a new query planning approach that balances the cost of communication against the cost of local processing at each server, as well as a new approach to partitioning RDF data that aims to increase locality in each server. The approach has been implemented in the RD-Fox data base, and empirical evaluation suggests that our techniques can outperform the state of the art by orders of magnitude in terms of query evaluation times, network communication, and memory use.

Johanna Beate Stumpf

On October 11th 2018, Johanna Beate Stumpf from the University of Oslo defended her dissertation for the degree of PhD. The title of the thesis was "Virtually Timed Ambients: A Calculus for Resource Management in Cloud Computing". The adjudication committee was Professor Uwe Nestmann (Electrical Engineering and Computer Science, Technische Universität Berlin), Associate Professor Anna Philippou, (Department of Computer Science, University of Cyprus) and Associate Professor Jon Henrik Forssell (Department of Informatics, University of Oslo). Johanna was supervised by Professor Einar Broch Johnsen and Professor Martin Steffen. Her work looked at developing a method of formally modelling analysing the behaviour of processes running on virtual machines running on a cloud computing.

Cloud computing is a paradigm of distributed computing in which users share resources by storing data and executing processes in common data centres. A key factor for the success of this paradigm is virtualization technology, which represents the resources of an execution environment as a software layer, a so-called virtual machine. Virtualization allows sharing of existing hardware and software resources, improves security by providing isolation of different users who share the same resource, and enables dynamic assignment of resources according to the demands of the user. This sharing of resources creates business drivers which make cloud computing an economically attractive way to deploy software.

Johanna introduces the calculus of virtually timed ambients: a formal model of hierarchical locations for execution with explicit resource provisioning. This calculus is based on the well-known calculus of mobile ambients and is motivated by the use of nested virtualization in cloud computing applications. The investigation of cloud computing from the point of view of process calculi provides a formal

specification of the subject, which is necessary in order to develop executable models for analysis and optimization.

The main contributions of the work are the definition of the calculus of virtually timed ambients, and the reasoning about its essential characteristics. In order to enable static analysis, we enhance the calculus with a type system. Furthermore, she defines a modal logic and a corresponding model checker, which is deployed in the definition of resource-awareness of virtually timed ambients, enabling dynamic self-management of processes. Lastly, she presents virtually timed ambients as a framework to analyse virtualization in cloud computing utilizing a prototype implementation. All concepts are illustrated by examples.

Leif Harald Karlsen

On November 20th, 2018, Leif Harald Karlsen defended his dissertation for the degree of Ph.D. The title of the dissertation was "A simple, General and Efficient Representation of Qualitative Spatial Information; An Approach Based on Bintrees". The adjudication committee consisted of Professor Anthony Cohn (School of Computing, University of Leeds, UK), Professor Hanan Samet (Computer Science Department, University of Maryland, USA) and Professor Vera Hermine Goebel (Department of Informatics, University of Oslo). Leif Harald was supervised by Professor Martin Giese and Professor Arild Waaler. This worked looked at new ways in which computers can represent and search for spatial data in databases. The approach Leif Harald used was to translate complicated spatial objects into simpler structures, bintrees, in a way that the qualitative information in the data is maintained. This allows faster and better access to this data.

Spatial data is used in a great number of highly valuable applications, like route planning, automatic navigation and modelling of physical processes. However, on a computer, spatial data is often represented as complex numerical objects, and therefore requires advanced numerical algorithms to process. Despite this numerical complexity, we humans tend to think of spatial data in qualitative rather than quantitative terms. Motivated by this observation, Leif Harald's work aims to improve the efficiency and applicability of answering queries involving qualitative relationships over spatial data. The novelty of the work is the development of an efficient algorithm that translates spatial objects into simpler structures, known as bintrees, in such a way that qualitative relationships are preserved. This simpler structure uses less storage space on disk compared to the original spatial data, and allows qualitative queries to be answered more efficiently. The thesis contains both a theoretical treatment of the problem based on logic, and implementation with experimental evaluation of the overall approach.

Governance

Operations Board / Centre Management

SIRIUS' General Assembly is the body that makes final decisions in the centre. It consists of one high-level representative from each partner. It held two meetings in 2018: one in on 23rd and 24th May, hosted by DNV GL at Høvik and the second on 30th October, hosted by IBM in Oslo. This year we have started to have these meetings hosted by partners. This reduces the cost of the meeting and allows closer contact between the centre and the hosting partner. We will continue with this practice in 2019.

As usual, the spring meeting focused on centre strategy and results, with a two-day program and poster sessions that allowed researchers and industry representatives to discuss their work. The autumn meeting was a business meeting to review and approve the work plan for 2019.

There has been one extraordinary General Assembly vote this year. This approved TechnipFMC joining as partner.

Executive Committee of the General Assembly

The Executive Committee of the General Assembly held six meetings in 2018. This group was formed as a result of a site visit by the Research Council in 2017. The executive committee acts on behalf of the General Assembly to exercise regular oversight over plans and progress for the centre.

This committee in 2018 consisted of:

- Knut Sebastian Tungland, Equinor, Chairman.
- Edo Hoekstra, Schlumberger, with Per Eivind Solum, Schlumberger as deputy.
- Frode Myren, IBM, with Nicolas Peels, then Ivan Mudron, OSIsoft as deputy.
- Einar Rustad, Numascale, with Hugo Kohmann, Dolphin as deputy.
- Eric Monteiro, NTNU, with Thomas Østerlie, NTNU as deputy.
- Arild Waaler, University of Oslo, with Einar Broch Johnsen, University of Oslo, as deputy.

A new committee will be chosen at the spring General Assembly in 2019.

Strategy Board

The Strategy Board defines the strategic plans for the SIRIUS centre and approves the projects in the SIRIUS project portfolio. Its members in 2018 were:

- Arild Waaler, University of Oslo, Chairman,
- Ian Horrocks, University of Oxford, Scientific Coordinator,
- Einar Broch Johnsen, University of Oslo, WP5 Leader,
- Eric Monteiro, NTNU, Pilots Coordinator,
- Geir Horn, University of Oslo, Faculty Research Strategist

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 roie and composition of the Operations Board will be reviewed in 2019.

Roles (from 2019)

The Centre made some organizational changes at the end of 2018. These are reflected in the following table. Changes from 2019 are as follows:

- Ingrid Chieh Yu becomes Deputy Centre Director
- The role of IPR Manager has been removed. This responsibility is taken by the centre coordinator.
- Work package leaders now have responsibility for the beacon projects in their work packages and delegate this to a leader for each beacon.
- Strand leaders have been replaced by leaders for research programs.

Roles

Role	Name	Affiliation
Chairman of General Assembly	Knut Sebastian Tungland	Equinor ASA
Centre Director	Arild Waaler	University of Oslo
Deputy Centre Director	Ingrid Chieh Yu	University of Oslo
Centre Coordinator	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
Exploration Beacons	Adnan Latif	University of Oslo
Operations Beacons	David Cameron	University of Oslo
Cross-Domain Beacons	Laura Slaughter	University of Oslo
Strategy and Outreach	Einar Broch Johnsen	University of Oslo
Analysis of Complex Systems	Lizeth Tapia Tarifa	University of Oslo
Ontology Engineering	Martin G. Skjæveland	University of Oslo
Data Science	Basil Ell	University of Oslo
Semantic Integration	Medha Atre	University of Oxford
Scalable Computing	Geir Horn	University of Oslo
Industrial Digital Transformation	Thomas Østerlie	University of Oslo

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 A 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.

Senior Staff		
Xing Cai	Simula Research	Scalable Computing
Martin Giese	University of Oslo	Data Science
Geir Horn	University of Oslo	Scalable Computing
Ian Horrocks	University of Oslo	Semantic Integration, Ontology Engineering
Einar Broch Johnsen	University of Oslo	Analysis of Complex Systems
Evgeny Kharlamov	University of Oslo	Ontology Engineering
Boris Motik	University of Oxford	University of Oslo
Eric Monteiro	NTNU	Industrial Digital Transformation
Tor Skeie	Simula Research	Scalable Computing
Laura Slaughter	University of Oslo	Ontology Engineering
Arild Waaler	University of Oslo	Ontology Engineering
Ingrid Chieh Yu	University of Oslo	Ontology Engineering

Research, Technical and Administrative Personnel		
Medha Atre	University of Oxford	Semantic Integration
David Cameron	University of Oslo	Centre Coordinator
Jiaoyan Chen	University of Oxford	Data Science
Crystal Chang Din	University of Oslo	Analysis of Complex Systems
Basil ell	University of Oslo/Universität Bielefeld	Data Science
Dag Hovland	University of Oslo	Semantic Integration
Ernesto Jimenez-Ruiz	University of Oslo	Data Science
Jia-Chun (Kelly) Lin	University of Oslo	Analysis of Complex Systems
Jacopo Mauro	University of Oslo	Analysis of Complex Systems
Elena Parmiggiani	NTU	Industrial Digital Transformation
Violet Pun	University of Oslo	Analysis of Complex Systems
Rudolph Schlatte	University of Oslo	Analysis of Complex Systems
Martin G. Skjæveland	University of Oslo	Ontology Engineering
Oliver Stahl	University of Oslo	SIRIUS Laboratory
Lizeth Tapia	University of Oslo	Analysis of Complex Systems
Evgenij Thorstensen	University of Oslo	Industrial Digital Transformation
Thomas Østerlie	NTNU	Industrial Digital Transformation

Publications in 2018

Agibetov, Asan; Jimenez-Ruiz, Ernesto; Ondrésik, Marta; Solimando, Alessandro; Banerjee, Imon; 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 2018; Volume 9:9., 1-22

Albert, Elvira; Correas, Jesus; Johnsen, Einar Broch; Ka I, Pun; Roman-Diez, Guillermo. Parallel Cost Analysis. ACM Transactions on Computational Logic 2018; Volume 19(4), 1-37

Benzmüller, Christoph; Otten, Jens. Proceedings of the 3rd International Workshop on Automated Reasoning in Quantified Non-Classical Logics. CEUR Workshop Proceedings 2018; Volume 2095

Bijo, Shiji; Johnsen, Einar Broch; Pun, Ka I Violet; Seidl, Christoph; Tapia Tarifa, Silvia Lizeth. Deployment by Construction for Multicore Architectures. I: Leveraging Applications of Formal Methods, Verification and Validation. Modeling - 8th International Symposium, ISoLA 2018. Proceedings - Part I. Springer 2018, 448-465

Cameron, David B.; Waaler, Arild; Komulainen, Tiina M. Oil and Gas digital twins after twenty years. How can they be made sustainable, maintainable and useful? The 59th Conference on Simulation and Modelling (SIMS 59); Oslo, 2018-09-26 - 2018-09-28

Cameron, David B.; Waaler, Arild; Komulainen, Tiina M. Oil and Gas digital twins after twenty years. How can they be made sustainable, maintainable and useful? Linköping Electronic Conference Proceedings 2018 (153) 9-16

Cameron, David B.; Østerlie, Thomas. Sirius: Collaboration across the digital divides in the oil and gas supply chain. 2018 University Industry Interaction Conference; London, 2018-06-20 2018-06-22

Chang Din, Crystal; Schlatte, Rudolf; Chen, Tzu-Chun. Program Verification for Exception Handling on Active Objects Using Futures. Lecture Notes in Computer Science 2018; Volume 10886, 73-88

Chen, Jiaoyan; Lécué, Freddy; Pan, Jeff Z.; Horrocks, Ian; Chen, Huajun. Knowledge-Based Transfer Learning Explanation. Sixteenth International Conference on Principles of Knowledge Representation and Reasoning; 2018-10-30 - 2018-11-02

Din, Crystal Chang; Johnsen, Einar Broch; Owe, Olaf; Yu, Ingrid Chieh. A modular reasoning system using uninterpreted predicates for code reuse. Journal of Logical and Algebraic Methods in Programming 2018; Volume 95, 82-102

Gkolfi, Anastasia; Johnsen, Einar Broch; Kristensen, Lars Michael; Yu, Ingrid Chieh. Using coloured Petri nets for resource analysis of active objects. Lecture Notes in Computer Science 2018; Volume 11222 LNCS, .156-174

Hähnle, Reiner; Johnsen, Einar Broch. A Model-Centric Approach to the Design of Resource-Aware Cloud Applications. I: Software Technology: 10 Years of Innovation in IEEE Computer. Wiley-Blackwell 2018, 315-326

Hu, Pan; Motik, Boris; Horrocks, Ian. Optimised Maintenance of Datalog Materialisations. Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18); 201802-02 - 2018-02-07

Johnsen, Einar Broch; Steffen, Martin; Stumpf, Johanna Beate. Virtually timed ambients: A calculus of nested virtualization. Journal of Logical and Algebraic Methods in Programming 2018; Volume 94. 109-127

Johnsen, Einar Broch; Steffen, Martin; Stumpf, Johanna Beate; Tveito, Lars. An Analysis Tool for Models of Virtualized Systems. NIK: Norsk Informatikkonferanse, Svalbard, 2018

Johnsen, Einar Broch; Steffen, Martin; Stumpf, Johanna Beate; Tveito, Lars. Checking Modal Contracts for Virtually Timed Ambients. Lecture Notes in Computer Science 2018; Volume 11187 LNCS. 252-272

Johnsen, Einar Broch; Steffen, Martin; Stumpf, Johanna Beate; Tveito, Lars. Resource-Aware Virtually Timed Ambients. Lecture Notes in Computer Science 2018; Volume 11023, 194-213

Kharlamov, Evgeny; Skjæveland, Martin G; Hovland, Dag; Mailis, Theofilos; Jimenez-Ruiz, Ernesto; Xiao, Guohui; Soylu, Ahmet; Horrocks, Ian; Waaler, Arild. Finding Data Should be Easier than Finding Oil. I: 2018 IEEE International Conference on Big Data (Big Data), Seattle, 10-13 Dec. 2018.

Kindermann, Christian; Lupp, Daniel P.; Sattler, Ulrike; Thorstensen, Evgenij. Generating Ontologies from Templates: A Rule-Based Approach for Capturing Regularity. CEUR Workshop Proceedings 2018; Volume 2211.

Lee, Ming-Chang; Lin, Jia-Chun; Owe, Olaf. EasyChoose: A Continuous Feature Extraction and Review Highlighting Scheme on Hadoop YARN. Advanced Information Networking and Applications 2018; Volume 2018-May, 996-1002

Lee, Ming-Chang; Lin, Jia-Chun; Owe, Olaf. Privacy Mining from IoT-based Smart Homes. Advances on Broad-Band Wireless Computing, Communication and Applications 2018, 304-315

Lin, Jia-Chun; Lee, Ming-Chang; Yu, Ingrid Chieh; Johnsen, Einar Broch. Modeling and simulation of spark streaming. Advanced Information Networking and Applications 2018; Volume 2018-May. 407-413

Lupp, Daniel Paul; Karlsen, Leif Harald; Skjæveland, Martin G. Making a Case for Formal Relations over Ontology Patterns. CEUR Workshop Proceedings 2018; Volume 2195. 87-91

Lupp, Daniel Paul; Thorstensen, Evgenij. Mapping Data to Ontologies with Exceptions Using Answer Set Programming. NIK: Norsk Informatikkonferanse, Svalbard, 2018

Mauro, Jacopo; Tapia Tarifa, Silvia Lizeth; Yu, Ingrid Chieh. Automatic Parameter Optimisation of Service Quality and Resource Usage. NIK: Norsk Informatikkonferanse, Svalbard, 2018

Monteiro, Eric. Reflections on digital innovation. Information and organization 2018; Volume 28(2), 101-103

Monteiro, Eric; Østerlie, Thomas; Parmiggiani, Elena; Mikalsen, Marius. Quantifying quality: Towards a Post-Humanist Perspective on Sensemaking. I: Living with Monsters? Social Implications of Algorithmic Phenomena, Hybrid Agency, and the Performativity of Technology. IS&O 2018. IFIP Advances in Information and Communication Technology, Volume 543, Springer, 2018, 48-63

Nieke, Michael; Mauro, Jacopo; Seidl, Christoph; Thum, Thomas; Yu, Ingrid Chieh; Franzke, Felix. Anomaly Analyses for Feature-Model Evolution. I: Proceedings of the 17th ACM SIGPLAN International Conference on Generative Programming: Concepts and Experiences. ACM Publications 2018, 188-201

Nooralahzadeh, Farhad; Øvrelid, Lilja. SIRIUS-LTG: An Entity Linking Approach to Fact Extraction and Verification. I: Proceedings of the First Workshop on Fact Extraction and Verification (FEVER). Association for Computational Linguistics 2018, 119-123

Nooralahzadeh, Farhad; Øvrelid, Lilja. Syntactic Dependency Representations in Neural Relation Classification. I: Proceedings of the Workshop on the Relevance of Linguistic Structure in Neural Architectures for NLP. Association for Computational Linguistics, 2018

Nooralahzadeh, Farhad; Øvrelid, Lilja; Lønning, Jan Tore. Evaluation of Domain-specific Word Embeddings using Knowledge Resources. I: Proceedings of the Eleventh International Conference on Language Resources and Evaluation. European Language Resources Association 2018, 1438-1445

Nooralahzadeh, Farhad; Øvrelid, Lilja; Lønning, Jan Tore. SIRIUS-LTG- at SemEval-2018 Task 7: Convolutional Neural Networks with Shortest Dependency Paths for Semantic Relation Extraction and Classification in Scientific Papers. I: Proceedings of the 12th International Workshop on Semantic Evaluation. Association for Computational Linguistics 2018, 805-811

Otten, Jens. Proof search optimizations for non-clausal connection calculi. CEUR Workshop Proceedings 2018; Volume 2162, 49-57

Otten, Jens. The Pocket Reasoner - Automatic Reasoning on Small Devices. NIK: Norsk Informatikkonferanse, Svalbard, 2018

Parmiggiani, Elena; Monteiro, Eric. Shifting Baselines? Recommendations for Green IS. I: Proceedings of the 39th International Conference on Information Systems (ICIS). Association for Information Systems 2018, 1-16

Pinkel, Christoph; Binning, Carsten; Jimenez-Ruiz, Ernesto; Kharlamov, Evgeny; May, Wolfgang; Nikolov, Andriy; Bastinos, Ana Sasa; Skjæveland, Martin G; Solimando, Alessandro; Taheriyan, Mohsen; Heupel, Christian; Horrocks, Ian. RODI: Benchmarking Relational-to-Ontology Mapping Generation Quality. Semantic Web Journal 2018; 9(1) 25-52

Potter, Anthony; Motik, Boris; Nenov, Yavor; Horrocks, Ian. Dynamic Data Exchange in Distributed RDF Stores. IEEE Transactions on Knowledge and Data Engineering 2018; 30(12), 2312-2325

Ronca, Alessandro; Kaminski, Mark; Grau, Bernardo Cuenca; Horrocks, Ian. The Window Validity Problem in Rule-Based Stream Reasoning. Sixteenth International Conference, KR 2018; 2018-10-30 - 2018-11-02

Ronca, Alessandro; Kaminski, Mark; Grau, Bernardo Cuenca; Motik, Boris; Horrocks, Ian. Stream Reasoning in Temporal Datalog. Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18); 2018-02-02 - 2018-0207

Presentations in 2018

Cameron, David B. Digitalization perspectives from the oil and gas domain. Digital Ship Maritime Big Data Forum; Oslo, 2018-05-29

Cameron, David B. Scalable, Useful and Maintainable Digital Twins: Cross-Sector Experience from the Oil and Gas Sector. Digital Twin technology in the steel industry: from concept to operational benefits; Charleroi, Belgium, 2018-11-21 - 2018-11-22

Latif, Adnan; Waaler, Arild; Giese, Martin; Hovland, Dag; Skjæveland, Martin G. Ontology-Based Data Access in Exploration. 6th November Conference for Norwegian/ Brazilian Energy Research; Rio de Janeiro, 2018-11-12 - 2018-11-13

Monteiro, Eric. Organizational and human aspects of digitalization in Oil and Gas. 6th November Conference for Norwegian/ Brazilian Energy Research; Rio de Janeiro, 2018-11-12 - 2018-11-13

Otten, Jens. Bibel's Connection Method – A General Framework for Automated Reasoning. Artificial Intelligence – Scientific Symposium on the Occasion of Wolfgang Bibel's 80th Birthday; 2018-11-23

Annual Accounts

Costs

All figures in 1000 NOK	**2015**	**2016**	**2017**	**2018**
Personnel and indirect costs	539	5188	10087	12450
Purchase of research services	-	600	2113	6232
Equipment	-	31	122	44
Other operational costs	62	9505	10424	16354
Total Sum	**601**	**15324**	**22746**	**35080**

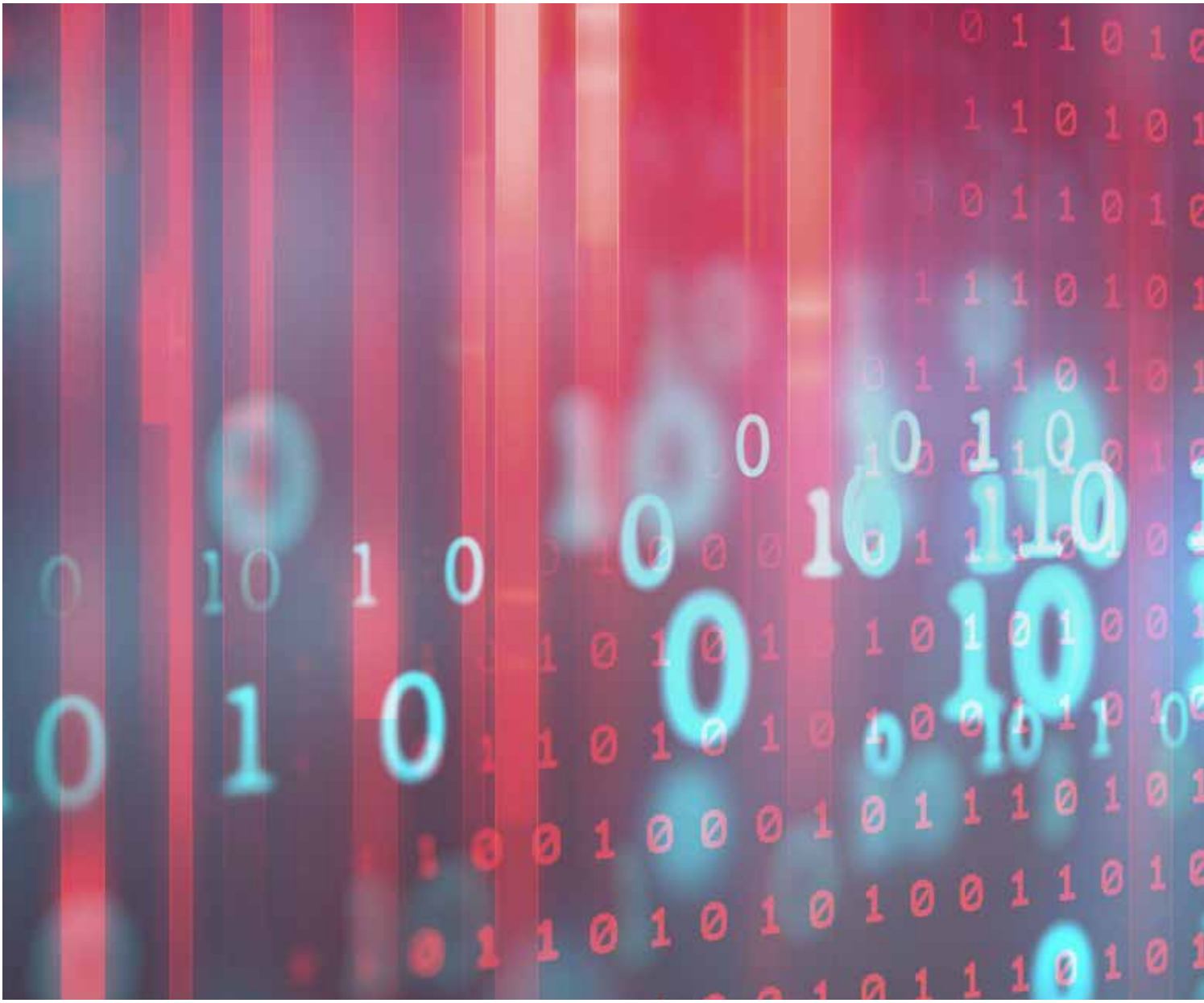
Funding

All figures in 1000 NOK	**2015**	**2016**	**2017**	**2018**
Research Council	-	4168	8180	12430
University of Oslo	601	2141	5593	5604
Public partners	-	144	813	1144
Private partners	-	8510	7910	15130
International partners		361	250	772
Total Sum	**601**	**15324**	**22746**	**35080**

Photo credits
iStock
Shutterstock
Equinor
Basil Ell, Susann Dietrich
and VG Bild-Kunst Bonn, 2018.

SIRIUS

ANNUAL REPORT 2018



Sirius-labs.no