SIRIUS ANNUAL REPORT 2019

Centre for Research-based Innovation
The Research Council of Norway

UiO : University of Oslo

Contents

SIRIUS aims to be an intellectual hub for applied industrial IT in Norway

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 does interdisciplinary research that leads to innovations that advance and support digitalisation in the oil & gas industry.

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 & 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.

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 energy value chain: including operators (Equinor), service companies (Schlumberger and DNV GL) and IT companies (Computas, Evry, Dolphin Interconnect Solutions, Operations AG, IBM, Kadme, Numascale, OSISoft and SAP). In 2019 Aibel, Aker Solutions and Bosch joined as partners. These companies work with researchers from the University of Oslo, NTNU, SINTEF, University of Oxford and Simula Research Laboratories.

SIRIUS researchers bring expertise in:

- **knowledge representation**
- **natural-language technologies**
- **databases**
- **scalable computing**
- **execution modelling and analysis**
- **working practices**

The centre provides the industry with better ways to access and use the massive amounts of data that are generated in exploration, projects and daily operations. The advent of big data, digitalisation, the internet of things and data science has made problems with data access more acute. Data access is the bottleneck that can prevent successful digitalisation. 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 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 started in November 2015 and has now finished four full years of work. 2019 saw a consolidation of the centre. We successfully completed the mid-term review process, with good results. Our financing for the whole eight years was confirmed. Four new partners joined the centre: Aibel, Aker Solutions, Bosch and SINTEF. Through this, we improve our coverage of the supply chain, with new engineering companies and equipment manufacturers. SINTEF has joined the centre to improve our European focus and find shorter ways to commercial innovation. SIRIUS was awarded a substantial Petromaks-FINEP Brazil project in digital twins. Our research continues to be recognized as excellent, with two best paper awards, several nominations and tutorials at world and European conferences. We have started collaborations in Europe, Brazil and Australia.

This report gives an overview of the centre's ambitions, scientific programme and activities for 2019.

Removing the Barriers that Block Innovation

SIRIUS is an endeavour where academic researchers join forces with specialists from partner companies to solve challenging problems in the digital transformation of the industry. There are various barriers between the academic research institutions and the industrial partner companies. Our mission can only be achieved by applying consistent effort to remove these barriers.

One barrier is the difference in culture and mentality between research institutions and companies. For example, the time span of projects is very different. Companies require project results in three months, while a PhD project takes three years. We need to find ways of linking these two timescales. A key to removing this barrier is to build relations and create mutual understanding. SIRIUS has implemented several measures to this end, in particular our mentoring program.

A second barrier is a difference in incentives. To build their careers, researchers need to publish. There are few incentives to develop techniques and prototypes beyond the minimum needed for publication. However, the maturity of the results is usually too low for practitioners who cannot adopt immature tools. We remove this barrier by using our public funding to reduce the risk of companies in joint research projects. We also finance a fourth year for our PhD fellows so that they can work on making their research results fit for innovation. SIRIUS also organises partner-led developments as joint industry projects, where risk is shared among several partners.

A final barrier lies in the differing attitudes of end users and IT experts to digitalisation. Digital transformation is first and foremost about a change of practice: it is something the end users own. A key to improved digitalisation is to allow end users to get control over their own data and give them more knowledge of the opportunities that digital technologies can bring. This is the heart of SIRIUS and we contribute in many ways, for example by developing standards-based technologies, by delivering research in domain adapted data science and in implementing strategies for transitioning from document-oriented work practices to digitally empowered workflows.

A key to improved digitalisation is to allow end users to get control over their own data and give them more knowledge of the opportunities that digital technologies can bring. This is the heart of SIRIUS and we contribute in many ways to achieve this objective.

All these barriers exist for good reasons, but through consistent work we will be able to overcome them. This is very much what research-based innovation is about. The value of achieving this is particularly significant for digitalisation. Digitalisation is a program of change that challenges established value chains and can bring problem owners much closer to the curiosity and the bright ideas of students and young researchers.

Arild Waaler

Arild Waaler is the director of SIRIUS. He has over the last decade been a driving force for establishing collaboration between academic research and the 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 the University of Oxford, in logic, and at the Søren Kirkegaard 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 the University of Oslo in 2006, when he started to build up the collaboration that eventually became SIRIUS.

Cultivating Collaboration Across Boundaries

2019 was a fruitful year for SIRIUS. The Faculty of Mathematics and Natural Sciences at the University of Oslo recognizes SIRIUS as being a central component in its research strategy. Our ongoing research effort in digitalsation of exploration and operations, in and beyond the oil & gas industry, demonstrates that we can bring together different IT disciplines to address core issues in making digitalisation successful across knowledge-intensive business sectors. During the past year, SIRIUS also successfully passed its mid-term evaluation by the Research Council, and as a result, the funding for its final three years of the project (2021 - 2023) was confirmed.

Our research programs in analysis of complex systems, data science, industrial digital transformation, ontology engineering, scalable computing and semantic integration continue to make advances in theory and tool development. Our researchers are actively engaged in the leading international communities. For example, we organized the 32nd International Workshop on Description Logistics in Oslo and the Semantic Web Challenge at the International Semantic Web Conference in Auckland. Our researchers have received paper awards in various conferences.

SIRIUS is expanding. We welcomed both new researchers and exchange students in 2019. We are looking forward to enlarging the SIRIUS team even further through newly funded research projects.

We have taken several steps to increase engagement and strengthen the centre's culture. We renewed our video-conference equipment and we can now hold multi-site lectures, seminars and conferences. The SIRIUS mentoring program continues its success, with ten new mentee-mentor pairs. We continue to be relevant for our industrial partners. We became the first academic partner in the Open Subsurface Data Universe initiative. We have presented our work at Schlumberger's SIS Global Forum. We worked with DNV GL to co-host the Industrial Ontologies Foundry. We co-organized and led the annual Conference on Norwegian-Brazilian Energy Research. In addition, we have held tutorials on our tools and methods at industrial meetups.

SIRIUS now has four more years to run. In this time, we will work to better define the impact of our research even better. Each research program now has a roadmap that covers these four years. We continue to pursue interdisciplinary research. In these ways, we will create value for our industrial partners while maintaining the high scientific quality of our research. Our research programs, under the leadership of our new program leaders, are seeking ways for long-term research in close interaction with the innovation ambitions of our partners. In this, we

Ingrid Chieh Yu

Ingrid Chieh Yu is the Deputy Director of SIRIUS and 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 also the Mentoring and Education Coordinator of SIRIUS. She was one of the Principal investigators of CUMULUS, an 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.

Going forward, understanding our research impacts, executing roadmaps and increasing our ability to pursue interdisciplinary research will be key elements in creating value for our industrial partners while maintaining high scientific quality in our research.

will be working together with the university's technology transfer organization and leading European research groups and innovation centres.

As we enter 2020, SIRIUS continues to evolve as we adapt and refine our collaboration model and continue to chart a path forward.

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 the Ole Johan Dahl's building in Oslo. The Faculty of Mathematics and Natural Sciences has four thematic initiatives: (i) life sciences, (ii) earth and space, (iii) energy and materials and (iv) digitalisation and computational science. SIRIUS is a core element in the university's digitalisation initiative and also contributes to the other three thematic areas.

SIRIUS at Half Time:

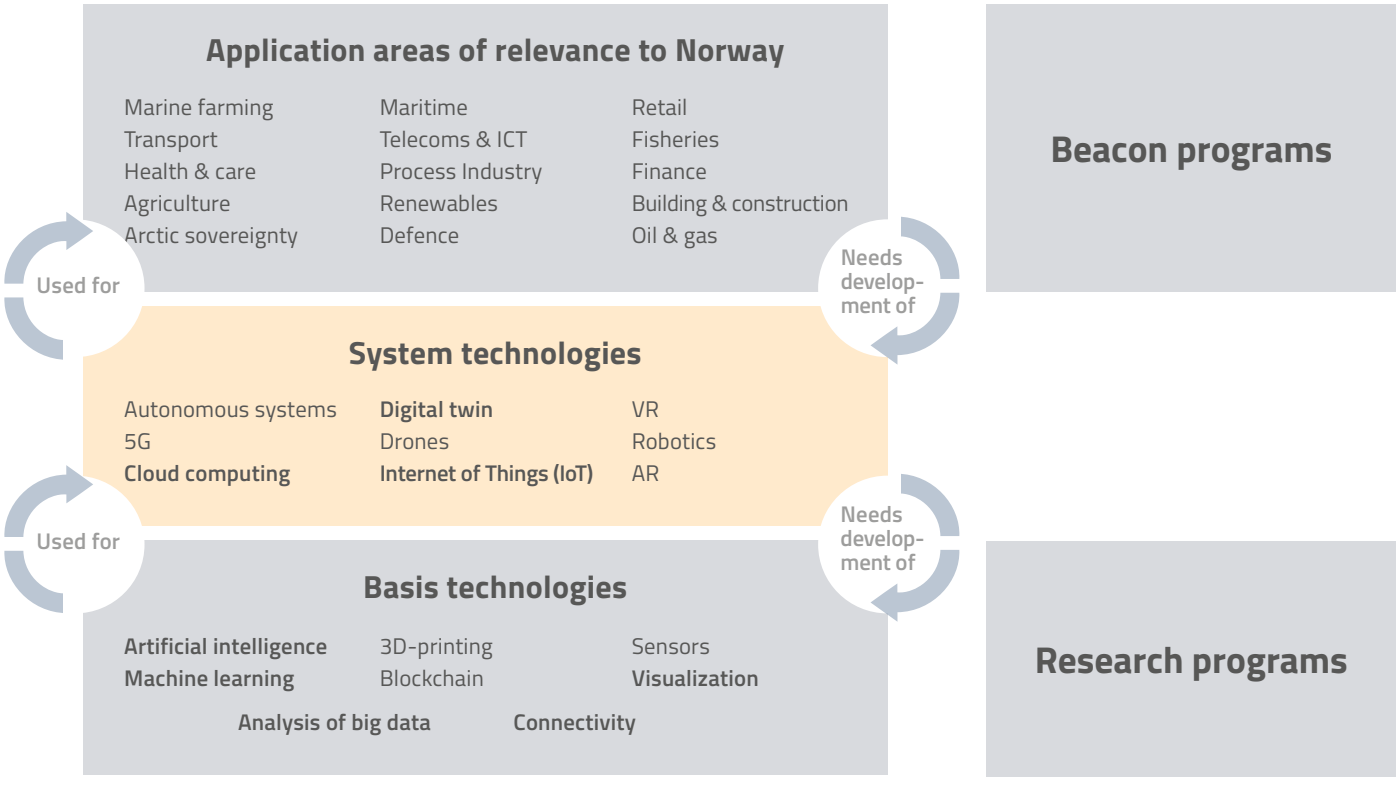
Realizing Digital21 and the AI Strategy

SIRIUS passed its halfway point in November 2019. We have four more years to achieve our research and innovation aims. We have identified eight areas where our research in scalable data access can fuel innovation. These are the SIRIUS beacons, which provide the applications that will drive innovation in the centre. The beacons are focused on business. Innovation in SIRIUS comes from linking these business challenges with the computer technologies that our research produces.

When we look at full time for the centre, at the end of 2023, we plan to have made industrial informatics scalable, standardised and sustainable. This supports the strategies for digitalisation, data and artificial intelligence that are being made at a European and national level. In Norway, this process was started, for the industrial sector, by the Digital 21 working party (https://digital21.no/#home). They chose a set of trend digital technologies, from the German *Industrie 4.0* initiative, and made a reference model of basic and system technologies. This approach was adopted, almost as a checklist, by many oil companies as the agenda for their digitalisation program. Fortunately, we seem to be moving beyond this now. SIRIUS is working on the practical challenges in making digitalisation work in and beyond the oil & gas industry. Instead of providing a grab-bag of hyped technologies, we are working on a set of research-based tools that can be used to build and maintain business-relevant systems.

No useful system is dependent on just a single system or basis technology. Digital21 proposed that research and development should concentrate on a selection of

Figure 1



technologies: namely artificial intelligence and machine learning, big data and data science, the Internet of Things, autonomous systems and robotics. However, useful systems need to draw upon all these technologies, and more. The technologies overlap. An autonomous system contains components from digital twins, virtual reality (VR), robotics, augmented reality (AR) and cloud computing. To design a research and development strategy on this selection of trend technologies is short sighted. It will enable Norway merely to be a good customer for foreign innovation. (See figure 1).

SIRIUS is building the computer science that lies behind trend technologies. We are working on systems that use automatic reasoning to support professional geologists and design engineers. Our researchers in complex systems are working on how the reliability of robots and autonomous systems can be proven. The scalable computing team is providing tools for effective use and management of cloud computing. The data science program is improving the usefulness of machine learning for the common situations where data sets are small. Our digital transformation workers look at how technologies are taken into use in industry.

The beacon programs ensure that our system technologies meet Norwegian industry needs and maintain competitive advantage. One-size, generic system technologies will not do this. For example, an Internet of Things built around training watches and household appliances has little to offer a high-risk, high-value industry like offshore petroleum. Our beacon programs lead to pilot projects that drive technology development from well-defined business problems. The system technology is sharpened so that it meets the specific business need.

Since the Digital21 report was published in August 2018 the Norwegian Government has worked further on a strategy for Artificial Intelligence. This presents a simple reference model (also used by the European Union) that tries to delimit what artificial intelligence is. They split AI into machine learning and machine reasoning and note that there is an overlap with robotics. (See figure 2).

Figure 2

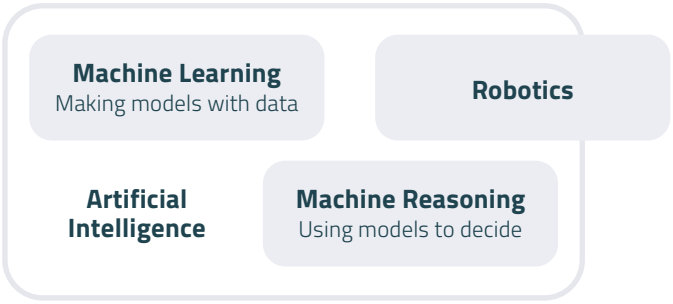
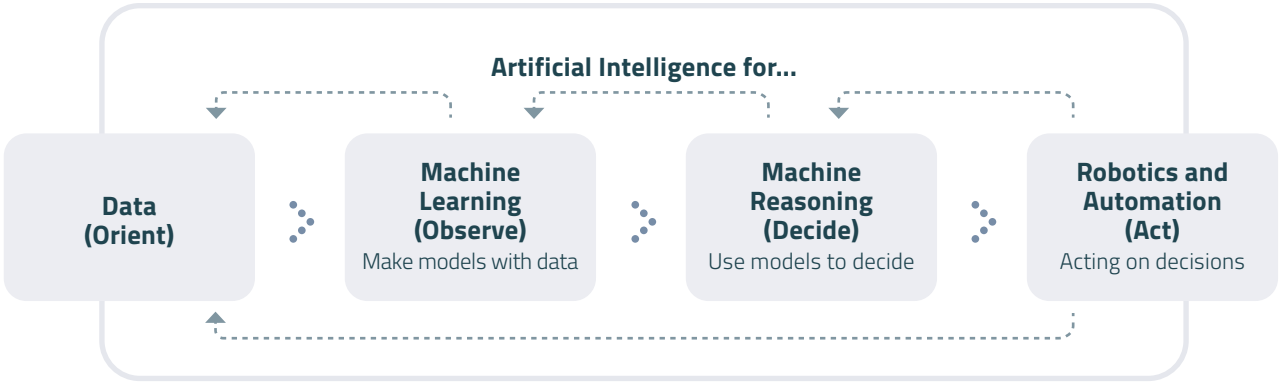


Figure 3



- Data semantics
- Data access
- Data cleaning
- Databases

- Domain-adapted ML
- ML with small industrial data sets
- ABS simulation
- Graph analysis
- Semantic modelling
- HPC for physical modelling

- Optimization
- Planning
- Prediction
- Semantic reasoning
- Computational logic
- Smart queries
- Explainability of AI

- Reliability of autonomous systems
- Design of cloud systems
- Effective user interfaces for using complex data sets.

At the end of 2023, we plan to have made industrial informatics scalable, standardised and sustainable – which supports the strategies for digitalisation, data and artificial intelligence that are being made at a European and national level.

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, the 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.

Photo: Ole Jørgen Bratland, Equinor.

How does SIRIUS' work fit into this vision? It is helpful to adjust the AI strategy's reference model to include the data that is needed to drive any AI application. We also show the feedback loops that are present in any AI system. This model fits well with the well-known OODA (orient-observe-decide-act) model for decision making.

We have noted in the figure (see figure 3) some of the areas where SIRIUS researchers are contributing to the AI vision.

Perhaps even more important, our beacon programs fill in the gap in "Artificial intelligence for...". SIRIUS and its partners are working on artificial intelligence for:

- **Reasoning around geological interpretations.**
- **Accessing and analysing subsurface data.**
- **Managing production from complex oil fields.**
- **Simulating and optimizing complex projects and logistics operations.**
- **Running equipment and production facilities in the best possible ways.**
- **Building high-quality capital works and facilities in a cost-effective way.**
- **Mobilizing health data to provide personalized medicine.**
- **Improving the quality of environmental monitoring, modelling and management.**

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 & 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, such as seismic data. Allow scalable access to real-time streams of sensor data. Make complex data accessible through end-user interfaces. Reduce the cost and risk of maintaining and changing systems.
Deliver a scalable, efficient and robust computational environment	Allow scalable processing and storage of big volumes of data. Process real-time streams of sensor data. Exploit affordable hardware platforms.
Reinforce mutual understanding and shared vision	Define and maintain shared vision. Establish new collaborations. Track evolving technologies and challenges. Achieve equal opportunity and gender equality within SIRIUS.
Establish SIRIUS as an internationally recognised centre of excellence	Attract additional funding (e.g. from EU programmes). Influence future research directions and funding policy. Influence society on big data, data access and digitalisation. Set up a PhD track that combines research and industry skills. Influence the international research community.

Highlights of 2019

1 Successful Mid-term Review

The centre was reviewed by an external panel of experts on 18th March 2019. SIRIUS received a letter from the Research Council on 9th September 2019 saying that the Board of the Research Council had received the report on the mid-term evaluation of the Centres for Research Based innovation. They decided based on this report that SIRIUS' contract will be extended to its full term of eight years. This allows us to proceed with our ambitions for demonstrated innovation in each of our beacon projects and research programs.

2 Spring General Assembly Hosted by Schlumberger

In May the Spring General Assembly was hosted by partner Schlumberger in Asker with 60 participants. This included both researchers and partners. The event was a mini-conference, with two full days of partner presentations, poster sessions and valuable discussions and networking. Plenary lectures were given by innovation leaders in Schlumberger, Equinor, IBM and SINTEF. These were followed by posters and demonstrations of work from the SIRIUS beacons. Day one concluded with demonstrations of partner's tools and integration frameworks: Kadme's WhereOil, Equinor's Sepes sandbox for data sharing, DNV GL's Veracity, IBM's Cognitive Discovery Pipeline and SAP's HANA integration platforms.

3 Reproducible simulation in ABS

Lars Tveito has been working to extend the ABS simulator with reproducible simulations. This makes it possible to record and replay a simulation run of a model. Reproducibility is especially useful for the models of parallel and distributed systems. It is also an important way of implementing model exploration, where we a model into previously unexplored behaviour. Model exploration gives us many of the benefits of full deductive verification of models, while being more applicable in industrial settings.

SIRIUS was well represented at the ISWC conference in Auckland, New Zealand. Some of our researchers won the Best In-Use Student paper. There was OTTR tutorial (Reasonable Ontology Templates), poster sessions and paper presentations.

4 SIRIUS becomes the first university member of OSDU

In mid-2019, the University of Oslo, through SIRIUS became a member of the Open Group's Open Subsurface Data Universe consortium. We were the first university in the world to join. This consortium is seeking to build a standard data platform for subsurface data in the oil & gas industry. This ambition is shared and supported by SIRIUS' subsurface beacon projects. Both Equinor and Schlumberger are active partners in OSDU. We will be aligning our research and innovation activities with the work done in OSDU.

5 Well-received OTTR templates tutorials at International Conferences

The OTTR templates, which simplify and structure the building of semantic models, have been well received in the Semantic Web Community. A full-day tutorial was held at the International Semantic Web Conference, held in Auckland. The tutors were Martin G. Skjæveland, Daniel Lupp, Leif Harald Karlsen, and Melinda Hodkiewicz from the University of Western Australia.

Martin G. Skjæveland and Daniel Lupp also held a tutorial on the OTTR templates for building ontologies at the European Semantic Web Conference held in Slovenia in June.

6 Workshop on Formal Methods and Artificial Intelligence in Logistics

SIRIUS helped organize the first international workshop on formal methods and artificial intelligence for logistics (FMAIL) in Bergen on December 2nd, 2019. The workshop discussed how current techniques from formal methods and artificial intelligence support the digitalisation of the logistics supply chain. Special focus was on distributed accounting and contract negotiations between stakeholders.

Melinda Hodkiewicz at the IOF 2019.

Mara Abel at the IOF 2019.

7 Industrial Ontologies Foundry (IOF)

SIRIUS and DNV GL jointly hosted the the Industrial Ontologies Foundry (IOF) and International Industrial Ontologies Workshop. The workshop was hosted at DNVGL's headquarters in Høvik. 130 participants attended from 17 countries, 34 companies and 16 research institutions. The IOF meeting was hosted by SIRIUS at the University of Oslo. At this meeting, the OTTR project was presented in a tutorial. Mara Abel, our collaborator at UFRGS in Brazil, was active in the IOF discussions. The IOF is a group working to co-create a set of open reference ontologies to support the manufacturing and engineering industry needs and advance data interoperability.

Ian Horrocks at the IOF 2019.

Melinda Hodkiewicz at the IOF 2019.

8 Collaboration with University of Western Australia

SIRIUS workers have started to work together with researchers from the University of Western Australia in Perth. The OTTR templates developed in SIRIUS have been applied by a group, led by Professor Melinda Hodkiewicz, to linking maintenance records to fault modelling analyses. There are many common interests between the Australian and Norwegian resource industries, so we see this as a fertile area for future work.

9 International Decision Logic Conference and Industry panel

The 32nd International Workshop on Description Logics was hosted by University of Oslo in June this year. The DL workshop is the major annual event of the description logic research community. It is the forum at which those interested in description logics, both from academia and industry, meet to discuss ideas, share information and compare experiences. An industrial panel session was moderated by Martin G. Skjæveland and featured presentations by industrial practitioners from DNV GL, Aibel, CapGemini and Festo.

10 Ontology Engineering roadmap workshop

A workshop was held on 20th June 2019, co-located with the international description logic conference, to discuss and agree on a roadmap for ontology engineering work in SIRIUS.

11 Daniel Lupp successfully defends his PhD thesis

Daniel Lupp successfully defended his PhD thesis on 4th May 2019. His thesis was entitled "A Higher-Level View of Ontological Modeling: Rule-Based Approaches for Data Transformation, Modeling, and Maintenance". Daniel is now employed as a post-doctoral researcher in the Ontology Engineering group.

12 Aibel joins SIRIUS.

Aibel joined SIRIUS in March 2019.

13 Aker Solutions joins SIRIUS

Aker Solutions joined SIRIUS in May 2019.

14 Presentation at Schlumberger Global Forum

A talk on "Geological multiscenario reasoning" was presented by Irina Pene and Ingrid Chieh Yu at Schlumberger's SIS Global Forum in Monaco. This presented results and a demonstration of the work done in the Geological assistant beacon project.

Crystal Chang Din, Irina Pene and Ingrid Chieh Yu at Schlumberger's SIS Global Forum in Monaco, 2019.

15

Crystal Chang Din with the Best Scientific Paper award at the annual NIK Conference in Narvik, November 2019.

15 Best Paper at Norwegian national ICT Conference

SIRIUS researchers won the Best Scientific Paper award at the annual NIK Conference in Narvik in November. The paper that won the award was called "Geological Multi-Scenario Reasoning" and was authored by Crystal Chang Din, Leif Harald Karlsen, Irina Pene, Oliver Stahl, Ingrid Chieh Yu and Thomas Østerlie.

16 November Conference in Rio de Janeiro

Once again, SIRIUS workers played a prominent role in the annual November Conference on Norwegian-Brazilian Energy Research. This year's conference was hosted by Petrobras, with plenary speakers from Equinor, IBM and Aker Solutions. David Cameron was chair of the program committee and co-organized the session on digitalisation and automation. Thomas Østerlie held a talk in the session on Energy Transitions.

17 November General Assembly Hosted by Computas

The November General Assembly, as usual, was a business meeting where we reviewed the annual work plan for the centre. This meeting was hosted by Computas at their new offices in the centre of Oslo.

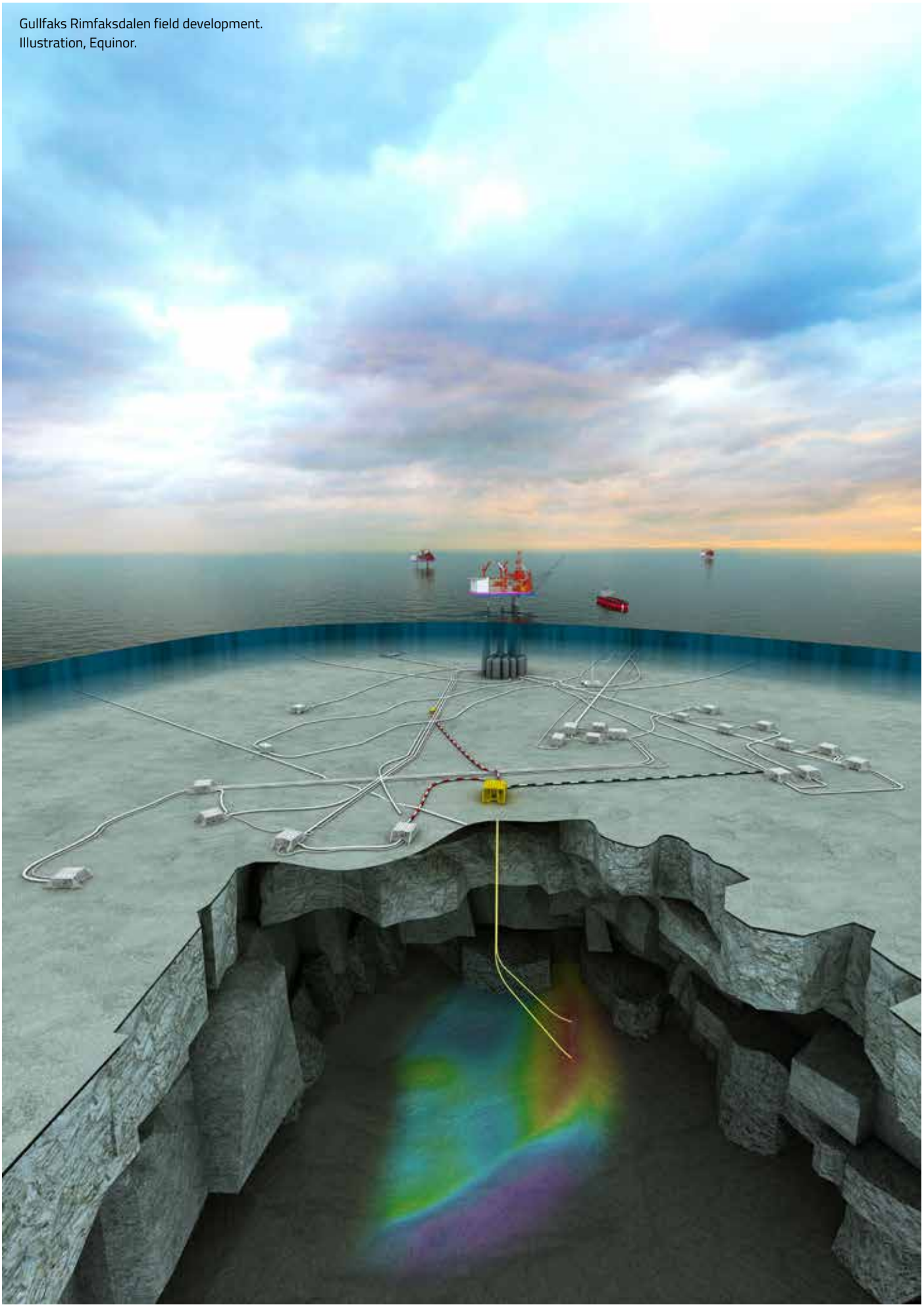
18 Robert Bosch GmbH joins SIRIUS

Robert Bosch GmbH joined SIRIUS in December 2019, through their Global Centre for Artificial Intelligence in Stuttgart, Germany. Bosch is well-known as a leading supplier of components and equipment for the engineering and automotive industries. Bosch will be working with SIRIUS on digital twins, industrial data science and semantic data integration along the supply chain.

19 Mina Haghshenas wins Best Paper Award at MCIS Conference

Mina Haghshenas won the "Best qualitative research paper" award at the Mediterranean Conference on Information Systems in Naples in September 2019. The paper draws upon her study of work processes in phase one of the READI project. READI is a collaborative project, managed by DNV GL, among operators, contractors, vendors, and service companies within the Engineering, Procurement, and Construction (EPC) sector. Phase one focused on developing a method for expressing requirements on a digital format that can be read and processed by computers.

Mina Haghshenas won the Best qualitative paper award at the Mediterranean Conference on Information Systems in Naples, September 2019.

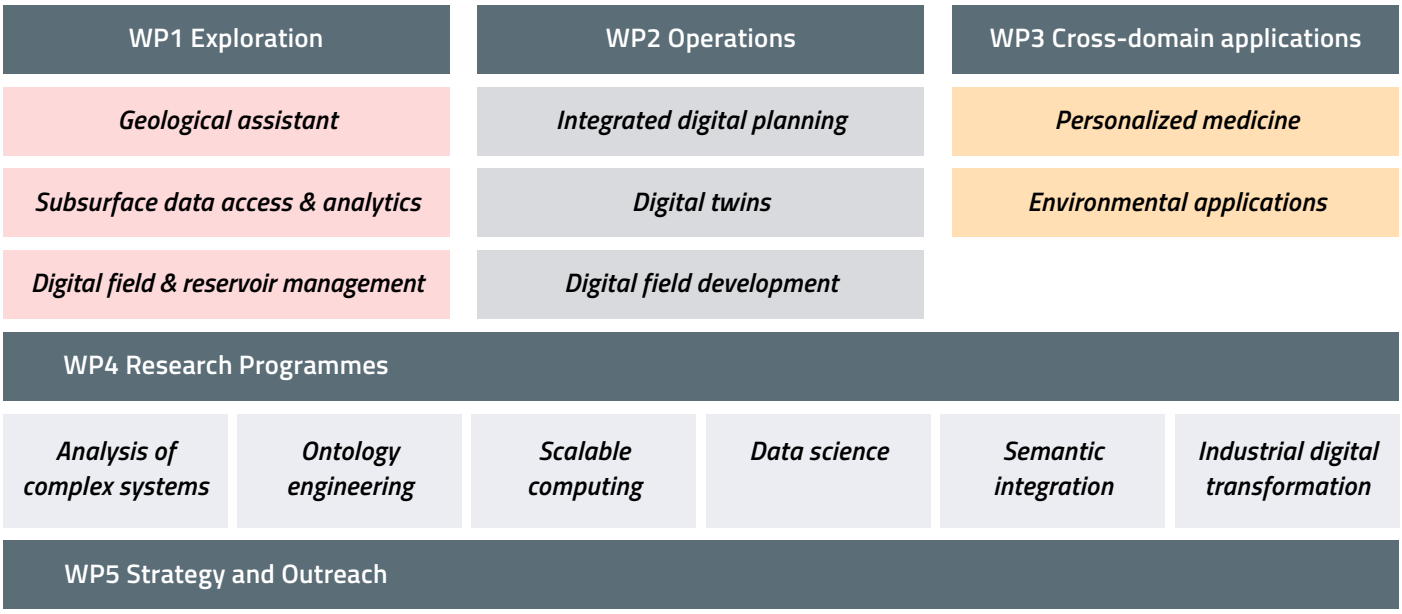


Gullfaks Rimfaksdalen field development.
Illustration, Equinor.

How SIRIUS is Organized

SIRIUS' projects are organised into either business-related work packages - beacon projects - or technically oriented research programs. Each researcher works in a research program and may be involved in one or more beacon projects. The research programs provide a foundation of methods and tools that can be prototyped and piloted in the beacon projects. Such a matrix organisation enables us to obtain the right balance of fundamental research and focus in innovation and industrial applications.

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

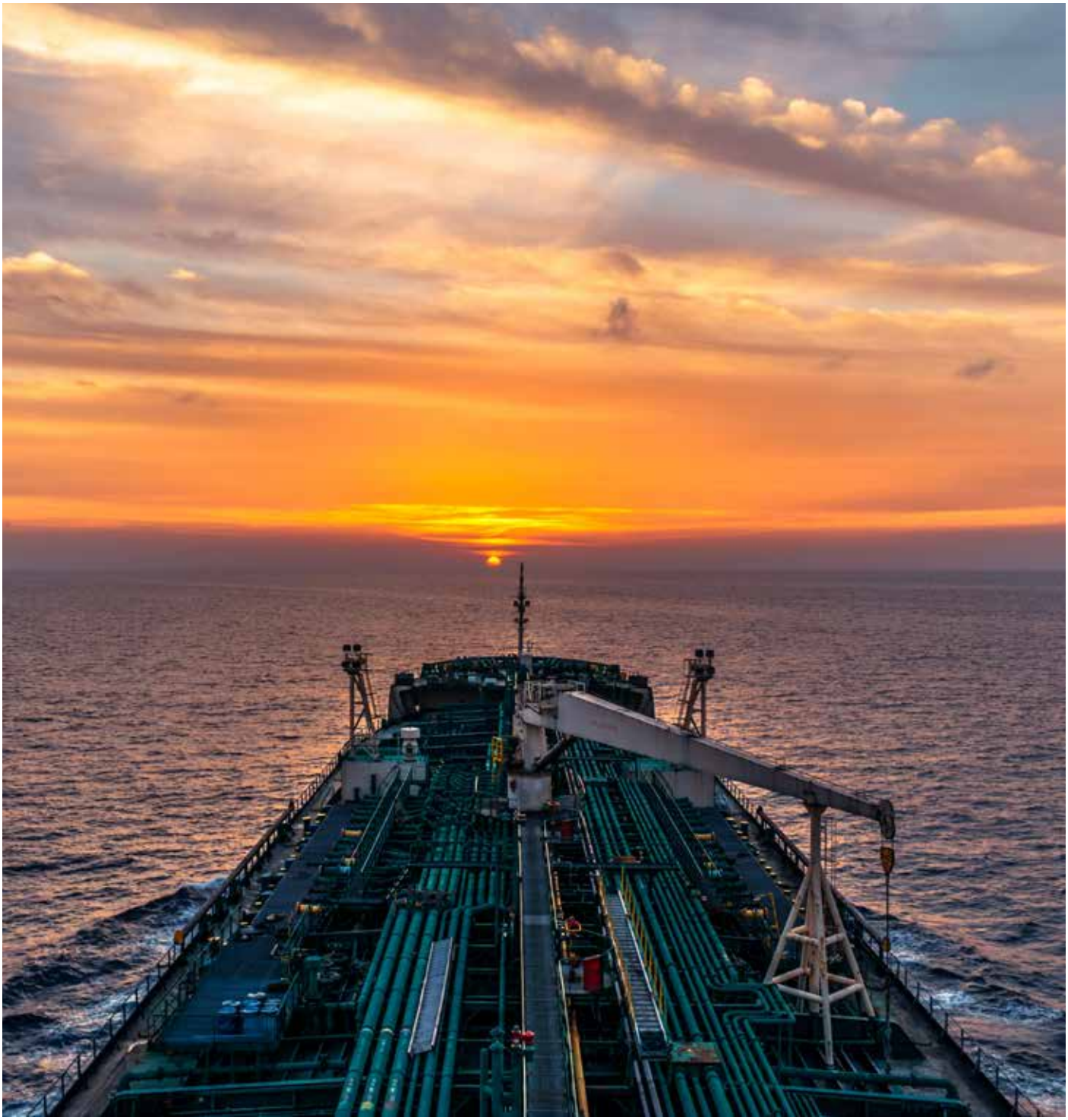


SIRIUS has five work packages:

- **Exploration:** beacon projects in exploration and subsurface applications.
- **Operations:** beacon projects in the design, construction, operation, maintenance and decommissioning of complex industrial facilities.
- **Cross-domain Applications:** applications in areas outside and beyond the natural resources industries, notably health and environmental applications.
- **Research Programs:** SIRIUS' computer science results that can be applied in projects in several of the business work packages. 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 fifteen industrial partners and five research 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.
- Simula Research Centre participates with work on high-performance computing and scalable computing.
- Finally, SINTEF joined SIRIUS in 2019.

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 Aker Solutions, Aibel and Robert Bosch GmbH joining the centre in 2019. Negotiations are continuing with several leading operating, service and software companies.

No single partner is interested in all the beacon projects, although a large operator such as 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						■								

At Melkøya. Photo: Equinor

New partners in 2019

Aibel is a leading service company within the oil, gas and offshore wind industries. They offer their customers optimal and innovative solutions within engineering, construction, modifications and maintenance throughout a project's entire life cycle. Aibel's 4000 skilled employees are located close to customers at the company's offices in Norway and South East Asia. In addition, they operate two modern yards in Haugesund and in Thailand, with complete prefabrication and construction capabilities.

Aibel's ambition is that the SIRIUS Centre will be able to contribute with a long-term perspective that gives Aibel a stronger basis for realization of improvements. The potential is mainly how to better handle and get more value from the huge amount of information processed in projects. Aibel's industry processes large amounts of information, and therefore a lot of research is carried out on how to best exploit the potential of this information. How to better manage, analyse and exploit information in the future is a clear goal for Aibel's collaboration with SIRIUS.

Aibel will by this agreement get access to relevant research and now also be able to direct research on topics that are important for our company. In addition, Aibel also participates in the READI joint industry project.

Aibel actively participates in the Digital Twin and Ontology Engineering research programs. Potential effects from those technologies have been demonstrated in the Johan Sverdrup P2 project.

Robert Bosch GmbH (Bosch) is a world-class manufacturing company working in four business sectors of mobility solutions, industrial technology, consumer goods, and energy and building technology. It operates hundreds of manufacturing plants with thousands of assembly lines providing the world with products that are "Invented for life". As a leading IoT company, Bosch offers innovative solutions for connected manufacturing and Industry 4.0. Moreover, cutting edge research at the Bosch Center for AI aims at achieving a leading position for Bosch in AI and transforming Bosch towards an AI-driven IoT company. Bosch is delighted to join the SIRIUS centre of excellence that brings together top researchers and businesses. The SIRIUS membership and SIRIUS-driven EU and Norwegian level research initiatives and projects is a great opportunity for Bosch to jointly drive innovations and value in the areas of industrial automation, AI-driven IoT, Digital Twins, semantic technologies, and industrial analytics.

Aker Solutions joined SIRIUS in 2019. This partnership formalized a long-standing relationship between SIRIUS researchers and the software development groups in Aker Solutions. Personnel from Aker Solutions have participated in the centre's Industrial Ontology Seminars. Aker Solutions and its new iX3 software subsidiary have built and maintain powerful IT systems that support engineering, procurement and construction processes. These tools already use cutting-edge semantic technologies. Aker Solutions is working with SIRIUS and in SIRIUS-linked joint industry projects to improve the effectiveness of existing tools, demonstrate their use in different contexts.

There is a need to build a set of industry commons that allow engineering firms and operators to communicate better and spend less time and money on documents and project-specific exchanges of data and specifications.

SINTEF joined the centre in mid-2019 through its SINTEF Digital business unit. SINTEF is one of Europe's largest independent research organisations and is a leading actor in European ICT projects. SINTEF Digital has its main office very close to SIRIUS and there are many SINTEF Digital personnel who have teaching or research roles at the University of Oslo. We have also been working together in the Gemini Centre for Big Data and in the European BDVA (Big Data Value Association). SIRIUS and SINTEF Digital share a digitalisation agenda, where we wish to apply computer science in industrial contexts. This means that we also have contact with the other units in SINTEF: Ocean, Industry, Manufacturing and Energy Research. SINTEF's participation in the centre is self-funded and focuses on working with us to bring our methods to higher Technology Readiness Levels and to participate together in the Horizon Europe programme.

IBM: The Advantages of Multiple Innovation Channels

In the SIRIUS research collaboration, IBM has engaged the Brazil Research Centres in Rio de Janeiro and São Paulo as IBM Research Brazil focuses on research related to oil & gas. IBM has been working with SIRIUS on various opportunities for collaboration in both exploration and production since 2016. Innovative technologies such as cloud-based platforms and artificial intelligence (AI) are already transforming industries and we value being able to collaborate with various partners within the petroleum sector to develop new and improved solutions for the industry.

A recent global IBM study based on interviews with more than 350 C-level executives states that the advantage of using multiple innovation channels is evident - and the collaboration within the SIRIUS-group is a great example. By exchanging and testing ideas together we are gaining valuable experience and developing assets for future exploration and production.

IBM would like to point to the following highlights of their activities with SIRIUS in 2019:

The Geological Image Annotation project

This was established as an innovation project with participation from **Equinor**, **SIRIUS** and **IBM Research**, the first phase of this project ran through all of 2019. The project developed a prototype capable of annotating selected image classes, including Geological Maps and Lithostratigraphic Charts, to be used for training image classification machine-learning models. Enhanced automated understanding of such images allows oil & gas companies to evaluate subsurface data more efficiently, for example when identifying analogues during the acquisition of new oil fields.

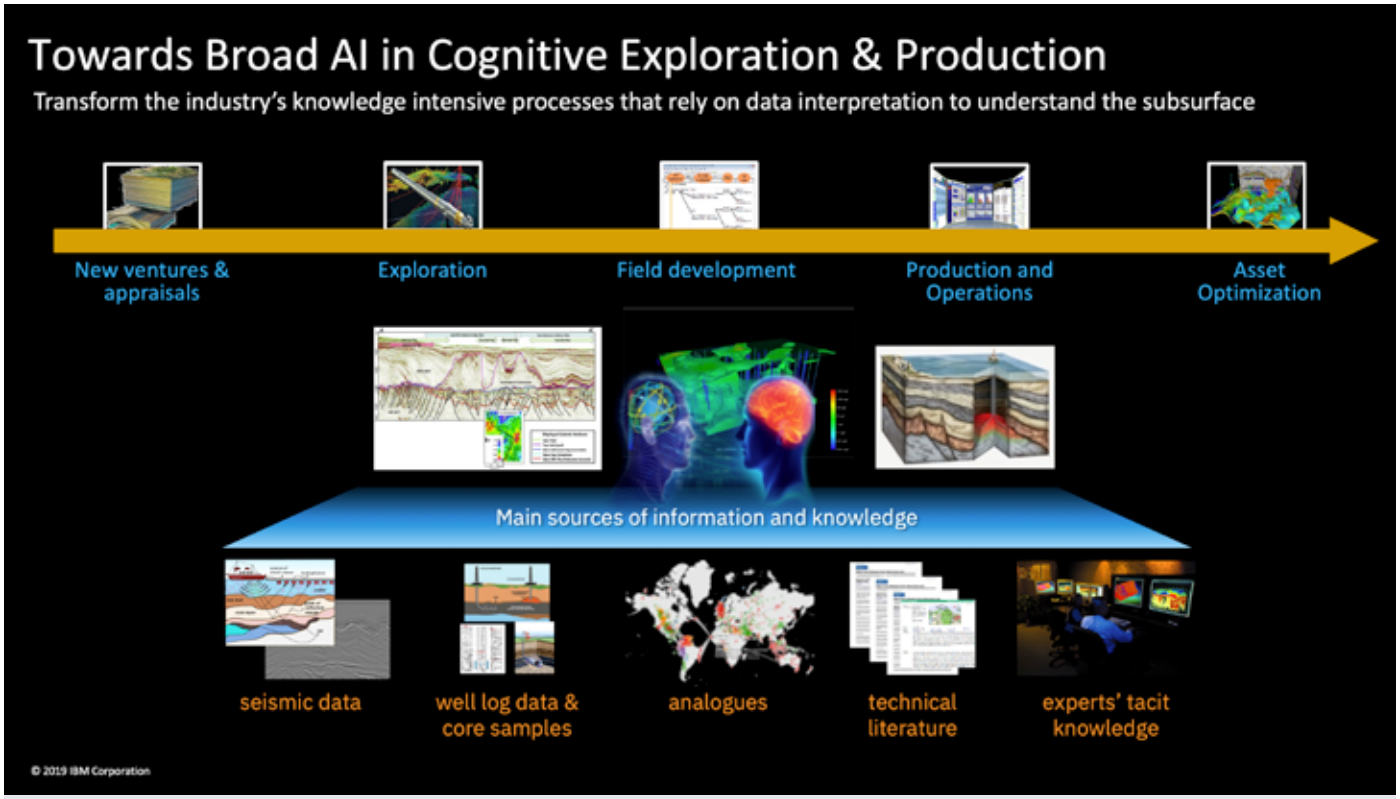
IBM Research values the collaboration in this project, and the knowledge and experience that we are accumulating through it. We have already:

- **increased our understanding of the oil & gas exploration domain.**
- **established further experience with geoscientific ontologies.**
- **participated in the design of the annotation tool.**

Allowing the annotation tool to be used by a larger number of people such as geoscience students, also known as crowdsourcing, rather than depend solely on professionals from the oil companies may contribute to accelerate the annotation task. For the next phase for the project, we propose further acceleration by integrating the tool with our Watson Discovery tool.

Mentoring

IBM has, since 2017, supported the **SIRIUS mentoring program** and participated with two mentors for 2019/2020. The experience is highly valuable for both mentors and mentees, and the benefits for the mentors includes opportunities to learn from your mentee's background and history to enhance professional and personal development, in addition to get to know and influence the SIRIUS group itself. For example, one of our experienced mentors was so impressed by the SIRIUS-team and the mentees he met, that he worked to establish a SIRIUS collaboration activity between his business unit and the SIRIUS centre itself. Furthermore, IBM is actively introducing our new employees into the work that SIRIUS is doing, and in 2019 we welcomed one our new graduates to the Geological Image Annotation project.



From Ulisses Mello's presentation at the November Conference in Rio de Janeiro November 2019

The Brazilian Norwegian Energy Research conference in Rio de Janeiro in November 2019

The November Conference in Rio de Janeiro has expanded to include all energy research, with special focus on the linkages between fossil energy and renewable energy. IBM sponsored the conference and Ulisses Mello, director of IBM Research Brazil and IBM Global Research Leader for the Petroleum Industry, gave a plenary speech. His presentation of *the Transformation in the Upstream and the role of AI*, was very well received. A path was outlined towards the future with broad AI, characterised by the ability to learn more broadly across multiple narrow AI domains and tasks.

2020 initiatives - The Open Subsurface Data Universe

SIRIUS has decided to join the Open Subsurface Data Universe (OSDU), an industry forum aimed at creating a cloud native system of record for well data and seismic data. IBM is already a member of OSDU, and we consider it as an enabler for subsurface AI applications in the industry. Together with SIRIUS partners, IBM recently participated in an OSDU workshop, and we are now investigating new initiatives for 2020 to explore and develop synergies between our approach to OSDU and SIRIUS with partners.

Rio De Janeiro. Photo: Davi Costa

Melkøya. Photo: Helge Hansen, Statoil

SIRIUS' Research Programs

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 & 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

Ian Horrocks

Ian Horrocks, FRS, is 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, optimization 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.

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 digitalisation efforts in organizations. This involves identifying the organizational and technical conditions for realizing the potential of digitalisation in practice. He has been project leader for a portfolio of research projects into digitalisation of oil & gas in close collaboration with industry. He is part of the management group on BRU21, a large-scale, fully industry-funded national initiative on digitalisation and automation in oil & gas. He has a robust academic publication track record (citations 6305/210, h-index 27/8, with Google Scholar/ISI).

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 with the oil & 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; and 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 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

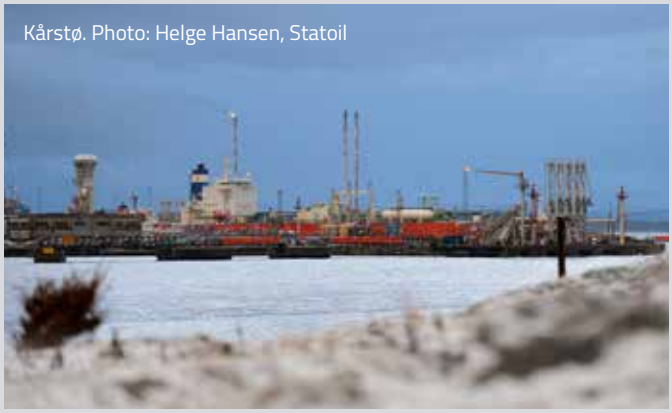
Industry-relevant data comes in many forms, from structured sources (e.g. databases) and unstructured sources (e.g. natural language documents intended to be read by humans). Having access to all this data in only as useful as the methods one has for evaluating and using this data to make decisions. 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 focuses 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.



Kårstø. Photo: Helge Hansen, Statoil

Analysis of Complex Systems

The oil & gas industry is full of complex systems. Designing a complex system is challenging: many possible settings and parameters can be tuned. Poor choices can result in system failure, high costs and displeased customers. These systems are technical, such as computer-based systems, or organizational, such as maintenance programs. Simulation and formal analysis can improve and optimize the performance of these systems. This makes operations safer, more efficient and more reliable.

Complex systems consist of many components with concurrent tasks and tangled interdependencies and interactions. The overall research conducted in this research program addresses ways to dynamically describe, predict and prescribe convoluted systems so that we can design them properly using behavioural models. A model is mainly a representation of some aspects of interest from a real-world system. Thus, models contain fewer details than such systems. By decreasing the amount of detail, we may be able to more easily describe and understand those aspects of interest, e.g., behaviours and interactions, and do predictive and prescriptive analysis. Model-based analysis using formal methods is particularly useful because formal methods use logic-based techniques, which are very effective for abstractly modelling the behaviour and interaction of such systems. In addition, techniques based on formal modelling can allow automation or semi-automation of the analysis and scalability of the targeted models.

The goal of the Analysis of Complex Systems Research program is to develop tools and techniques within formal methods. We research on formal languages for system specifications and techniques to describe, predict and prescribe the behaviours and interactions of system executions based on the analysis of models. The research contributes to various methods and tools. In particular, we are building and maintaining ABS, a modelling and analysis framework for distributed systems.

The ABS Modelling Language

(http://www.abs-models.org)

Silvia Lizeth Tapia Tarifa

Silvia Lizeth Tapia Tarifa is the leader of the Analysis of Complex Systems research program. Her main research area is formal methods for parallel and distributed systems. She has a Young Research Talent grant from the Research Council of Norway, the only grant in computer science for this call in 2017. In this project she is applying static analysis techniques to approximate parallel data access in architectures with shared memory. She has worked in several EU projects including FP7 ENVISAGE, FP7 FET UpScale and FP7 FET HATS. She had a research stay in Imperial College London in 2016 and she was a fellow at the United Nation University, International Institute for Software Technology in Macao in 2007. She has been PC co-chair of iFM'19, PC co-chair of PhD-iFM 2018 and 2017, PC member of FASE'21, SEFM'20, ICE'20, DV-SoA20, FASE'19, ICE'19, F-IDE'19, DISE'19, DEVOPS'18 and DISE'18. She has more than 25 co-authored peer-reviewed papers and Journals. Her scientific interests are model-based analysis techniques using formal methods, programming language semantics, modelling and predictive analysis of complex systems.

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.

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

Our aim is to predict the behaviour of complex systems using the analysis of models.

Spotlight on ABS: Reproducing and Replaying Simulations of 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. Using the ABS modelling language and analysis framework, 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.

Simulation and formal analysis can improve and optimize the performance of complex systems. This makes operations safer, more efficient and more reliable.

Gianluca Turin

Project: ADAPt: Exploiting Abstract Data Access Patterns for Parallel Processing. Which main goal is to improve data processing, by systematically extracting data-access patterns from applications and by matching and customising task scheduling and data allocation using such patterns.

What I am working on:
Applying techniques well known in the field of formal methods (type checking, model checking, static analysis in general) to the empiric world of parallel computing. Trying to find which kind analysis can actually lead to a benefit in terms of performance, and for in what sort of parallel problems/algorithms/architectures they can be exploited.

How this is connected to SIRIUS?
Well, Centre for Scalable Data Access already says a lot. Furthermore many companies involved may have a direct or indirect interest: Numascale which is in the business of parallel architectures, Equinor as end user of parallel developing techniques (for geological simulations, mere exploitation of big data), Simula as research company mostly focused on scientific computations. And probably even more that outgrow my knowledge of the company landscape behind SIRIUS.

In 2019, we augmented the ABS analysis framework with reproducible simulations. This new technical feature makes it possible to record and replay a particular simulation run of a model to analyse and share its behaviour. Reproducibility is especially useful for the models of parallel and distributed systems that the ABS language specializes in. It is also an important foundation for implementing model exploration, a technique that guides a model into previously-unexplored behaviour. Model exploration will give us many of the same benefits as full deductive verification of models, while being significantly more applicable in an industrial setting such as SIRIUS.

From Project-driven to Community-driven Research

We are continuously moving ABS from a project-driven to a community-driven technology. To this end, there has been a series of international workshops for ABS. The first international ABS workshop was held in Oslo, Norway (May 2017), the second international ABS workshop was held in Darmstadt, Germany (May 2018), the third international ABS workshop was held in Amsterdam, The Netherlands (May 2019), and the forth international ABS workshop will be held in Torino, Italy (June 2020). 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.

Chi Mai Nguyen

Chi Mai Nguyen is a postdoctoral researcher at the SIRUS centre, Department of Informatics, University of Oslo. She obtained her Masters degree in Theoretical Computer Science at Aalto University, in 2012. In April 2017, she received her PhD degree at University of Trento with a thesis entitled "Efficient Modelling and Reasoning with Constrained Goal Models". Her work mainly focuses on formal verification, model checking, and applying automated reasoning techniques (mostly SAT and SMT) to goal models.

Lars Tveito

Project: Rapid Verification of Concurrent and Distributed Specifications

What are you working on?

I am working on strengthening the testing- and debugging facilities of the ABS modelling language, by offering reproducible simulations and systematic model exploration. The techniques and tools developed are very general and may be applied to many other computational models and languages.

How is this connected to SIRIUS?

SIRIUS projects that leverage the ABS modelling language may enjoy the strengthened testing and debugging facilities provided by the tools I develop.

ADAPT: Exploiting Abstract Data-Access Patterns for Better Data Locality in Parallel Processing

(https://www.mn.uio.no/ifi/english/research/projects/adapt/)

The ADAPT project started in 2018, with Silvia 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 to

coordinate systematically 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 combines formal models of parallel systems with basic research in programming language theory to (1) capture abstractly the interaction of workflows with dynamically created tasks and memory locations on parallel computers, and (2) combine formal analysis with model-based simulations to explore how to systemically calibrate schedulers and allocators for specific applications. The result of this project will complement SIRIUS' current toolbox on model-based prediction.

Violet Ka I Pun

Violet Ka I Pun is currently associate professor at Western Norway University of Applied Sciences and at the University of Oslo, associated to SIRIUS centre. Before that, she was a postdoctoral researcher in the research group Reliable systems at the University of Oslo. Her research to date has investigated a wide range of challenges in the area of formal modelling and analysis. In particular, her expertise focuses on verifying behaviour of concurrent programs based on formal methods, using especially static analyses to guarantee safety properties. Application domains include worst-case execution time execution for services running on distributed systems, multicore architectures, and automation and tool-support for planning in oil industry and healthcare domain. Her early research interests include analysing traffic intensive web-services based on workflow patterns, and prototyping intelligent question and answering system based on computer-human interaction and speech recognition. She has been involved in COST Actions as well as several national and EU-level projects, including Intromat, COEMS, UpScale, and Envisage.

Operational Centre. Photo: Helge Hansen, Statoil

Ontology Engineering

The digital transformation of the industry depends on rich information models in order to support automation of specialized and knowledge intensive tasks. These models must be intelligible to both computers and humans and 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.

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. This makes maintaining the ontology a real issue.

Martin G. Skjæveland

Martin G. Skjæveland is a researcher at the University of Oslo, leader of the SIRIUS research program on Ontology Engineering and project leader of the Reasonable Ontology Templates project. Martin is also associate professor in data science at the University of Stavanger. Martin's research interests are ontology engineering and ontology-based systems 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.

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 the 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 become experts in logic and semantic technologies, whereas ontology experts and information modellers have the easier task of working with abstractions over the given domain. 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 and exploiting the knowledge captured in the ontology with existing software platforms.
- equipping ontology experts with **powerful tools** to oversee the development of the ontology

Projects

Reasonable Ontology Templates (OTTR)

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

easier to comprehend and maintain and ensures that any changes made to a pattern are automatically propagated 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. The project is working with Aibel, DNV GL and Aker Solutions to further develop the functionality of OTTR and its open-source reference implementation used for building and maintaining ontologies.

Ontology modularity

The project aims to develop efficient procedures for maintaining large and complex ontologies using ontology modularity techniques. These techniques extract a significant smaller module, a subset of an ontology, which users are interested in and still preserves the same semantics as the whole ontology. Therefore, users can focus on smaller modules and safely replace the full ontology with such smaller modules for many use cases.

The project is working with Aibel to apply these techniques to their Material Master Data (MMD) ontology. The MMD ontology is an OWL 2 compliant ontology with around 100k classes and 1.8m axioms. It is a representation of multiple industry standards and requirements and is used by Aibel to help automate the selection of valid design artefacts and products in their engineering projects.

Levels of granularity

The aim of this project 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. To aid domain users in the ontology design process, this project is developing a methodology aimed at reducing the gap between intuitive, 'box-and-arrow' diagrammatic modelling and their logical implementation in industrial ontologies. The approach will provide a formalism capable of capturing ontologies at various levels of abstraction, ranging from intuitive patterns between entities to their implementation in OWL. Thus, users will be able to interact with the model at a high conceptual level which more closely resembles their understanding of the domain, as opposed to the underlying logic.

To support users in querying large-scale data sets, this project is continuing development on the OptiqueVQS, a

Martin Giese

Martin Giese is a professor of informatics at the University of Oslo. He has a Ph.D. 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 thirteen years, he has been working in projects to bring ontologies and other knowledge representation techniques to applications in the 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

Leif Harald Karlsen's research in SIRIUS is connected to the following two SIRIUS projects:

OTTR Templates:
A project that introduces a novel framework for scaling up ontology construction to real-world domains. In this project I work on the implementation of Lutra, the software driving the framework. Ontologies are central to solving many of the problems SIRIUS is addressing, and this project helps in constructing scalable and robust ontologies.

The GeoAssistant project, which works on combining knowledge representation and formal methods in order to capture static and dynamic knowledge about geology and to do multi-scenario reasoning on geological processes. In this project Leif Harald work on spatial representation and reasoning and on combining ontologies with rewriting logic. The GeoAssistant project is a beacon project for SIRIUS.

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. 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 (2019) on ranking of user interface suggestions based on past queries.

Events in 2019

1. Full-day tutorial on OTTR templates at ISWC 2019 located in Auckland, New Zealand. Tutors were Martin G. Skjæveland, Daniel Lupp, Leif Harald Karlsen, and Melinda Hodkiewicz from the University of Western Australia.
2. The Industrial Ontologies Foundry (IOF) organized via SIRIUS and DNV GL and International Industrial Ontologies Workshop hosted at DNV GL's headquarters in Høvik. 130 participants attended from 17 countries, 34 companies and 16 research institutions. The OTTR project was presented in a tutorial by Martin G. Skjæveland, Daniel Lupp and Leif Harald Karlsen.
3. Half-day tutorial on OTTR templates at ESWC 2019. Tutors were Martin G. Skjæveland and Daniel Lupp.
4. Industry panel at the international Description Logic workshop, moderated by Martin G. Skjæveland, featuring presentations by DNVGL, Aibel, CapGemini and Festo.
5. Ontology Engineering roadmap workshop co-located with the international Description Logic Workshop.
6. Daniel Lupp successfully defended his PhD thesis: "A Higher-Level View of Ontological Modeling: Rule-Based Approaches for Data Transformation, Modeling, and Maintenance".

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.

Daniel Lupp

What are you working on?
A formal methodology for constructing, maintaining, and interacting with information models at varying levels of abstractions. Additionally, extending ontology-based information systems to allow for reasoning over requirements.

How is this connected to SIRIUS?
The requirements work is associated with the READI project, whereas the modelling methodology is linked to multiple projects, including READI and planning. Neither of these research directions are domain-specific, so they are applicable to a number of activities within SIRIUS.

Jieying Chen

I am a postdoctoral researcher at SIRIUS, University of Oslo (Norway). My research interests are ontology modularity, description logics and KR. Before I was a research associate at the University of Manchester (UK) and worked on the EPSRC IAA Project "Comparison and Abstraction of SNOMED CT Ontologies". I obtained my PhD degree at Université Paris-Saclay/Paris-Sud (France) on "Knowledge Extraction from Description Logic Terminologies". During my PhD, I was a visiting doctoral student at Vrije Universiteit Amsterdam (Netherlands) and TU Dresden (Germany). When I did my master in China, I was an exchange student at TU Dresden (Germany) funded by Erasmus Mundus and DAAD.

Turnaround at Melkøya. Photo: Harald Pettersen, Statoil

Vidar Klungre

Project: Query Extension Suggestions in OptiqueVQS.

What are you working on?
I work with OptiqueVQS, which is a tool that allows users to construct SPARQL queries and run them over RDF data.

In this setting, I try to predict future queries based on query history.

This, combined with the underlying RDF dataset, can be used to present better query extensions to the users, which makes them more efficient.

How is this connected to SIRIUS?
The project is funded by SIRIUS because it solves real industry cases using semantic technologies.

OptiqueVQS is built and developed to fit an industry setting, and it helps domain experts to formulate queries over large datasets without having to know a technical query language.

Ratan Bahadur Thapa

Project: Integrity Constraints in Ontology Enhanced Knowledge Representation

What are you working on?
Semantic preserving constraints rewriting and reasoning in ontology-based data access and integration. The notable problem in this setting is the satisfaction and implication of integrity constraints between closed-world and the open-world interpretation given the semantics of data mappings and the domain overview of the ontology.

How is this connected to SIRIUS?
The PeTWIN NFR project affiliated with the SIRIUS Centre aims to provide the infrastructure and best practices for next-generation Digital Twins for field digitalisation. This endeavour requires precise domain models for semantic integration and interoperability of data from diverse facilities, and the integrity constraints will be a critical piece of information both to standardize and optimize those domain models.

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.

Digitalisation of oil & 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 & gas industry traditionally resides in database management systems (DBMS). Recent developments, such as OSDU, are moving towards less structured data sources. However, the complexity of end users' data needs will not diminish, so the lack of structure must be compensated. We want to allow the user to ask for the data using their description of reality. This is the vision behind semantic integration. In SIRIUS we are working on different ways of providing semantic integration, using *query rewriting and materialization*.

DEPARTMENT OF COMPUTER SCIENCE

UNIVERSITY OF OXFORD

The Information Systems group at the University of Oxford's Computer Science Department participates closely with the 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.

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 edgewise 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 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.

Materialization: An alternative to query rewriting is materialization of a query. This involves making a copy, materializing, 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

Alessandro Ronca

What are you working on?
I am working on processing streaming data via rule-based logical formalisms. In this context, the way streams are to be processed is specified via if-then rules, and then the stream processor takes care of computing the processed stream according to the specification. Specifically, I have been investigating fundamental problems underlying the realisation of such a stream processor for logical rules. For instance, a core challenge is to determine whether the input data received so far suffice to correctly compute any prefix of the output stream.

How it is connected to SIRIUS?
In wellsites there is the need to continuously monitor and analyze sensor data in order to detect equipment malfunction and predict maintenance needs. Solutions that address this task would benefit from using logical rules. They allow an engineer to abstract from lower level details, speeding up development and maintenance, and increasing the confidence in the correctness of the solution.

generated, giving us more insight into the existing data. Consider a simple example: a compressor is labelled in one data-base 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.

> **We want to** allow the user to ask for the data using their description of reality. This is the vision behind semantic integration. In SIRIUS we are working on different ways of providing semantic integration, using query rewriting and materialization.

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 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 2019 we have made significant progress both on distributed RDF stores and on reasoning over streaming data, with many results already published and some preliminary implementations in *RDFox*. We have improved existing algorithms for reasoning over distributed stores to enable higher expressivity in terms of the rules and queries over the data, and we have studied Stream Reasoning in an extension of Datalog for *metric temporal data*.

Dag Hovland

Dag Hovland is the Program Leader for the Semantic Integration Research Program and has published research in several areas of computer science, including logics, security, parallel programming and database theory. Currently his focus is on semantic technology with applications to user-oriented data access. The SIRIUS centre and the Optique project has enabled ongoing cooperation with the University of Bolzano and Oxford University on these issues.

We have improved the ontology reasoner Sequoia, which we first developed in 2018. Sequoia applies consequence-based reasoning to the entire OWL language. Consequence based reasoning was until recently only applicable to subsets of OWL. Sequoia already outperforms state-of-the-art reasoners on hard ontologies on some important tasks.

We have also worked more generally on the structure of knowledge bases, specifically on the recognition of strings of words ("literals") that can be made into identifiers, and on the treatment of tabular data and compressed data.

Federico Igne

I Am a DPhil student at the University of Oxford, conducting research as part of the Semantic Integration research group in SIRIUS. I am currently working on an extension of PAGOdA, a reasoner for conjunctive query answering over OWL2.

Sigurd Kittilsen

Project: Formalization and Reasoning on Requirements

What are you working on?
Using modal/deontic logic to formalize and allow reasoning on requirements.
The project is still in an early phase at the time of writing.

How is this connected to SIRIUS?
SIRIUS is working on requirements through the READI-project. The project will address SIRIUS's Digital Field Development beacon.

Temitope Ajileye

Project
Temitope is working on datalog materialisation in distributed RDF stores with dynamic data exchange, pulling together contributions from the fields of concurrency programming and data structures, distributed computing, datalog materialisation and RDF store implementations. The first version of the algorithm was published in ISWC 2019. He is currently working on a new partitioning scheme of the input data to improve the performance and resilience of the system.

How is this connected to SIRIUS?
Distributed Datalog materialisation will enable scalable reasoning of a very important OWL profile. Furthermore, datalog is a very popular logical language and the results of this work will easy to translate to other deductive database systems.

David Tena Cucala

Project
Sequoia: a reasoner for OWL 2 DL ontologies.

What are you working on?
I am developing an inference calculus based on "consequence-based" reasoning for the Description Logic underpinning OWL 2 DL, which is implemented in the Sequoia reasoner.

How is this connected to SIRIUS?
SIRIUS regularly needs to check ontologies for inconsistencies, classify them, and use them in data-intensive applications. Sequoia provides fast and robust support for these reasoning services.

JSDP at night. Photo: Øyvind Sætre, Aibel

Shuwen Liu

Project
Knowledge embedding for link prediction using Graph Neural Networks

Networks Scientific interests: Knowledge Graph, Machine Learning, Graph Neural Networks

What triggers me scientifically
To teach neural network models to gain human intelligence.

How is this connected to SIRIUS?
This belongs to the Semantic Integration research program.

Stefano Germano

Project
Logic-based Stream Reasoning

What are you working on?
Effective techniques to reason on 'data streams' in order to support real-time decisions.

How is this connected to SIRIUS?
The oil & gas companies, as well as many other companies nowadays, need to integrate, manage, and reason over a huge amount of data received continuously and in real-time to drive their business decisions and improve their processes.

Data Science

Data Science at SIRIUS is about accessing data, making use of data, and making approaches explainable. We deal with the extraction of structured data (e.g., RDF graphs) from unstructured data (e.g., text), so that this data becomes available to analytics. Furthermore, our vision is to advance the potential of data science by creating machine learning approaches that can exploit structured data (e.g., to realise distantly supervised data mining and machine learning approaches) and that take into account the structure and the meaning of structured data. We envision approaches that learn from and with ontologies: structured data where elements can have formally defined meaning - thus, domain-adapted machine learning will be enabled.

Our ambition to advance the potential of data science has three major parts:

A. Approaches based on data mining and machine learning process structured data. However, information may not always be available in structured form. For example, it may have to be identified in and extracted from natural language text, so that it can be expressed as structured data and thereupon be analysed via data mining or machine learning methods and in general be available to analytics. However, to date, due to the complexity of natural language, extracting information from text is still a hard problem and an active research area - natural language understanding is far from being solved. Besides text as a source for information, information can also reside in more structured forms such as tables (think of tables created with the intention to be read primarily by humans and tables created with the intention to be read primarily by machines, such as database tables) or in geological diagrams. Thus, it is our ambition to develop approaches that can extract information from unstructured data and represent it in structured form.

Ole Magnus Holter, our newest group member, who started his PhD in August 2019, is addressing the task of semantically parsing textual requirements in the engineering domain, which means, to transform textual statements, such as "*Shell boilers with a shell diameter of 1400 mm or greater shall be designed to permit entry of a person and shall be provided with a manhole for this purpose*" into a machine-readable representation. **Farhad Nooralahzadeh** and **Basil Ell** work on information extraction from text, see below.

B. Machine learning approaches usually require input in the form of vectors: each data point (e. g., an observation) is represented as a vector where each value corresponds to a certain feature or class. Given structured data, such as an ontology, or in general data that has the inherent structure of a graph or tree, these approaches can only work with the data once it is transformed into vectors. But then, the original structure is not available (or only to a certain extent) to the approach and thus cannot be taken into account / the approach cannot reason about the structure. In the case that structured data is available and the data mining or machine learning approach cannot handle structured data and requires the data to be flattened, the approach may be underperforming. Moreover, besides the structure, approaches typically cannot make sense of the formal semantics of data, such as axioms that can be expressed in ontologies, which is a fundamental limitation. Approaches that take into account all available data including the structure and the semantics of the data might perform better and may even require less training data. Thus, it is our ambition to develop domain-adapted machine learning approaches - approaches that can deal with, exploit and reason about the structure of data without being deprived of its inherent structure and its formal semantics.

Photo: Ole Jørgen Bratland, Equinor

Being able to combine unstructured data and structured data, especially formally represented knowledge, can help to compensate for insufficient training data. Machine Learning is data hungry, but often sufficient amounts of data are unavailable. This combination of ML and knowledge representation (KR) may enable zero-shot learning and transfer learning. For example, consider a system that was trained with one million images known to show horses or dogs so that the system learns to identify and distinguish these animals in images. One would now like to add the knowledge that a giraffe looks like a horse with a very long neck, even though training images for giraffes are not available. Instead, one would like to have an approach that exploits this knowledge so that it can identify giraffes without ever having seen one / being trained for. Furthermore, domain knowledge can help to reduce the search space in machine learning applications. If we can specify what we already know about a domain, ML approaches do not need to spend time on finding out what we know already. For example, that sandstone is a type of stone is nothing we would need an ML approach to figure out for us. Furthermore, domain knowledge can help to reduce the search space by specifying what we believe is not true so that the approach does not need to try to figure out whether it is true.

Approaches in this direction can be found under the terms neural symbolic integration, where focus is put on ML approaches based on artificial neural networks, or hybrid learning and reasoning. **Ernesto Jiménez-Ruiz** is a member of the organising committee of the 15th International Workshop on Neural-Symbolic Learning and Reasoning.

Farhad Nooralahzadeh addresses the challenge of identifying named entities in domain-specific natural language texts by making use of existing domain specific knowledge resources such as ontologies. Currently, he is finishing his PhD thesis.

Ole Magnus Holter tackled in his Master's thesis the embedding of OWL ontologies, which means to transform complex knowledge representations into vector data so that they can be processed by machine learning approaches for analytical tasks and for ontology alignment tasks. Currently, he is investigating how domain knowledge can be exploited for the task of semantically parsing of textual requirements.

In her recent publication, **Summaya Mumtaz** developed similarity functions for categorical variables that take into account structured knowledge (such as taxonomies, ontologies about categorical variables) for the task of reservoir analogue identification.

Basil Ell is developing an algorithm that mines frequent patterns from a large RDF graph. These patterns can then, amongst other things, be used to create structure-aware vector representations of such graphs. Currently, he applies the algorithm to bridging the lexical gap between natural language and ontologies via unsupervised induction of correspondences between

linguistic patterns and knowledge graph patterns. These patterns can then be used for information extraction from text and for natural language generation from ontologies.

Erik Bryhn Myklebust is focusing on ecological risk assessment. This task consists of estimating how a chemical or a mixture of chemicals can affect an ecosystem such as a lake. In his recent paper, for which he received the Best Student in-use paper award at ISWC 2019, he explored the suitability of using a knowledge graph embedding approach for ecotoxicological effect prediction.

Jiaoyan Chen is working on knowledge base curation such as erroneous fact correction, table (plain RDF triples) to knowledge graph matching and ontology matching, by utilizing both semantic reasoning and deep learning. He is also developing new neural symbolic methods for machine learning challenges like sample shortage. They are applied to solve not only the above knowledge base curation tasks, but also more general machine learning problems like transfer learning and zero-shot learning.

Daniel Bakkelund is working on increasing the power of statistical algorithms in machine learning through augmentation with ontology theory.

C. Explainability

In recent years, the explainability of complex systems such as decision support systems, automatic decision systems, machine learning-based/trained systems, and artificial intelligence in general has been expressed not only as a desired property, but also as a property that is required by law. For example, the General Data Protection Regulation's (GDPR) "right to explanation" demands that the results of ML/AI-based decisions are explained. The explainability of these complex systems becomes increasingly relevant as more and more aspects of our lives are influenced by these systems' actions and decisions. Our ambition is to make machine decisions more transparent, accountable, re-traceable, comprehensible, interpretable, explainable, and reproducible.

Basil Ell has initiated the first workshop on Semantic Explainability, which took place at ISWC 2019 in Auckland, New Zealand. Several workshops address the problem of explainable AI. However, none of these workshops has a focus on semantic technologies such as ontologies and reasoning. We believe that semantic technologies and explainability coalesce in two ways. First, systems that are based on semantic technologies must be explainable like all other AI systems. In addition, semantic technology seems predestined to support in rendering explainable those systems that are not themselves based on semantic technologies.

Peyman Rasouli and **Ingrid Chieh Yu,** who are working in the field of eXplainable Artificial Intelligence (XAI), are developing methods to explain black-box decision systems. In order to make a black-box decision system truly explainable, domain knowledge needs to be incorporated for improving the validity and fairness of explanations. Currently, Peyman and Ingrid are aiming at combining explainable artificial intelligence with semantic web technologies.

At Melkøya. Photo: Helge Hansen, Statoil

Activities in 2019

Ernesto Jiménez-Ruiz and **Jiaoyan Chen** organized the challenge SemTab: Semantic Web Challenge on Tabular Data to Knowledge Graph Matching at ISWC 2019 in Auckland, New Zealand. This challenge aims at benchmarking systems dealing with the tabular data to KG matching problem, so as to facilitate their comparison on the same basis and the reproducibility of the results (https://www.cs.ox.ac.uk/isg/challenges/sem-tab). A paper about SemTab 2019 will be presented at the ESWC 2020 Resources track. This challenge will continue at ISWC 2020.

Basil Ell, Philipp Cimiano, Agnieszka Ławrynowicz, Laura Moss, and Axel-Cyrille Ngonga Ngomo organized the first workshop on semantic explainability at ISWC 2019 in Auckland, New Zealand. This workshop will continue at ISWC 2020. (http://www.semantic-explainability.com/).

Farhad Nooralahzadeh finished his 3 months research visit in CopeNLU working on meta-learning with Isabelle Augenstein and Johannes Bjerva.

Ole Magnus Holter successfully defended his Msc thesis in May: "Semantic Embeddings for OWL 2 Ontologies". This has led to an ISWC poster paper: "Embedding OWL ontologies with OWL2Vec". He started his PhD in SIRIUS in August 2019.

Daniel Bakkelund

Project
Relation preserving clustering for identification of functionally equivalent machinery on stock

What are you working on?
I am currently working on a theory for clustering of directed acyclic graphs that preserves the graph structure. For exploratory data analysis, it is customary to discard the structure during the clustering, and this theory shows how we can adapt the existing theories in order to preserve the structure.

How is this connected to SIRIUS?
The work aims on identifying functionally equivalent pieces of machinery on stock. Machinery is composed in part-of hierarchies, where one part consists of several sub-parts. At a certain level, equivalent machinery exhibits similar inner structure. We believe that by comparing this inner structure of machinery, we can detect which pairs of machinery that are potentially functionally equivalent. But a part is never equivalent to its sub-parts: a pump cannot be replaced by one of the valves it contains. Therefore, it is necessary to preserve the part-of structure during this identification.

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 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 Ph.D 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) in May 2016 as a postdoctoral researcher. He joined SIRIUS labs at University of Oslo in September 2018 where he is leading the Data Science Research Program. Currently, he is 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) as well as domain-adapted machine learning.

Our ambition is to develop approaches that can extract information from unstructured data and represent it in structured form.

Peyman Rasouli

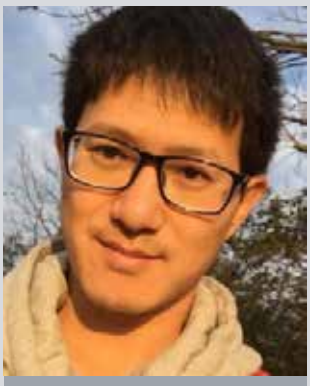
Project: A Formal Interpretable Framework for Opening Black-Box Decision Systems

What are you working on?
Recent Machine Learning (ML) algorithms like deep neural networks have been seen as black-boxes that provide impressive prediction performance while they suffer from opacity. Lack of transparency is an obstacle for their applicability in safety-critical applications such as self-driving cars and medical diagnosis. Indeed, it is important to understand the behavior of a black-box model and the reasons behind its decisions. Such an understanding of the model is useful in assessing trust, providing insights, and extracting knowledge. eXplainable Artificial Intelligence (XAI) proposes to make a shift towards more transparent AI. During this project, I research the XAI domain to create a framework for opening black-box decision systems that make them explainable, transparent, and accountable whilst maintaining high-performance levels.

How is this connected to SIRIUS?
SIRIUS partners use AI decision systems for making life-changing decisions. Providing explanations for the decisions is necessary to avoid undesired risks, expenses, and consequences. Specifically, explanations help to assess trust, which is fundamental if one plans to take action based on a decision, or when choosing whether to deploy a new system. Moreover, explanations provide insights into the system, which can be used to transform an untrustworthy system into a trustworthy one or to achieve new knowledge about the modeled process.

Main objectives for 2020

- Create a repertoire of recipes for approaches that make use of structured and unstructured knowledge together
- Develop/publish an algorithm for mining frequent graph patterns from a large RDF graph (Basil Ell)
- Carry out experiments with textual requirements (Ole Magnus Holter, Basil Ell)
- Develop a novel way to integrate taxonomies into predictions (Summaya Mumtaz)
- Develop an approach to identify inconsistencies and anomalies in knowledge graphs (Ernesto Jiménez-Ruiz)
- Extend a knowledge graph with chemical interactions and predict concentrations (Erik Bryhn Myklebust)
- Develop a method to exploit domain knowledge for explaining black box models (Peyman Rasouli)
- Study knowledge base alignment and alignment correction with deep learning and reasoning, for higher accuracy and better scalability over big data. (Jiaoyan Chen)
- Collaborate on the topic AI-Methods for Knowledge Graph Comprehension with Bosch Center for AI, Nanjing University, NTNU, and SINTEF



Jiaoyan Chen

In 2019 I mainly worked on knowledge graph (KG) curation, including extracting KG from tabular data, KG correction and canonicalisation, together with Ernesto Jimenez-Ruiz and Ian Horrocks. I finished two studies — Knowledge Base Literal Canonicalization (published in ISWC 2019) and Correcting Knowledge Base Assertions (accepted by The Web Conference 2020). These works can be applied in data management and integration, and thus are very relevant to the SIRIUS Data Science Program. I also participated the organisation of ISWC Challenge on Matching Tabular Data with Knowledge Graph which is supported by SIRIUS. It will increase the impact of our work and our research group in the Semantic Web community.

Our ambition is to develop domain-adapted machine learning approaches - approaches that can deal with, exploit and reason about the structure of data without being deprived of its inherent structure and its formal semantics.

Farhad Nooralahzadeh

Project
Adapted Information Extraction tools for Low-Resource Domains.

What are you working on?
I worked on two research areas transfer learning and distance supervision in low-resource NLP.
The results of my research are as follows:

1. Evaluation of Domain-specific Word Embeddings using Knowledge Resources
2. Named entity recognition in Low-Resource Domains
3. Relation extraction and classification in Scientific papers.
4. Cross-Domain and Cross-Lingual transfer with meta-learning

How is this connected to SIRIUS?
Central to successful data access is the production of a structured representation of relevant information on which to base subsequent action. The main focus of this project is to construct a system that will extract and make searchable factual information from large quantities of unstructured, natural language text.

In all my work, I investigate how to extend the capability of language technology tools in low-resource settings, when little annotated data is available for underlying machine learning method.

Gullfaks B. Photo: Ole Jørgen Bratland, Statoil

Scalable Computing

The Scalable Computing (SC) research program is about making data access faster; it aims to provide more efficient processing to SIRIUS projects. This is achieved by building knowledge in high performance computing (HPC) and coupling 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 automatic cross-cloud application deployment and reconfiguration.

High Performance Computing (HPC)

High-performance computing (HPC) is one form of scalable computing. The target problems of HPC involve large scale computations that are beyond the capabilities of laptop PCs or desktop computers. Moreover, close interactions often exist between the inherent components of these computations, thus the required hardware platforms for HPC are tightly coupled computer clusters, consisting of many powerful computers. The research topics of HPC encompass parallelization schemes, partitioning algorithms, communication overhead reduction strategies, software implementation and optimization techniques, use of heterogeneous clusters that consist of both conventional CPUs and cutting-edge hardware accelerators, in addition to adopting HPC for real-world applications.

The Simula team that participates in SIRIUS constitutes an integral part of SIRIUS' Scalable Computing research program. The Simula members contribute with their expertise and knowledge throughout the entire value chain of HPC: algorithms, software, middleware, hardware and applications. Moreover, the Simula members are also the connectors between SIRIUS' HPC industrial partners (Dolphin and Numascale) and the HPC research activities by assisting the testing and utilization of the industrial partners' technologies in SIRIUS' research infrastructure and concrete applications. The particular beacon project that has seen substantial contributions from the Simula team is the Digital Field and Reservoir Management project, where Simula researchers collaborate with Equinor researchers and engineers to study how to provide better HPC support for reservoir simulation. These simulations are the cornerstone in the workflow of oil reservoir management. The complexity and uncertainty of the subsurface geological properties lead to large-scale numerical computations that can only be met with HPC. Specifically, Simula researchers investigate how to improve the various algorithmic and implemental aspects of such

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 research 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 adaption, autonomic decisions, self-awareness, and emergence. Geir has previously held positions as senior scientist and research director at SINTEF in Oslo, before spending four 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.

Dolphin PCIe Gen 3.0 equipment installed in eX3

The Scalable Computing (SC) research program is about making data access faster; it aims to provide more efficient processing to SIRIUS projects. This is achieved by building knowledge in high performance computing (HPC) and coupling this with scalable Cloud computing to support scalable big-data application processing.

computations, so that reservoir simulations can be carried out faster and become capable of efficiently using more compute nodes.

Read more about the Digital Field and Reservoir Management beacon project on page 64.

Also worth reporting is that the newly established national HPC research infrastructure eX3, (**https://www.ex3.simula.no**) hosted at Simula, has been extensively used in connection with the reservoir simulation research activities in 2019. Apart from the various cutting-edge CPU architectures, the eX3 infrastructure also facilitates testing with different interconnect technologies, (including those from Dolphin), hardware accelerators, (such as Nvidia graphical processing units GPUs), Artificial Intelligence (AI) specific processors, and high-performance memory and storage technologies. All of these are of general interest to other SIRIUS researchers.

In 2019, Dolphin worked on up-scaling the PCIe (Peripheral Component Interconnect Express) systems and did accomplish to establish a PCIe Gen 3.0 network consisting of 60 nodes. This system has been used for experimental trials of new routing algorithms and techniques to improve the next generation PCIe 4.0 network. The main focus has been solving how to optimize demanding applications and increase the flexibility and dynamics for sharing PCIe IO (Input and Output), calculation and storage units.

Dolphin's PCIe Gen 3.0 network is installed on the eX3 machines at Simula and is accessible for the SIRIUS partners. The plan is to upgrade the eX3 system to PCIe Gen. 4.0 during 2020. The existing 8 node PCIe Gen 3.0 machine at Simula will be converted into an 8 node machine and used for educational purposes at UiO. It consists of 6 Nvidia Tegra Xavier nodes and 2 x86 machines. It will also be available for SIRIUS partners who want to experiment with hetero-geneous architectures.

Currently several master projects are working on software protocols for optimized use of PCIe networks.

In 2019 Numascale worked on the next generation hardware for scale-up systems. This resulted in the release of servers with 16 sockets, in collaboration with Numascale's system integration partner Atos, on the latest generation Intel platform. These servers are available in Google Cloud, AWS and at the Atos laboratory.

Numascale wants to contribute to allowing the SIRIUS partners having access to these systems. We have thus set up a protocol where the partners develop and test that their software is ready for large Scale-up systems by testing it on a previous generation Scale-up server – which is located at the University of Oslo and is maintained by Numascale in collaboration with IFI. When the software is ready for reading and using that capacity, the partners get access to the Atos servers that are at Numascale's disposal for development purposes. Equinor is very interested in testing OPM on the latest generation Scale-up. Thus, Numascale, Equinor and UiO initiated a collaboration in a master student project. Numascale also collaborates with SIRIUS related to a model based analysis of data access and further plans for memory modelling and optimization, data location, and placement of tasks connected to Numascale's work in profiling. In addition, Numascale has made a number of support libraries that enables a more efficient exploitation of the Numascale machines regarding memory allocation, linear algebra, MPI, as well as Numascope – a profile tool that follows the vmstat style and is able to provide indications of the performance related to our connection technology.

Tor Skeie

Tor Skeie is professor in computer architecture at the University of Oslo, and adjunct research scientist at Simula Research Laboratory. He is a co-founder of Fabriscale Technologies developing an innovative fabric manager and analytics platform to ensure network reliability and efficient operation of cluster and cloud datacenter infrastructures. Skeie's research has mainly contributed to the High Performance Computing (HPC) field. But, Skeie is also a researcher in the Industrial Ethernet and wireless networking areas. Some of his research results have been published in several of the most respected IEEE transactions and magazines. Skeie has for years had a strong focus on collaboration with prominent industrial partners, these collaborations have resulted in adoption of solutions and algorithms developed by Skeie. He has participated in a number of EU projects related to computing systems: MACRAMÈ (1993-1995), ARCHES (1995-1997), SIVSS (2004-2006), and NaNoC (2010-2013). Skeie has been the initiator of the following RCN-funded projects: ScaleM (verification of an innovative fabric management platform through a proof-of-concept prototype, 2012-2016), ERAC (Efficient and Robust Architecture for Big Data Cloud, 2013-2017); and the following innovation projects for the business sector: Numa-MultiConnect (Cost-efficient HPC technology with mainframe capabilities, 2014-2017), NexaScale (developing an efficient multi-port node controller technology to build scale-up servers, 2019-2021). In SIRIUS, Skeie is managing the Simula Research Laboratory involvement in the centre.

Cloud computing

The main activity of scalable computing at UiO has been centred on autonomic application management in Cloud computing through the Horizon 2020 project MELODIC. **(https://melodic.cloud/)**. Together with SIMULA Research Laboratory and partners around Europe, UiO has led the research and development of a middleware platform that is capable of deploying any application base on a model of

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 such a large number of CPU cores.

its components and their interaction, and then continuously adapt the Cloud resources given to the application based on intelligent monitoring of the application's execution context and key parameters like the number of application users, and the location and volume of the data processed by the application. MELODIC is capable of optimizing the application's resources across any combination of Cloud infrastructures, including computing resources private to the application owner organisation. Furthermore, MELODIC supports the deployment of virtual machines (VMs), containers, and serverless functions.

The MELODIC project ended in January 2020, and its successor project MORPHEMIC **(http://morphemic.cloud/),** also led by UiO, will extend the MELODIC platform to support application management and optimisation using hardware accelerators like GPUs, AI accelerator application-specific integrated circuits (ASICs) like the tensor processing unit (TPU) from Google, or field- programmable gate arrays (FPGAs) already offered by Cloud providers as part of their infrastructures. This will allow true scalability for a Cloud application both horizontally, e.g. by letting the application use more VMs, and vertically by letting the application use better and more powerful processors or HPC. The ambition is to start testing MORPHEMIC on the eX3 infrastructure at SIMULA Research Laboratory late 2020 or early 2021.

Numascale contributed substantially to the Horizon 2020 project ACTiCLOUD **(https://www.acticloud.eu/)** where Numascale facilitated the running of a numa-aware hypervisor with the Kernel-based Virtual Machine (KVM) on Numascale systems. This can be used for smart placement of virtual machines where the numa-topology is intact within each VM, through an OpenStack **(https://www.openstack.org/)** based component called ACTiManager. Given that MELODIC aleady has full support for OpenStack, another ambition for 2020 or early 2021 is to let MELODIC deploy and manage applications through the ACTiManager component and thereby create an even more powerful platform for managing HPC Cloud applications.

Marta Różańska

Project
Utility-based optimisation of Cloud application resources

Marta is a part of the scalable computing group in the SIRIUS project. She started in September 2019. She is currently working on multi-objective optimisation of Cloud application resources. More precisely, she is investigating algorithms for collocating the Cloud application components based on the coalition structure generation for games in partition.

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 is an important prototyping environment for large scale testing and benchmarking of Dolphin's infrastructure with real world applications from SIRIUS' partners.

Industrial Digital Transformation

Digital technologies offer the potential to transform organisational practices, roles and tasks. More often than not, however, the realization of this potential falls significantly short of expectations. The central challenge industrial companies face is to identify and cultivate the organizational 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.

Projects

Geological Assistant. A key challenge in subsurface evaluation is that geodata are uncertain, intermittent, sparse, multiresolution, and multi-scale. Geodata underdetermine concrete models and interpretations. This means that even though geodata may refute particular interpretations of the subsurface, they can never verify the correctness of a single interpretation. We seek to further theory on virtual work by studying how industrial geoscientists deal use digital tools and representations in their everyday work.

> **The central** challenge industrial companies face is to identify and cultivate the organizational pre-conditions necessary for realizing the potential of digital technologies.

Data-Driven Delivery, from Engagement to Steady State is a collaboration with IBM to evaluate digital tool support for improving scalable data access across all phases of an IT Strategic Outsourcing (SO) venture. Preliminary findings from this project suggest that the predominant focus on technology and tools need to be supplemented with understanding and appreciation of key mechanisms for experience and knowledge sharing in SO. Master student Trine-Lise Helgesen conducted the first stage of this study by mapping mechanisms and challenges for SO in practice.

Digital Project Delivery studies industry-wide efforts at standardization of data and information exchanged throughout EPC projects. While practically every individual company have digitalized their EPC project activities, digital delivery is the use of integrated software and processes across the project ecology, where 'integrated software' is an interconnected set of applications giving access to a shared dataset through a single user interface and delivery involves design, coordination, project management, and governance. Our emphasis in this project is to develop empirically grounded insight into the

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.

Best Scientific Paper at NIK 2019 Conference, Narvik: "Geological Multi-Scenario Reasoning".

Mina Haghshenas

Project
Digitalizing the requirements pipeline in Engineering, Procurement and Construction (EPC) projects

What are you working on?
I am involving with the READI Joint Industry Project (JIP), attempting to investigate the development of a digital ecosystem from the socio-technical perspective.

How is this connected to SIRIUS?
My engagement with the READI project can give SIRIUS an empirical-grounded knowledge on the organizational aspects of large-scale projects. This is very much in-line with one of SIRIUS aims of focusing on the industrial work practices and their improvement.

sociotechnical aspects of transitioning towards a shared digital infrastructure for delivery of EPC projects.

2019 Highlights

Best Qualitative Research Paper at MCIS conference, Naples: "Coordinating Innovation in Digital Infrastructure: The Case of Transforming Offshore Project Delivery" by Mina Haghshenas and Thomas Østerlie.

The paper reports from a large-scale collaborative project among operators, contractors, vendors, and service companies within the Engineering, Procurement, and Construction (EPC) sector of the oil & gas industry. The project's main goal is to digitalize the exchange of requirements and technical documentation throughout the EPC project lifecycle. The paper draws implications for coordinating large-scale collaborative digital innovation projects where the transformative and possibly disruptive potential of digitalisation is constrained by and need to be

Thomas Østerlie

Thomas Østerlie leads the SIRIUS research program on Industrial Digital Transformation. He is a research scientist at NTNU's Department of Computer Science. He holds a Ph.D. 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 the industry. Østerlie has worked extensively with the oil & gas industry since 2009 and has previously led strategic initiatives on innovation in organizations at NTNU Social Research.

Elena Parmiggiani

Elena Parmiggiani is associate professor in Computer-supported Collaborative Work (CSCW) and Digital Collaboration at NTNU. She is currently involved with two activities in SIRIUS. First, researching data science work practices in exploration. Second, exploring opportunities for enhancing data governance in service outsourcing organizations.

negotiated against the participating companies' existing technical and organizational investments and commercial interests.

Best Scientific Paper at NIK 2019 Conference, Narvik: "Geological Multi-Scenario Reasoning".
SIRIUS researchers won the Best Scientific Paper award at the annual NIK Conference in Narvik in November. The paper that won the award was called "Geological Multi-Scenario Reasoning" and was authored by Crystal Chang Din, Leif Harald Karlsen, Irina Pene, Oliver Stahl, Ingrid Chieh Yu and Thomas Østerlie.

Photo: Ole Jørgen Bratland, Equinor

SIRIUS Beacon Projects

The innovation work that goes on in SIRIUS is organised into beacon projects. These beacon projects are identifying broad but distinct areas where scalable data access can benefit both the industry and the wider society. The beacon projects are the defined areas where our research and skills can best solve our partner companies' industrial problems.

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.

Exploration Beacons

The Exploration beacon projects comprise a portfolio of experiment, prototype and pilot projects that support business processes in the subsurface part of the petroleum business: exploration, geosciences, reservoir modelling and wells.

- **Geological Assistant** (see page 56)
- **Subsurface Data Access & Analytics** (see page 60)
- **Digital Field & Reservoir Management** (see page 64)

Operations Beacons

The operations beacons cover the parts of the oil & 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.

- **Integrated Digital Planning** (see page 68)
- **Digital Twins** (see page 72)
- **Digital Field Development** (see page 76)

Cross-Domain Beacons

Scalable data access is a challenge in all complex organizations. The Cross-Domain Applications work package has a portfolio of projects that demonstrate the transfer of technology from SIRIUS oil & gas solutions to other economic sectors and organizations. Currently, we have projects within the areas of:

- **Personalized Medicine** (see page 80)
- **Environmental Science** (see page 85)

The aim is to set up cross-fertilization of ideas and methods between the oil & gas sector and other sectors.

Mariner field. Photo: Equinor

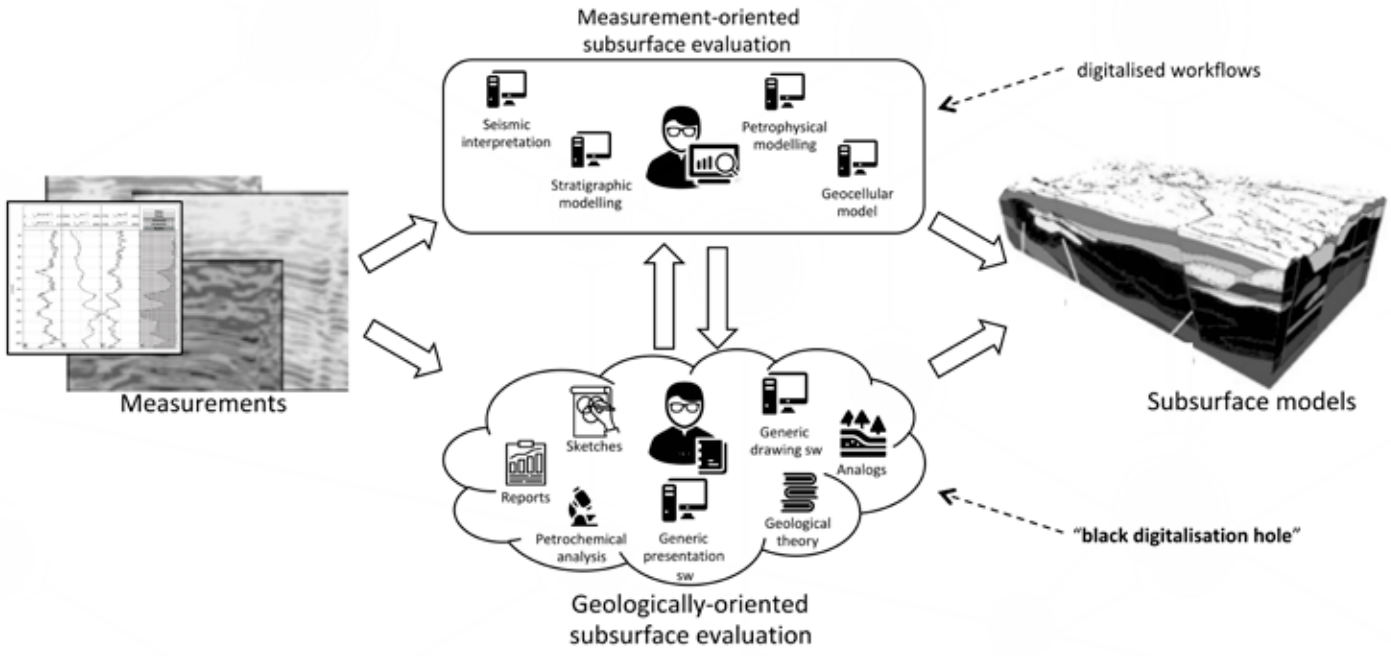
Geological Assistant

The Geological Assistant is a SIRIUS innovation project with Equinor and Schlumberger. The digital tools available for exploration geologists are not always a good match for their workflows. Most existing tools are measurement-based, and as such give better support to geophysicists than geologists. Geological thinking, rather than being measurement-based, lies closer to forms of analogical reasoning. As such, there is a need to fill a digitalisation black hole for digital tools that can support even more of the exploration geologists' workflows.

With the currently in use methodologies and tools, geologically oriented workflows in exploration prospect assessment rely on ad hoc manual work practices and tools, use pen and paper along with computer drawing and presentation tools. The Geological Assistant beacon project aims to provide the end-users with the efficient means to make the fullest use of state-of-the-art digital technologies to formulate, systematically compare, and assess different conceptual (geological) models.

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.

The GeoAssistant project is divided into a long-term research project and a short-term innovation project; Geological multi-scenario reasoning (GeMs). In this short-term innovation project, we developed a Rule-based reasoning engine that can simultaneously compare multiple geological scenarios and automatic identification



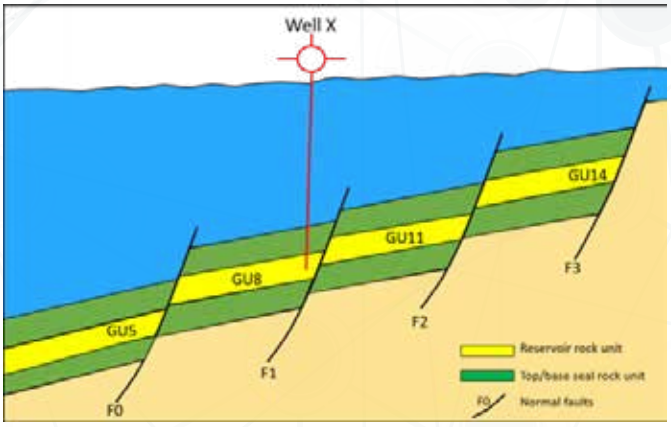
The Geological Assistant beacon project aims to provide the end-users with the efficient means to make the fullest use of state-of-the-art digital technologies to formulate, systematically compare, and assess different conceptual (geological) models.

of anomalies in these scenarios. In 2019, this methodology (and tool) was demonstrated at the Schlumberger's SIS global forum and a publication at the NIK conference which received the best paper award.

Following this success, we are now working to broaden the scope of machine-assisted support to the Geologists by populating the Geological Assistant beacon project with several complex use cases.

Use Case: Multi-scenario reasoning of hydrocarbon migration and accumulation
Users: Explorationists
Goal: Leads maturation

Geological setting – Figure 1



- A series of rotated fault blocks represented by 3 geological units
- Carrier beds / reservoirs deposited in a marine depositional environment as submarine fan
- Source rock has generated oil and gas
- Well X: found no accumulation

Unknown information to the geologists:

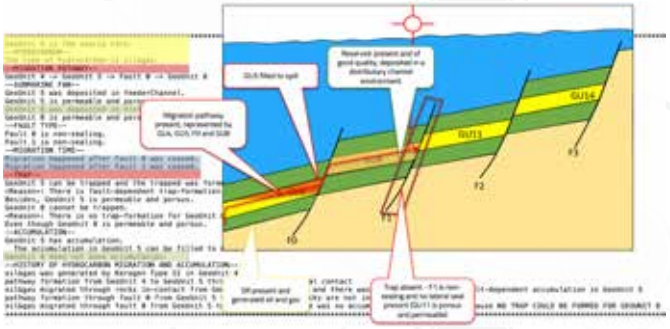
- Locations of the 3 geological units inside the submarine fan and their porosities and permeabilities.
- Sealing capacity of the faults.
- Traps configuration.
- Hydrocarbon migration pathways and migration time.

Question to be answered:
Is it possible to have an accumulation in GU11, while having no accumulation in GU8? And WHY and HOW?

Solution approach:
We apply formal methods and logic-based techniques to

- Capture underdetermination as discrete scenarios with branches of potential alternatives.
- Dynamically compute scenarios based on formal semantics of geological processes.

Multi-scenario generation: Our engine generates multi scenarios by instantiating the unknowns based on the given observations, evidence, assumptions, and the geological domain knowledge. Each scenario includes information about the migration pathway, which is represented by carrier beds and/or faults, the sealing capacity of the faults, how a reservoir is trapped or why a reservoir cannot be trapped, whether a reservoir has hydrocarbon accumulation or not, the maximum accumulation volume if any, and the simulation history of hydrocarbon migration and accumulation. Figure 2 shows how a scenario can be visually represented.



Answer to the question:
There are many scenarios showing that it is possible to find hydrocarbon in GU11; Each scenario explains why and how.

Our goal is to develop a tool-supported method for exploration geologists to better assess and evaluate exploration projects.

Jens Otten

Group/Project
Semantic Integration, Formalizing Requirements

What are you working on?
My research is focused on automating logical reasoning. More specifically, I am developing tools for automated reasoning in classical and non-classical logics, such as intuitionistic and modal logics. I have developed the leanCoP series of tools that belong to the most compact and most efficient automated reasoning tools currently available for these logics.

How is this connected to SIRIUS?
Automating reasoning is an essential task in many applications that formalize human knowledge. Within SIRIUS, ontologies are used in many projects in order to formalize knowledge. Hence, my objective is to integrate state-of-the-art reasoning technologies into these applications or integrate them into existing tools. For example, within the "Requirements" project we develop methods in order to specify formal requirements with deontic logic, a logic used to capture concepts such as obligation and permission. We are also implementing tools in order to reason in this logic. Furthermore, I develop methods and tools to extend and combine languages used to formalize ontologies, for example description logics, with more expressive languages in order to simplify the formalizing and reasoning tasks.

Crystal Chang Din

I am a researcher in formal methods.

This includes formal modelling, simulation, model checking, and deductive verification for software systems. I am part of the Geological Assistant beacon project at the SIRIUS centre. It is an interdisciplinary project. I collaborate with geologists and computer scientists from different backgrounds to develop a geological multi-scenario reasoning engine. My main contribution in the project is to formalise geological processes in rewriting logic and use it to simulate and analyse multi-scenarios. This engine assists explorationists to explore, explain, and constrain scenarios based on observations, evidence and assumptions about the subsurface.

I presented our work at the Norwegian Informatics Conference (NIK), from which we won the best scientific paper award.

Oliver Stahl

Oliver Stahl is supporting research in SIRIUS as Technical Programmer with implementation work and advice on software architecture, integration, deployment, software quality and project organization. His is or was involved in several projects such as the Geological Assistant, Geoscience Image Annotation and MELODIC. Before joining SIRIUS he worked in the industry for years as in-house developer, consultant and team lead in industry sectors including telecommunications, public health insurance, intralogistics as well as food production and logistics.

Photo: Ole Jørgen Bratland, Equinor

Schlumberger

The Schlumberger story begins with what it truly means to be a technology innovator. Our common sense of purpose unites 105,000 people representing 170 nationalities with products, sales and services in more than 120 countries. We supply the industry's most comprehensive range of products and services, from exploration through production, and integrated pore-to-pipeline solutions that optimize hydrocarbon recovery to deliver reservoir performance sustainably.

Schlumberger is raising the benchmark for continuous innovation. Our technology development, manufacturing, and lifecycle management teams work seamlessly to transform the toughest E&P challenges into ground-breaking technology solutions. Our DELFI cognitive E&P environment 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 & 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.

Subsurface Data Access and Analytics

SIRIUS is building a pipeline for data wrangling by employing ontology-based data access technology. It will demonstrate how repositories such as DISKOS can be developed into digital platforms for exploration, research, and innovation. Once this data is made available, it needs to be analysed. For this reason, we are also working with image analysis, data science, and natural language applications in sub-surface data management.

Subsurface Data Access: Data Wrangling

Digital transformation of sub-surface work processes is about overcoming the bottlenecks of data access pipelines and increasing the quality of interpretations by means of a better use of data. Data access bottlenecks mean 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 upfront, though having a Project Data Manager (PDM) helps. It is difficult to extract data from databases. Should complex 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 geoscientists must do themselves. Last but not least, it is difficult to extract data and knowledge from the text documents, as 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, the valuable analysis does not include all essential data, and conclusions may be inaccurate or, at worse, erroneous, due to incomplete data foundation.

Kvitebjørn. Photo: Harald Pettersen, Statoil

SIRIUS is working on a vision of providing a platform for sub-surface innovation. 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.

In early 2020, a partner workshop is planned to explore research opportunities related to OSDU and populate the beacon with related research/Innovation projects. Potential research areas around OSDU are currently being worked out with the industrial partners in SIRIUS. Research areas are expected to be in the intersection between digital geoscience and the SIRIUS research programs Analysis of Complex Systems, Data Science, Ontology Engineering or Semantic Integration.

Subsurface Data Access: SIRUS subsurface Lab

SIRIUS is working on a vision of providing a platform for sub-surface innovation. 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, such as DISKOS, have the potential to provide such a platform. For this to be done, however, we need to improve access to the data and allow it to be linked with data in other databases. We also need to improve access to unstructured text information in these databases.

SIRIUS has built a prototype 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 2020 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.

Irina Pene

More than 15 years of experience as a geoscientist from the oil industry, working with both operators and service companies. Since October 2017 I have joined SIRIUS, working as a researcher in multi-disciplinary teams.

I have two main projects that I am working on
Geological Assistant and Geological Images Annotation, but I am advising in all the projects where geology knowledge is required.

Geological Assistant project
I am creating models for geological subsurface data and enabling reasoning over geological processes and facts, creating ontology as input for proto-scenarios generation and working closely with the other team members when formalizing geological knowledge into rules.

Geological Images Annotation projects
I have constructed ontologies for representing the geological knowledge embedded in geological images, in order to be used in the annotation tool that out team has been developed and for creating knowledge graphs.

Subsurface Data Analytics

Faster access to relevant data is of interest only if the data can be used to generate insights and guide decisions. Emerging sciences such as 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 focusing on developing methodologies and tools for data analytics, comprising topics such as data analysis, machine learning, natural language processing, and visualization.

Applications currently in focus are:
Machine Learning-based annotation system for geological Images: Workflows used in subsurface evaluation generate a tremendous amount of image data. These images become then part of several different types of documents, e.g. PDF reports, PowerPoint slides, scientific articles and standalone images. Most of the time, information embedded in these images is complex, condensed and only in a 'Geoscientist readable form'. These images and the embedded information is extremely useful in subsurface evaluation, e.g. identifying analogues or retrieving geologic information for a specific area. To retrieve the correct images and its corresponding information from a large image database, is a difficult task.

The origin of the images is also essential for backtracking information.

Information extraction from geological images is a challenging task as these images comprise extremely complex assembles of information that is based on different conventions and standards. The challenge of information extraction stems from the fact that interpreting the information embedded in geological images requires high-level expert knowledge. For one, this creates ambiguity in the annotation process in which experts might not agree with each other on possible interpretations relative to particular features and consequently the annotations. Two, the quality of the interpretation rests on the experience and knowledge of the expert, which might vary across annotators. This creates one of the most significant problems in training machine-learning based algorithms by creating biased models.

During 2018 and 2019, we worked with IBM and Equinor to develop a tool-based methodology to convert geological images into knowledge graphs, driven by the domain ontology. This system allows end-users to manually annotate features on geological images, guided by the geoscience ontology, and generate semantic data. Further, a semantic search engine is deployed on these generated knowledge graphs to make factual information and corresponding images searchable. The results for this project were presented at the ECIM conference in 2019.

In 2020, we aim to extend this methodology to automate the process of feature annotation using Machine Learning techniques. Annotated images and the semantic data generated by the existing tool will be used as training data to train the machine learning model.

Domain-adapted information extraction for the oil & gas domain: Central to successful data access is the production of a structured representation of relevant information on which to base subsequent action. The main focus of this project is to construct a system that will extract and make searchable factual information from large quantities of unstructured, natural language text. In the 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.

Knowledge-based Machine Learning Models: The Exploration scoping workshop identified a challenge that finding and managing analogues (the geological area that resembles an area of exploration interest in some way) is a complex and time-consuming activity for exploration personnel. Methods that allow search and retrieval for analogues based on various types of similarity would improve work processes and allow faster evaluation or prospects.

The main objective of this project is to investigate techniques to identify and quantify formal domain knowledge, thus predicting more accurate parameters to be used in exploration modelling. This will enable the Machine learning model to incorporate Oil & Gas domain information and recommend analogues to a reliable extent.

The secondary objective is to propose general methods in the data science domain that can combine knowledge-based systems with machine learning systems. In complex domains, the predictions based only on data-driven approaches are often unable to quantify contextual similarities.

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 from the University of Oslo and MBA degree in marketing and finance and PMP from the 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.

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 management applies our skills in semantic technology, data science and scalable computing to improve 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 (open porous media, **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, such as reservoirs. An extensive software framework has already been developed in OPM.

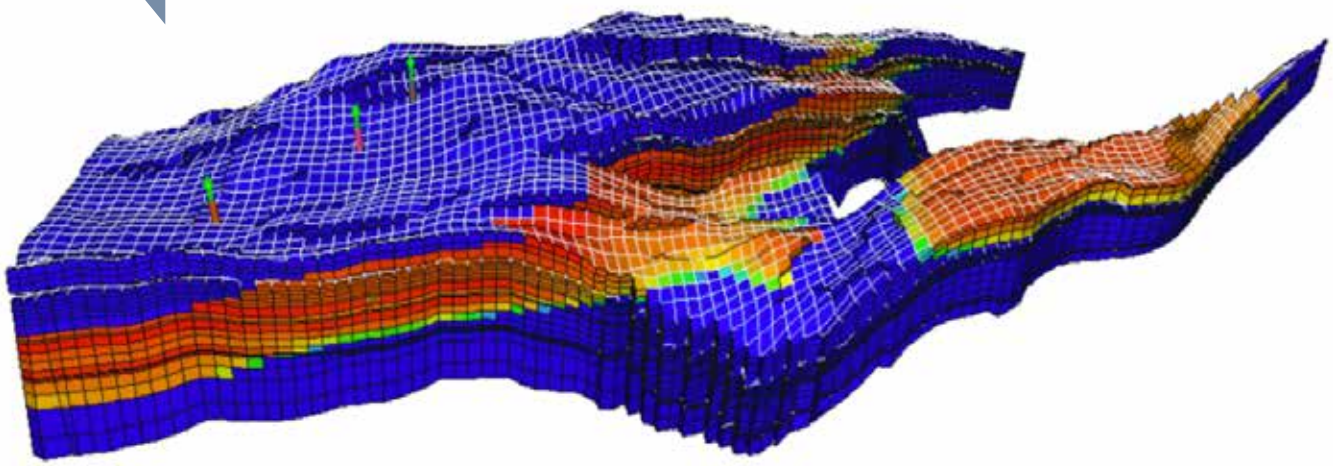
Progress/Activities 2019

The Simula team, with assistance and guidance from Equinor researchers and engineers, worked on improving the open-source software framework of OPM in 2019. The SIRIUS PhD candidate Andreas Thune continued his research work which led to several partitioning schemes, impacting numerical effectiveness and parallel efficiency, being implemented and incorporated in the OPM software framework. Implementation of optimisations in the linear solver further improved the parallel performance of the Flow reservoir simulator. These new research results have led to improvements of both the underlying numerical algorithm and the partitioning scheme adopted. The software implementation of the corresponding parts of the reservoir simulator also underwent an overhaul. Extensive experiments, involving different test problems and hardware scenarios, were done to validate the algorithmic and implemental changes.

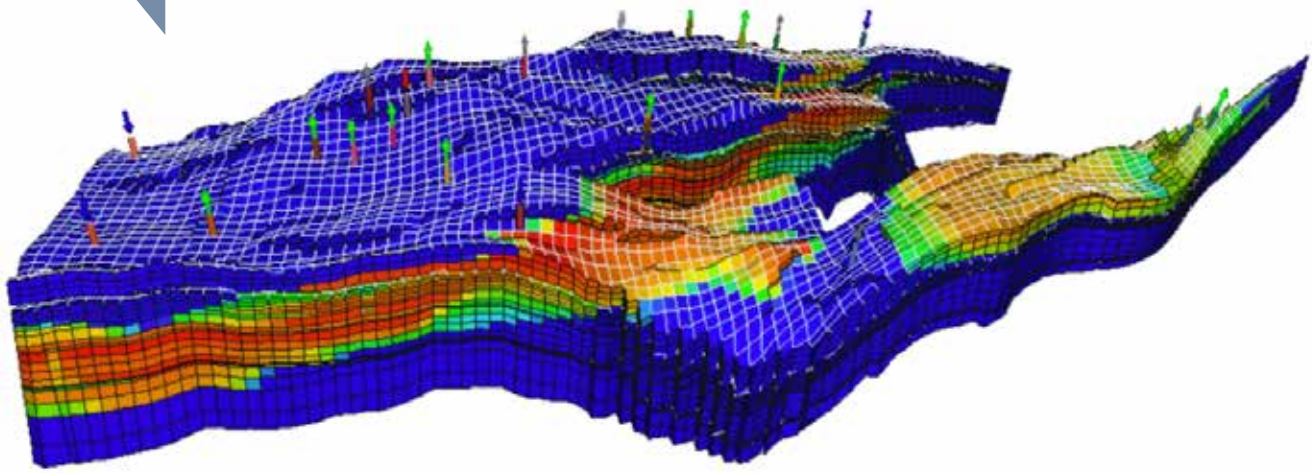
In a very fresh back-to-back comparison between the improved OPM software and a well-established industrial reservoir simulator, the Equinor staff observed a reduction of the simulation time by more than 40% due to using the improved OPM simulator. Moreover, the OPM simulator is able to utilize a larger number of processors. This substantial performance advantage and considerably better scalability of the OPM simulator are key factors that can result in a much faster workflow for the reservoir engineers.

Thus, we are now starting to see concrete and measurable results from these previous years of research and improvements in the OPM reservoir simulator capacity. For the oil & gas industry these results mean potentially big cost reductions as their operations can become much more efficient in the future.

Norne 1997



Norne 2001



Also worth reporting is that the newly established national HPC research infrastructure eX3 (https://www.ex3.simula.no), hosted at Simula, has been extensively used in connection with the reservoir simulation research activities in 2019. Apart from the various cutting-edge CPU architectures, the eX3 infrastructure also facilitates testing with different interconnect technologies (including those from Dolphin), hardware accelerators (such as Nvidia GPUs), AI-specific processors, and high-performance memory and storage technologies. All of these are of general interests to other SIRIUS researchers.

The OPM initiative represents a new community effort to promote openness in reservoir simulation research.

Summaya Mumtaz

Project
Semantic-based Machine Learning Models using domain information

What are you working on?
Machine learning algorithms are able to show good performance on data-intensive learning tasks; however, many tasks require large amounts of training data. In real-life scenarios, mostly the datasets are noisy, missing information and consists of events that occur rarely. Many business users demand background knowledge and explanations of the predictions based on ML models. These facts require specific domain knowledge to be represented and incorporated into machine learning process. Currently, I am working on developing techniques to calculate semantic similarity for categorical variables based on knowledge representation. The main focus is to integrate knowledge engineering and machine learning for complex business use cases.

How is this connected to SIRIUS?
SIRIUS provides scalable data solutions for complex domains such as oil & gas and medicine. In order to make quality predictions in these domains, there is an increasing demand to embed knowledge representation in ML. My project aims at solving these challenges for the industrial partners of SIRIUS.

Andreas Thune

Project
High-performance computing for reservoir simulation

What are you working on?
I am working on software and numerics for large scale parallel reservoir simulation. Focus is on efficient use of modern computing platforms.

How is this connected to SIRIUS?
In my project we work with digital models of real-world subsurface reservoirs. The simulation of these models are important components of the reservoir engineer workflow, and efficient reservoir simulation software is of great importance to the petroleum industry.

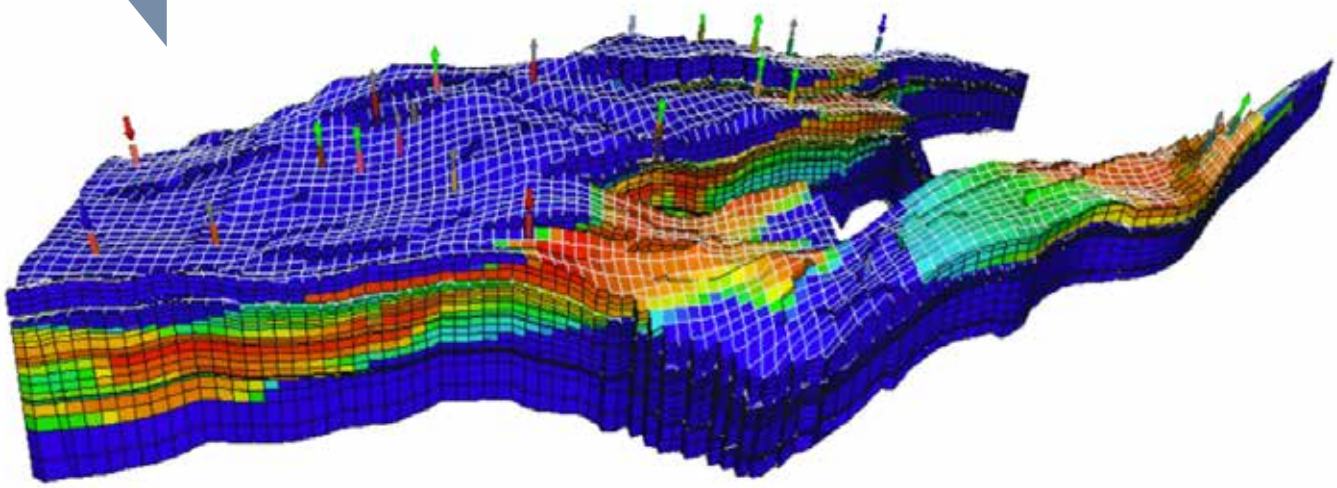
simula

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.

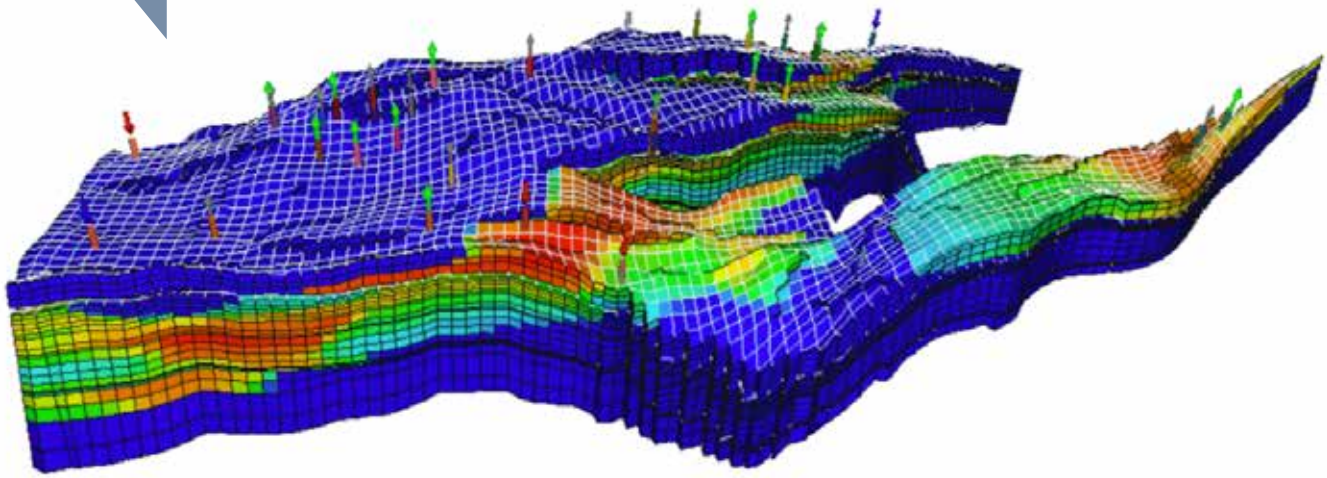
Xing Cai

Xing Cai is a professor in Scientific Computing at the 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 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 optimization". His main research interests center 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 research program of Scalable Computing through hardware-compatible numerics, parallel programming and high-performance computing.

Norne 2003



Norne 2006



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.

The four figures show the simulated oil saturation for the Norne reservoir case in, respectively, 1997, 2001, 2003 and 2006.

Integrated Digital Planning

SIRIUS' planning project has started by developing a case study that looks at vessel movements and cargo transport in the North Sea. The goal is to improve the workflow of planners at 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 first 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 are ordered by a dependency relation (see Figure 1). 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 (see Figure 2).

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

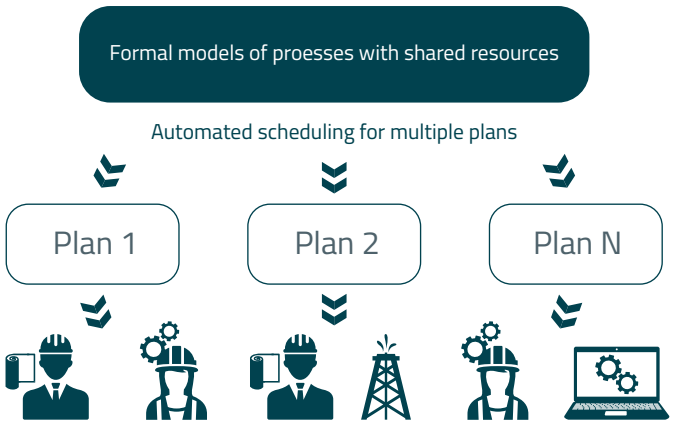
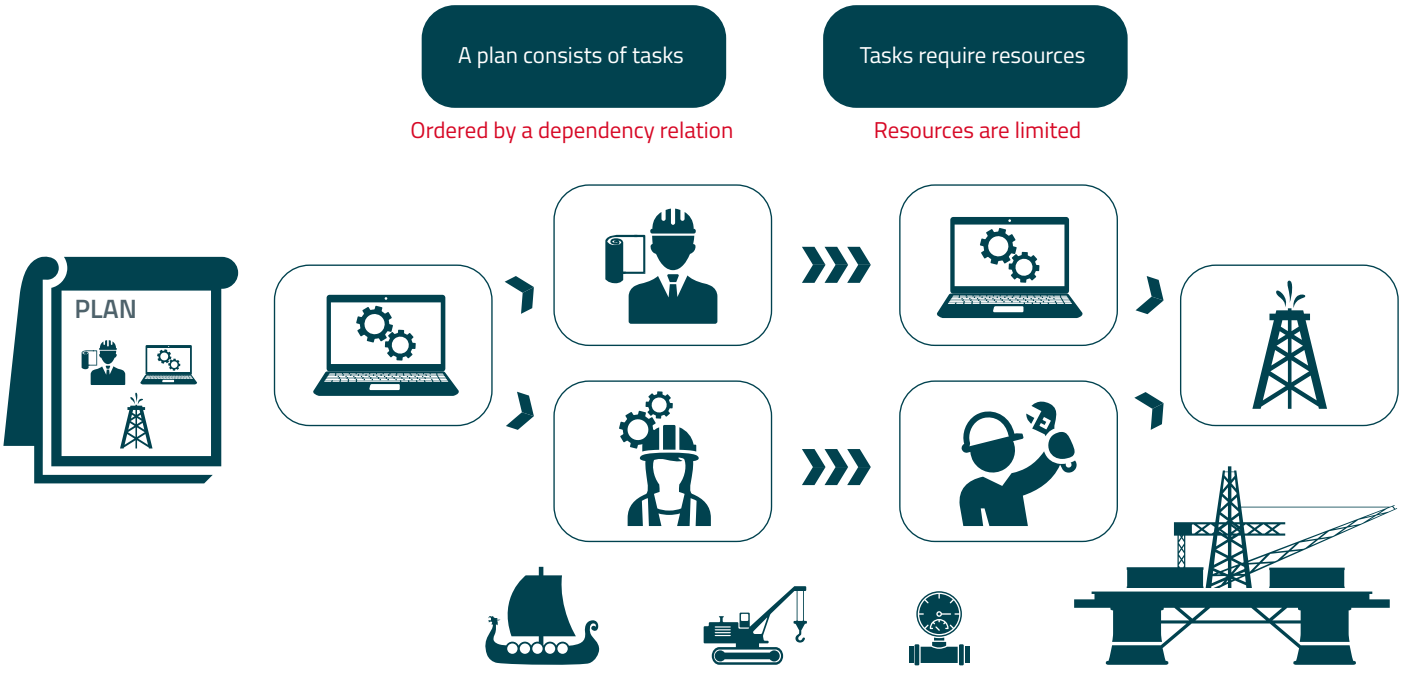


Figure 1: Plans, tasks, and resources.



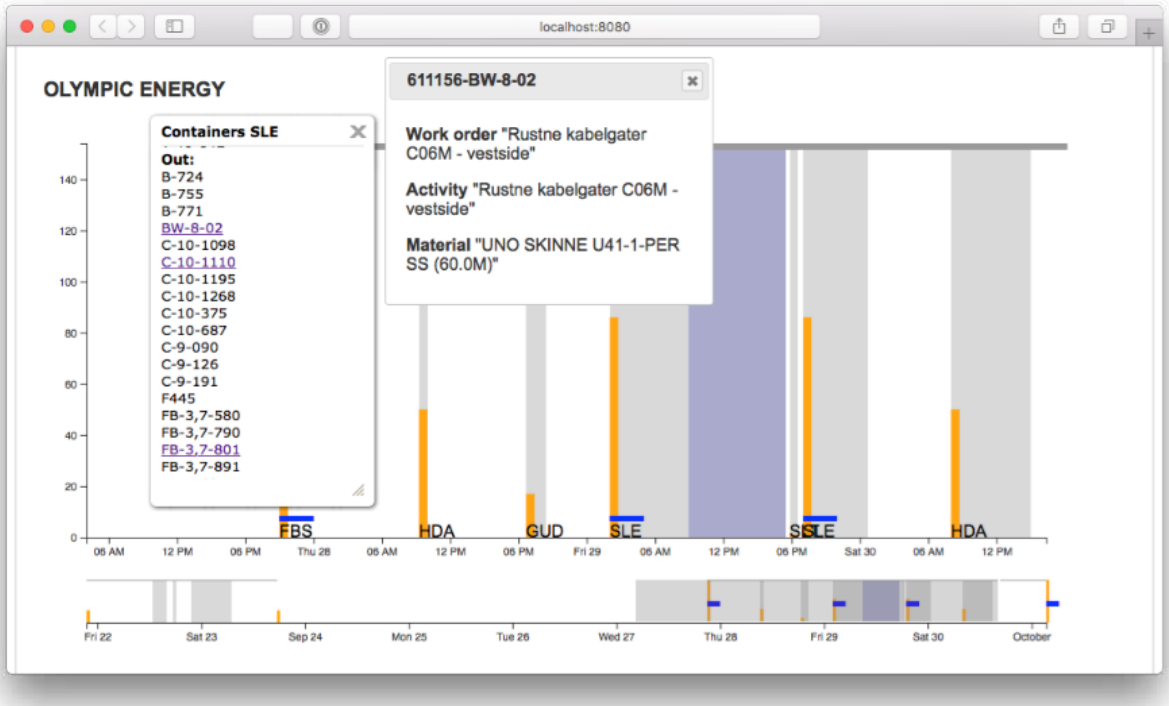
Our initial case study already illustrates the general usefulness of Real-time ABS modelling, beyond the realm of computing systems. Thus far, we have only used Real-time ABS for simulations in this case study. The case study has also driven the development of new input and output facilities for the Real-time ABS simulator, to better facilitate the interaction between model simulation and real data about 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. 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. We have begun work on exploring the use of ontologies for an overarching data model for the case study and will focus this work in 2020.

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.

The model is populated by specific data, representing a concrete plan. This is currently done by moving the data from Excel into a SQL database, then generating ABS data structures that correspond to the industrial data set. Thus, the industrial data set acts as the driver for the simulation of the ABS model. Work is underway to simplify this workflow by adding more powerful data import functionalities directly in the simulation engine, and to connect the running model with live data sources.

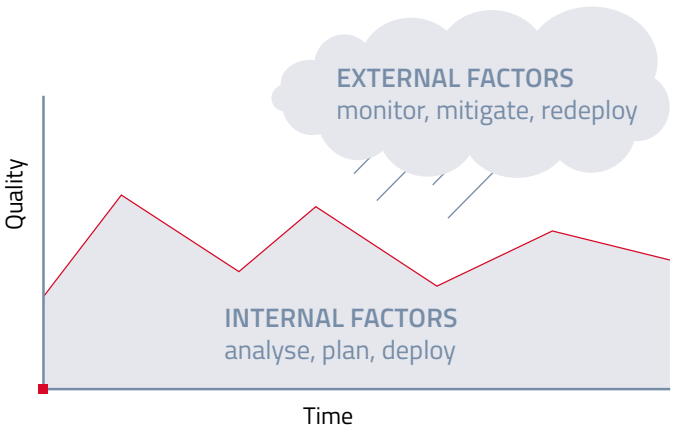
Figure 3: Visualization of time series data depicting vessel movements



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, modelling and simulation of cyber-physical processes, and analysis of operational data. 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. Rudolf Schlatte is currently working on the Integrated Digital Planning beacon project, focusing on semantic data integration processes.

Figure 4: External factors
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.



To run a simulation, the modeller specifies a time window, and data for this time interval is automatically extracted from the SQL database, converted into ABS data structures, and fed into the model as the concrete plan is simulated. This allows the ABS model of the concrete plan for the given time window to be simulated with up-to-date data. The planner is presented with a graphical view of the simulated plan, as depicted in Figure 3. Our current 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 have begun to 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. There are very promising early results in analysing the operational data using SMT (Satisfiability Modulo Theories) and OMT (Optimization Modulo Theories) techniques – we managed to automatically locate errors and ambiguities in the provided data by checking against its expected logical structure, and to calculate maximum values for various parameters. We will continue these activities in 2020 with larger data sets. Finally, it is important to remember that we can only plan up to a certain level of accuracy, and that external factors will always perturbate the plans (see Figure 4). 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 in dialog with TechnipFMC and SAP to broaden the scope of planning activities in SIRIUS.

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 behavior. He was the coordinator of the EU FP7 project ENVISAGE about the design of resource-aware services for deployment on the cloud and the analysis of service-level agreements for such systems. He 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.

SAP is the global leader in business applications and analytics software as well as 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 a SAP system. Our continued growth is attributable to relentless innovation, a diverse portfolio, our ability to anticipate ever-changing customer require-ments, 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 a 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.

Topics of specific interest for SAP are within Integrated Digital Planning, Digital Twins and Digital Field Development. For all beacons we are interested in enabling SAP Data Hub as the platform for collaboration.

Digital Twins

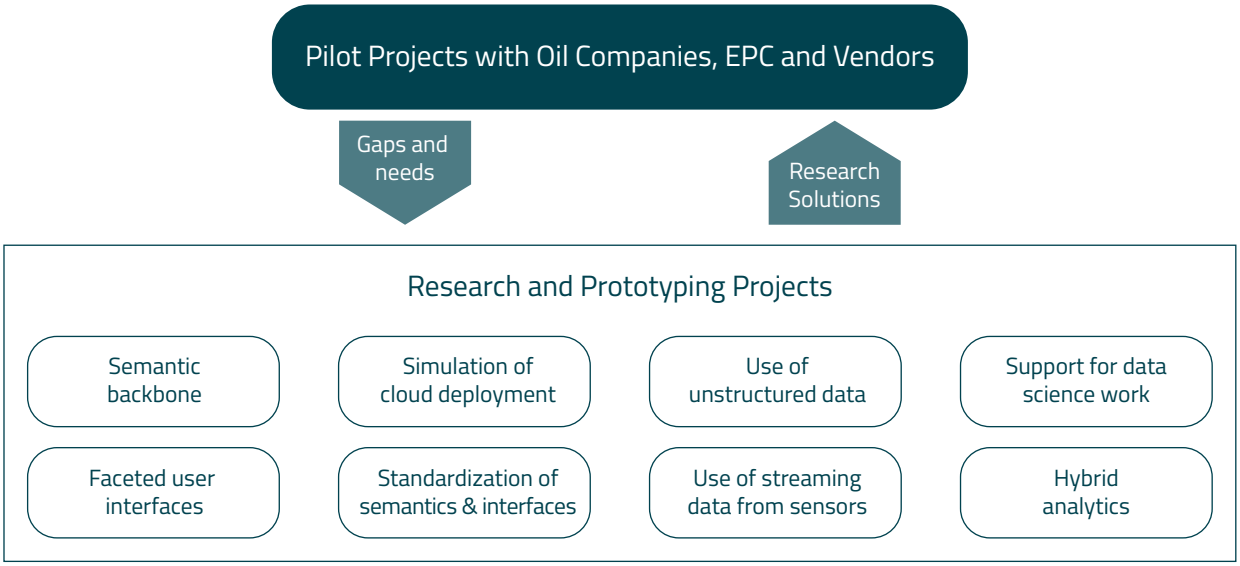
SIRIUS' work on digital twins in 2019 built on the workshop on digital twins held in 2018. We worked with existing and new partners this year to build projects that will demonstrate our vision of digital twins. This work was helped by three new companies joining SIRIUS. Aker Solutions, Aibel and Robert Bosch all bring different perspectives to what a digital twin is and what it can do. These perspectives complement Equinor's end-user view and the vendor perspectives of TechnipFMC, Computas, IBM, SAP and OSIsoft. We have also worked with other operating companies, Shell and Petrobras, to define digital twin projects for 2020. SINTEF also joined SIRIUS in 2019. They are coordinators in a European Union project called COGNITWIN. This project started in October 2019 and focuses on the process industry.

Interest in, and hype about, digital twins reached new heights in 2019. This means that SIRIUS' research agenda for digital twins is relevant for many organizations. We have presented on digital twins at a well-attended Tekna seminar held in Trondheim in April. We also presented a keynote at the SINTEF Petroleum Conference in March. The research agenda was further refined and developed through seminars and discussions with Petrobras in September.

SIRIUS follows the US Government's definition of a twin: "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.

A digital twin is a construction that stands on three pillars: operational data from the system, a description of the system's design, and simulations of the system. These three elements work together provide a mirror of the system's behaviour in real time.

However, a digital twin depends on good design data and operational data. Again, using US Defense Department terminology, a digital twin is embedded in a Digital System Model: a collection of all the data about a plant. The trouble in practice is that every source of data in a digital system model uses its own schema for the data. An item of



equipment will have a tag name in the control system, a functional location in the maintenance system and another tag name in engineering systems. All these different names need to be mapped so that the information about a piece of equipment can be integrated.

Effective, fast and accurate integration of data and models needs what we call a *semantic backbone*. This is an agreed (hopefully standard) description of the components in a plant and how they relate to each other. This is often called an asset model, a structuring of the life-cycle information about a plant.

Consider a typical piece of equipment in a plant: a pump

A semantic backbone is also needed to manage requirements in a project. Every requirement has implications for, and needs to be linked to, equipment in a facility. This means that work on digital twins overlaps and contributes to work digital field development.

This single piece of equipment needs to meet requirements on function, structural integrity and safety. The pump is an element in the overall plant design. It is linked to other equipment by piping and to the plant's electrical system. It is designed using a variety of simulations of the structure, fluid dynamics, heat transfer and electrical load. These simulations can also be used in operations for trouble-shooting and improvement. Once in operation the pump produces operational data (sensor values, alarms and events and log entries). It is also maintained through a maintenance system. Maintenance and operation tasks depend on easy access to detailed technical documentation.

A semantic backbone ties this information together and allows it to be combined in new ways. SIRIUS' research programs have found mays to make building a semantic backbone easier. The OTTR tool for building semantic models allows non-experts to work on modelling without needing to master complex, low-level coding. The OTTR templates hide the repetitiveness and wordiness of the modelling language, by using a higher-level, more concise language and reusable modules of low-level modelling language. An engineer can build a model by creating tables in a spreadsheet. These tables are then converted into a working semantic model by our software.

Semantic models are best handled by triple-store databases. This is a database that is designed to store data as groups of three data sets: a subject, an object and a predicate. For example, a pump (subject) has tag name (predicate) 21PA1002A (object). The RDFox database, developed at the University of Oxford, is a powerful and efficient database of this type. The developers of the database work on SIRIUS projects and SIRIUS-financed PhD students and researchers work on extending the database's abilities.

A digital twin is not merely a collection of all the sensor data, alarms or maintenance records in a plant. Similarly, it is not the just all the design data and 3D models for a plant. It is only as these data sources are linked to simulations that we can talk of a twin of plant operations.

Streaming data is an important element in a digital twin. Here we need to differentiate between time-series data, produced by sensors, and discrete packets of data, such as are produced by alarm systems. When computer scientists talk about streaming data, they usually mean the latter. Work in 2018 and 2019 with DNV GL, Computas and OSIsoft showed that there are many challenges in using streaming data from industrial facilities. Time-series data from specialized databases, like OSIsoft's PI system, needs lots of work to make it suitable for use in machine learning. The sensor data must be put into context. For example, what does a value of zero mean? Is the sensor faulty? Was the equipment shut down? Has the equipment failed? Is the sensor properly configured in the database? Sensor values must be seen in context with each other and with the operational history of the plant. Here we need to work on hybrid analytics: where machine learning is enhanced by the insights from physical models and semantic models.

Much information produced in design and operation is text and other unstructured data. Requirements decisions, logs, reports and maintenance requisitions are all text. These use technical terminology. Operational records can be in Norwegian dialect. A move to a semantic backbone will need quick and easy ways of converting existing text and documents to structured data. Fortunately, many technical documents and specifications are, by their nature, formulaic and structured. Our data science program is working on improving the performance of language recognition on technical language. We are also looking at how language processing can be used to convert requirement specifications into a semantic database representation.

2019 was a year where SIRIUS established projects that will drive our digital twin activity and its innovation goals. The following innovation projects were defined and will be run in 2020.

SIRIUS workers have been part of a Konkraft working group that has discussed how semantic technology could be used to implement a digital design basis for early-phase field development. Here an oil company prepares an information package that describes the requirements and scope for a field. This package is sent to suppliers, who then work with the oil company to investigate concepts and prepare a costed conceptual design. This design is used to decide whether the project will proceed or not. The working group contains representatives of Aker Solutions, TechnipFMC, Aibel, Equinor, Lundin and AkerBP. In 2020 we expect to demonstrate the feasibility of this approach through a SIRIUS innovation project.

Our research strategy is to work alongside our partners' i nnovation and digitalisation initiatives. In 2019 we worked with Aibel to introduce SIRIUS tools into their development of semantic systems for managing materials in engineering projects and creating asset models of field developments. Aibel and Equinor are working together on a semantic digital twin for an ongoing field development project. SIRIUS researchers will be providing our tools and this work in 2020.

Robert Bosch GmbH joined SIRIUS in December 2019. As a leading supplier of components for engineering and automotive, they have an ambitious research program in artificial intelligence and digital twins. In 2020 we will start projects, with Bosch-financed PhD students working on developing our semantic technologies using Bosch products and software for prototyping.

Finally, in mid-2019, we were awarded a large Norwegian-Brazilian research project. This was a knowledge-building project for the petroleum industry and was financed by the Research Council of Norway and FINEP, the Brazilian federal innovation agency. The project, called PeTWIN, aims to provide a research-based foundation for digital twins used in the management of complex oilfields. The University of Oslo is partnered with the University of Rio Grande do Sul and three oil companies: Equinor, Shell and Petrobras. This project finances four researcher positions in Oslo and a corresponding set of resources in Brazil. Our plan is that the PeTWIN activities will act as a hub for all our other digital twin projects.

Evgeny Kharlamov

Evgeny Kharlamov is a Research Scientist at the Bosch Centre for Artificial Intelligence and an Associate Professor at the University of Oslo. Before joining Bosch he was a Senior Research Fellow at the University of Oxford. Evgeny's work is focused on AI, knowledge graphs, semantics technologies, data mining, machine learning, and graph analytics, with applications in Industry 4.0 and digital twins. Evgeny has 120+ publications, won the best in-use paper award at ISWC'17, was a runner-up at ISWC'16, and won the best demo award at ISWC'15. He was in PC of various international venues including ISWC'19, ISWC'20, CIKM'20 and gave keynotes at KESW'13 and JIST'19. Evgeny played a leading role in Oxford's participation in a large-scale EU funded Optique project on end-user oriented access to industrial Big Data. He was the (co-)PI on several projects including an industry-funded with Siemens Corporate Technology on semantic diagnostics of complex industrial production systems and on analytics and model management.

Founded in 1980, OSIsoft has consistently focused on a single goal: to develop software that transforms data streams from critical operations from sensors, devices and industrial processes into rich, real-time insights to help people save money, save lives, increase productivity and create connected products. Over 1,000 leading utilities, 80% of the world's largest oil & gas companies and 65% of the industrial companies in the Fortune 500 rely on OSIsoft's PI System and other OSIsoft technologies. Worldwide, over 2 billion sensor-based data streams are managed by the PI System. For more, please visit www.osisoft.com.

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% have 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. In 2019 we participated in the digital twin beacon as well as the digital field development and personalized medicine beacon projects. These initiatives are well aligned with the company's business strategies and will further strengthen our commitment to SIRIUS.

Digital Field Development

The oil & gas industry has recognized that it needs to change the way of work in order to remain competitive. In our 2018 report we presented SIRIUS' view of the report from the Konkraft initiative on Competitiveness – changing tide on the Norwegian continental shelf. This report led to the creation of industry-led initiatives to revise and digitise standards for project documentation. SIRIUS' Digital Field Development beacon is tightly linked to the initiatives for revising and digitizing the NORSOK standards for technical information.

This set of standards includes the standards for handover of information from contractor to operator, the code manual that is the basis for tag numbering systems and the standards that regulate the flow of technical information.

The vision is shown in the accompanying figure, which is taken from the Konkraft report above. The figure depicts a scenario, with high level of machine support, where requirements and specifications flow downwards in a system engineering's V-model and are verified against documentation of deliverables.

In order to concretize and realize this vision, a broad group of companies established the READI joint industry project. The project is led by the SIRIUS partner DNV GL. In preparation for READI, DNV GL invested in a SIRIUS project that developed and prototyped a tool, called DREAM, for defining and accessing digital statements of requirements.

SIRIUS researchers and practitioners from SIRIUS partners have made essential contributions to the methodology

> **The SIRIUS** Digital Field Development beacon channels research of direct relevance to READI from the ontology engineering, industrial digital transformation, domain adapted data science and semantic integration programs.

Equinor (former Statoil) has a leading role in SIRIUS. As the major operating company on the Norwegian Continental Shelf, they set the agenda for the 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 the 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.

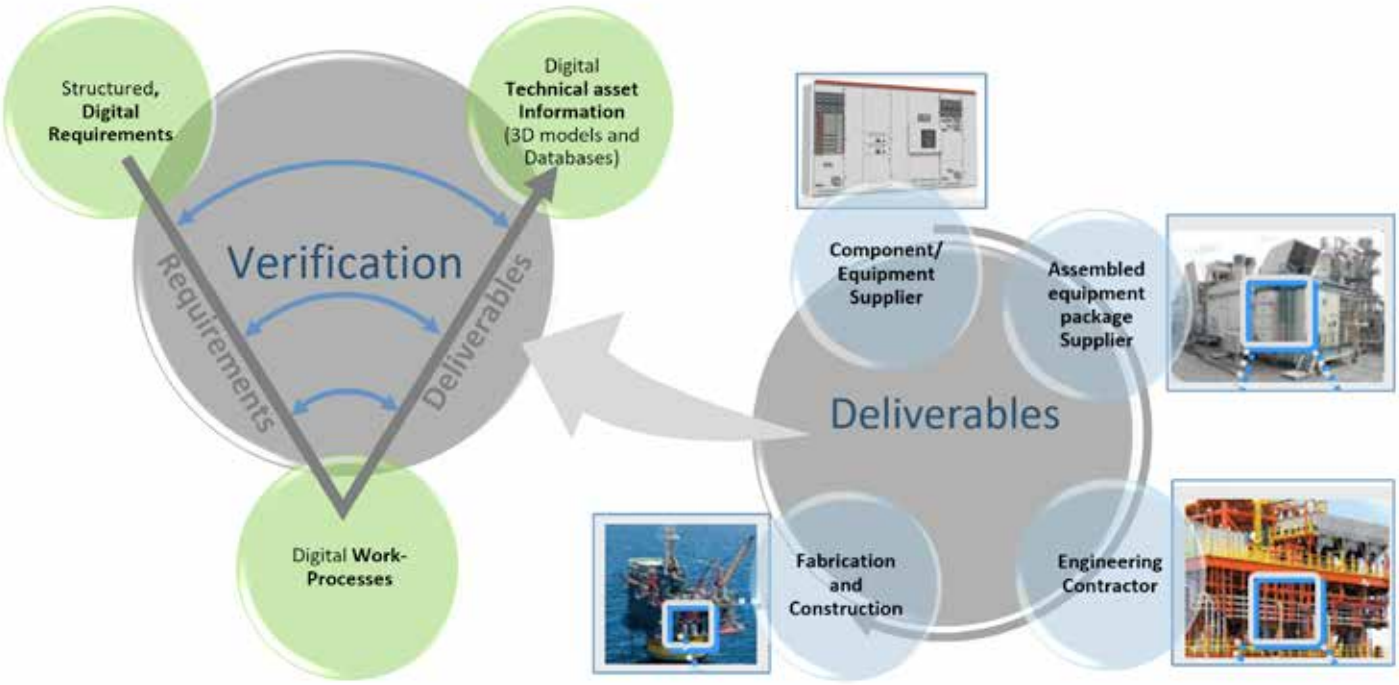


Figure 1: The V-model verification process flow vision as illustrated in the Konkraft report

of READI and in developing tools to support the work. The SIRIUS Digital Field Development beacon channels research of direct relevance to READI from the ontology engineering, industrial digital transformation, domain adapted data science and semantic integration programs. Our research addresses fundamental questions related to representation and verification of requirements, methods for design of shared asset models, tools to support the workflow, and transition strategies for digital transformation of industry practice.

At present, most requirements, specifications and documentation of deliverables are documents that are made to be read by human experts. To increase the level of machine support, requirements must be represented in a structured, machine readable form. This means we need a shared, rigorous vocabulary for classes, properties and relations: - a reference data library. We are building on experience from SIRIUS partners Aibel, Aker Solutions and DNV GL to use ontologies for building reference data libraries that are consistent with the international standard ISO 15926.

Rustam Mehmandarov

Project
Digital Compliance Management for Operations (DCMops)

What are you working on?
Creating a framework for making the engineering requirements machine-readable. We are using semantic technologies and modern architecture and data integration patterns to enable the information model integration and exchange in the industry.

How is this connected to SIRIUS?
SIRIUS is all about enabling digitalisation in and beyond the oil & gas industry. This is one of the building blocks to achieve this goal.

We use a simple encoding of requirements and specifications as rules using the **Scope-Condition-Demand (SCD)** pattern. Scope defines what the requirement applies to. It is given by a class that refers to a thing or place: a valve, a cooling system, a location that requires gas detection. The condition expresses when or where a requirement applies. For example, one requirement for a valve is needed only if the valve is in an explosive area. Likewise, another requirement may apply only if an item exceeds a certain weight. Finally, the Demand term describes the required, or specified, state of affairs. This usually means that the item in Scope has some property.

Semantic technology is an ideal match for this form of representation. It offers in-built mechanisms that can be applied for reasoning over requirements and for inheriting requirements. We have demonstrated that this enables powerful automatic verification of requirements, provided that we have an asset model in an appropriate form. However, it is not easy to design asset models.

Our research addresses fundamental questions related to representation and verification of requirements, methods for design of shared asset models, tools to support the workflow, and transition strategies for digital transformation of industry practice.

There is no single source of truth for an asset. There is also no single name or representation for a component in a system. Information about an asset is fragmented and spread over many different sources. Just within on EPC contractor there are more than 300 specialist applications in use, each of which contains a small part of the total body of information about the asset. We therefore need to address the problem of how to put different pieces of information together in a coherent way to obtain a wider view than one can get from individual sources in isolation.

Our approach builds on two principles: - (1) top-down modelling and (2) reuse of model blocks.

Top-down modelling uses systems thinking and the ISO/IEC 81346 standard Industrial systems, installations and equipment and industrial products – Structuring principles and reference designations. This leads to a step by step decomposition of systems into finer system elements in a flexible way. The method allows us to introduce new breakdown structures as needed to reflect new perspectives. In this way, we can extend the asset model incrementally from the conceptual design basis through the EPC process and further into operation. Importantly, we aim to model the asset at different levels of granularity to reflect the nature of the engineering process.

Ole Magnus Holter

Project
Semantic parsing of textual requirements

What are you working on?
I am a PHD candidate and my project is involved with the topic of using Natural Language Processing (NLP) techniques on domain-specific textual data and with the automatic generation of structured data that can be integrated into an existing ontology. More concrete, I investigate how we could use NLP together with semantic technologies to extract information and generate structured data from existing requirements documents.

How is this connected to SIRIUS?
The PHD project is positioned in the Data Science research program and is relevant for the Digital Field Development beacon project, where SIRIUS is already involved with work on digitalizing requirements.

We are currently collaborating with Stiftelsen Det Norske Veritas (DNV GL) and Equinor.

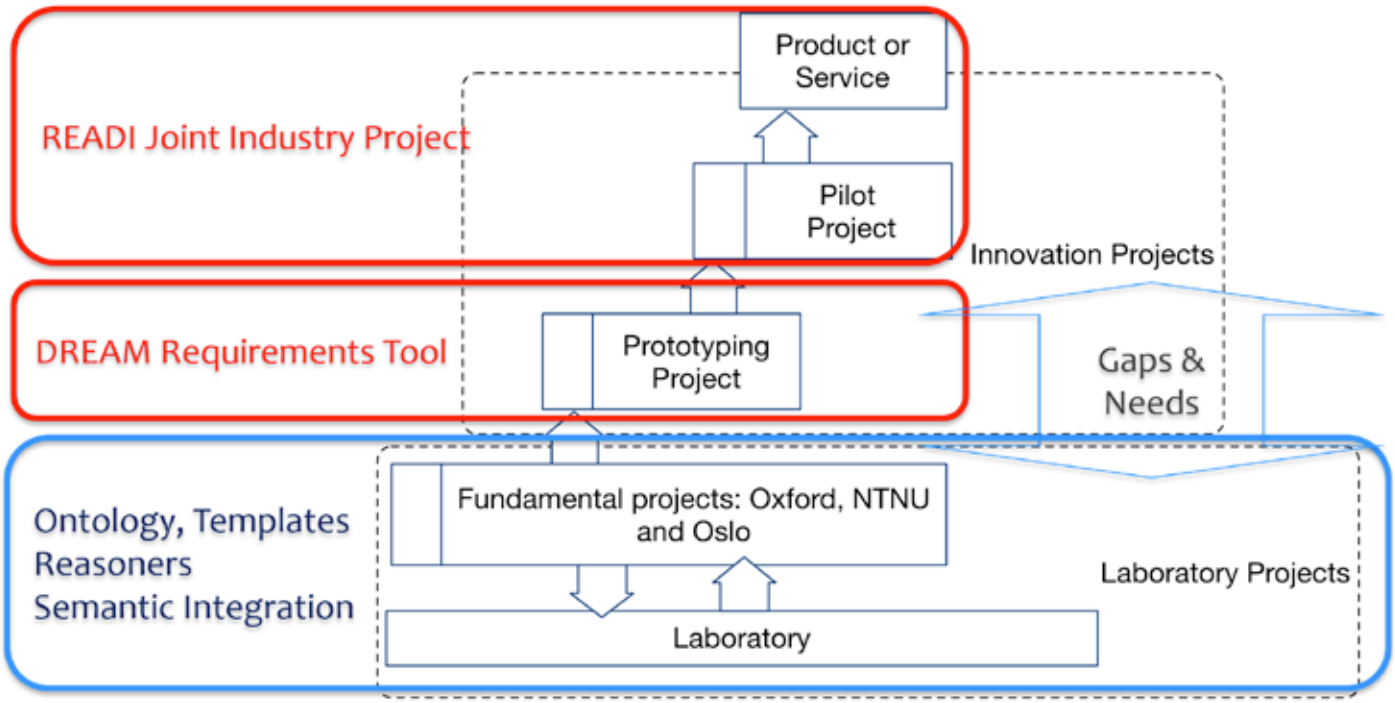


Figure 2: The SIRIUS innovation model demonstrated through the Digital Field Development beacon project

An asset model of an oil & gas facility contains many similar blocks of semantic modelling. We reuse these model blocks using ontology templates. The idea is that partial asset models can be identified, with appropriate interfaces. These partial models can then be plugged into another partial asset model. This approach mimics the way engineers define, specify and refine the design on an asset. The OTTR templates technology developed at SIRIUS provides a unique way of doing this.

The Digital Field Development beacon demonstrates SIRIUS' innovation model in practice. Our fundamental work on ontology templates and reasoning was prototyped by DNV GL in their DREAM tool. This successful prototyping was important for the funding of the READI joint industry project, which works with piloting and preparations for commercialization.

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 systems 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.

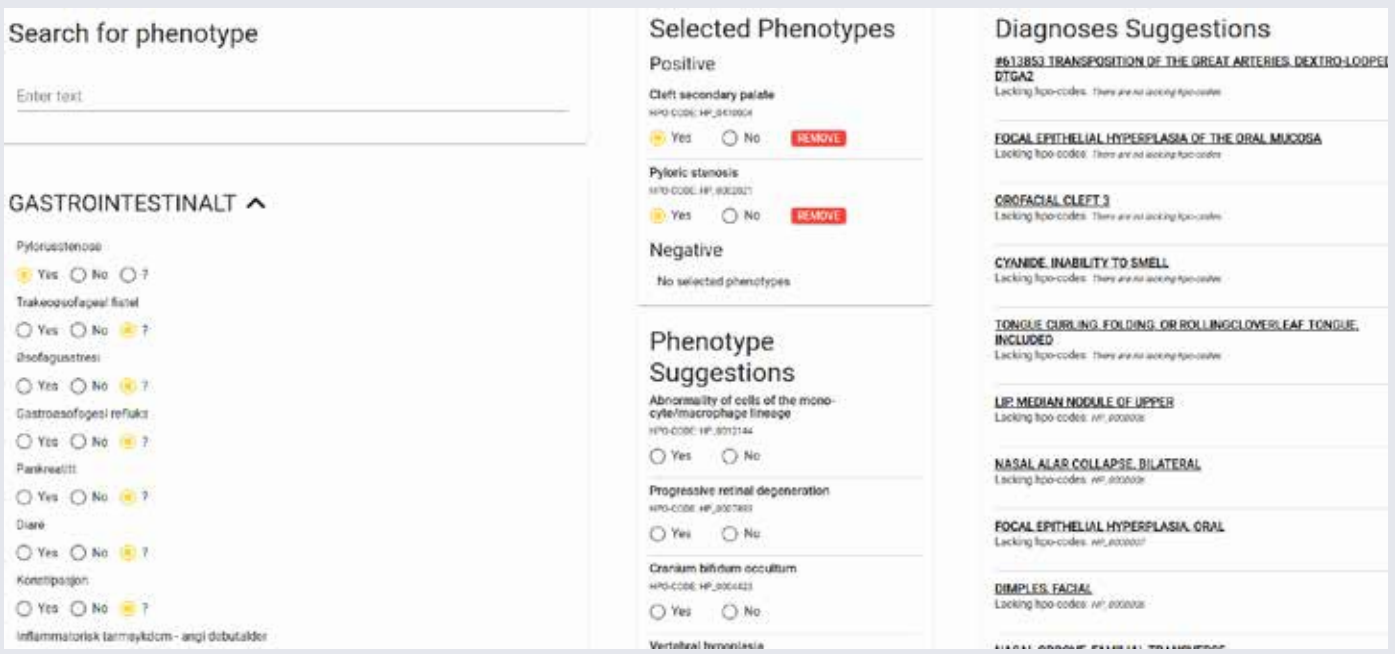
Personalized Medicine

Personalized Medicine (PM) is an approach to medicine that takes into account individual variability in genes, environment, and lifestyle. The goal of PM is to improve treatment outcomes and reduce adverse events. This approach attempts to predict an individual's response to specific treatments and prevention strategies. Personalized medicine is also known as precision medicine and has focused on big data and use of machine learning methods, studying more common complex conditions such as heart disease, diabetes, and high blood pressure. Currently, the role of precision medicine in day-to-day healthcare has not been widely implemented. In the coming years, the hope is that this approach will expand to many areas of health and healthcare.

BIGMED Beacon Project

We have been working on a tool to improve the process of capturing patient phenotypes and communicating these with the genetics laboratory. This work has been done in collaboration with paediatricians at Oslo University Hospital. We have created a suggestion service for providing pointers to further work-up needed to diagnose newborns suspected of having a genetic disease. We are working towards improving the process of creating a more complete representation of the patient characteristics using the Human Phenotype Ontology (HPO).

This work was completed within the Semantic Integration research group. We have gained experience working with a large ontology, the Human Phenotype Ontology (HPO). The HPO provides a standardized vocabulary of phenotypic abnormalities encountered in human disease, for example code HP:0001631 Atrial septal defect. The HPO currently contains over 13,000 terms and over 156,000 annotations to hereditary diseases. We have implemented consistency checking using the Sequoia Reasoner, a consequence-based ontology reasoner from the Oxford group. We are completing the project by conducting an evaluation of our suggestion tool. Through



The Genetic Testing Requisition User Interface

collaboration with DIPS AS, a leading supplier of eHealth systems to Norwegian hospitals, we have contributed to the design of the user interface that implements the suggestion service as part of their genetic test requisition tool.

Explainable AI in Clinical Decision Support

Traditional Clinical Decision Support Systems (CDSS) have been black-boxes, leaving medical professionals without a clear idea of how the machine calculated specific diagnoses or treatment recommendations. Physicians are ultimately responsible for the medical care of their patients, and therefore will override CDSS output when they feel it is wrong. Explainable Medical AI (XAI) systems improve the usability of CDSS by providing clinicians an explanation of the reasoning and the logic behind recommendations and results given.

SIRIUS researchers have begun exploring XAI methods that incorporate new explanation techniques with the results produced by the machine. The objective is to clarify and expand on underlying models and mechanisms built into the CDSS that doctors use as a part of their work-flow. Our research is showing us better ways to support problem solving and decision making within the domain expert's work tasks. This includes question anticipation and inferencing, as well as argumentation systems for identifying arguments and counterarguments instead of displaying one final machine-produced solution.

Scoping Activities

SIRIUS researchers have participated in scoping activities related to health registries. The ontology engineering group has tested an application of Reasonable Ontology Templates (OTTR) to the management of health registry data. We are also further exploring applications of OTTR within other medical areas, specifically pharmacogenomics.

The Data Science group at SIRIUS has explored the use of machine learning and predictive analytics using the MI-MIC-III dataset. This dataset is openly available, developed by the MIT Lab for Computational Physiology, and comprises de-identified health data associated with ~60,000 intensive care unit admissions. It includes demographics, vital signs, laboratory tests, medications, and more. Summaya Mumtaz conducted initial explorations, including network-building meetings with researchers working on the same dataset at the Health Informatics and Technology, University of Tromsø.

Laura Slaughter

Laura Slaughter is the Cross-Domain Applications Leader at SIRIUS. Her main role in SIRIUS is to establish projects that support further dissemination and uptake of SIRIUS technologies by promoting their use more generally. Her research interests fall within the areas of semantic integration, ontology engineering, and data science. She received her PhD in Information Science from the University of Maryland in 2002. This was followed by post-doctoral training at the Department of Biomedical Informatics, Columbia University, New York. She has 15 years of experience in Healthcare Informatics. This past year, her research focus has centred on the BIGMED Beacon Project and Precision Medicine.

With a then-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 & gas and healthcare.

Evry is one of the operators of the DISKOS system and wishes to contribute to making it a platform for the petroleum 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.

We have created a suggestion service for providing pointers to further work-up needed to diagnose newborns suspected of having a genetic disease. We are working towards improving the process of creating a more complete representation of the patient characteristics using the Human Phenotype Ontology (HPO).

DNV·GL

DNV GL is a global quality assurance and risk management company. Driven by our purpose of safeguarding life, property and the environment, we enable organizations to advance the safety and sustainability of their business.

Combining technical, digital and operational expertise, risk methodology and in-depth industry knowledge, we empower our customers' decisions and actions with trust and confidence. We continuously invest in research and collaborative innovation to provide customers and society with technological and operational foresight.

With origins stretching back to 1864, and operations in 100 countries, our professionals are dedicated to helping customers make the world safer, smarter and greener.

DNV-GL's primary involvement in SIRIUS is in the area of ontology engineering and its application of field development and digital twins. We have also collaborated in the Personalized Medicine Cross-Domain Beacon through the BigMed project.

Environmental Science

SIRIUS is active in building up a portfolio of projects within the broad area of environmental research. We apply techniques and methods to help extract the most relevant data, predict extreme events, and accelerate our ability to make predictions and model impact. Our most recent project activities have focused on environmental impact and pollution. At SIRIUS, we hope to stimulate discussion of new ideas and foster new collaborations between environmental science and informaticians.

Environmental projects can be related to forestry, agriculture, climate, pollution monitoring, and weather. Related to this, we wish to contribute to a new subfield, Climate Informatics, which is gaining momentum. The work done in this field refers to any research combining climate science with approaches from statistics, machine learning and data mining. AI techniques are used to understand the effects of climate change on weather patterns and ecosystems.

Artificial intelligence aids Environmental protection

As many as 150 million natural and synthetic compounds are currently registered, whereof more than 200 000 of these are used in high volumes for a variety of human activities. Many of these compounds find their way to our waterways, oceans, soil and air, and are potentially hazardous to humans and wildlife after exposure. It is becoming increasingly evident that these challenges require new approaches that can handle increasing amounts of data and address complex environmental issues. Artificial intelligence (AI) is rapidly paving the way for a new generation of hazard and risk assessments, where computer-aided approaches are expanding the ability to identify and characterize potential environmental impacts. SIRIUS collaborates with the Norwegian Institute for Water Research (NIVA) to develop tomorrow's solutions by educating today's students within the expanding field of AI. This training and R&D activity have developed and implemented approaches such as graph embedding to infer and extrapolate missing data in order to expand the knowledge domain for chemical hazards (toxic effect). This effort typically involves the integration of disparate data sources with well-established methods and tools to support data-gap filling in hazard and risk assessment processes. Furthermore, state-of-the-art machine learning models are applied over the integrated data to estimate chemical toxicity. The work by PhD student Erik Bryhn Myklebust and others is anticipated to dramatically enhance our ability to perform sound impact and risk assessments, and to strengthen the application of AI into more cost-effective environmental protection initiatives.

Typical cleaning and settling ponds receiving road pollution run-off (Photo: Sondre Meland, NIVA)

Pollutant emissions from natural and human activities are challenging environmental and human health (Photo: NIVA)

Erik Bryhn Myklebust

Project
Knowledge Graph Embedding and Chemical Effect Prediction

What are you working on?
The project aims at developing models which learn structured data (graphs) and that in turn can be used in conjunction with chemical data to predict hazardous compounds for species and ecosystems.

How is this connected to SIRIUS?
The project requires large amounts of data integration, which is related to the ontology engineering work at SIRIUS.

In addition, the combination of structured and unstructured data in machine learning approaches is one of the major objectives of the data science group.

Finally, this project provides a novel application for the work at SIRIUS outside the oil & gas domain.

Ernesto Jimenez Ruiz

Ernesto Jimenez Ruiz is a Lecturer in Artificial Intelligence at City, University of London affiliated to the Research Centers for Machine Learning and Artificial Intelligence. He is also a researcher in the Centre for Scalable Data Access (SIRIUS) at the University of Oslo, Norway. His current research interests focus on the application of Semantic Technology to Data Science workflows and the combination of Knowledge Representation and Machine Learning techniques. His work is very relevant to the Data Science research program and the beacon project on Environmental Applications (collaboration with the Norwegian Institute for Water Research, NIVA).

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.

Large-scale deaths of frogs after emissions of road pollution run-off. (Photo: Susanne Lund Johansen, Statens Vegvesen)

NIVA

NIVA is not a SIRIUS partner in the traditional sense. SIRIUS is supervising and funding Erik Bryhn Myklebust's PhD at NIVA. Erik collaborates tightly with the Data Science researchers in SIRIUS, sharing methodology and application areas.

Norwegian Institute for Water Research

The Norwegian Institute for Water Research (NIVA) was founded in 1958 and is Norway's leading institute for fundamental and applied research on marine and freshwaters. The institute's research comprises a wide array of environmental, climatic and resource-related fields and combine research, monitoring, evaluation, problem-solving and advisory services at international, national and local levels.

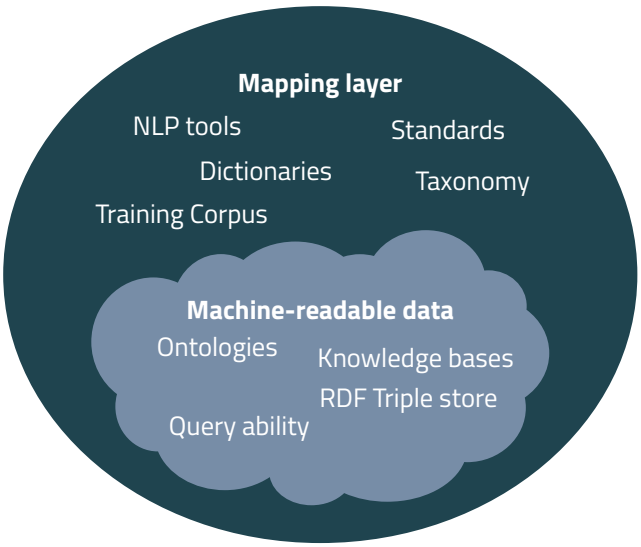
International Activity and Dissemination

Semantics for the Resources Industries

In June 2019 SIRIUS was visited by Professor Melinda Hodkiewicz from the University of Western Australia. She leads an Australian Research Council centre that works on maintenance in heavy industries. Here Melinda explains her work and how her group is collaborating with SIRIUS.

The economies of both Australia and Norway are underpinned by their resources industries. In Norway, oil companies expect to spend 175.3 billion kroner in 2019[1]. In Australia, there has been A$514 billion investment in the resources sector since 2009[2]. Getting value from these investments depends on how they are operated and maintained. We are seeing the impact of technology and data science in the design of new facilities, but it is hard to get feedback from in-service records to validate the performance of existing plant. To do this we need to create a digital thread that starts in design. We then need to understand the system's operating profile, what maintenance was performed and when, and then its eventual retirement. All of this data is currently collected but it is locked away in different systems, often with different users and usually in the form of semi-structured or unstructured text. Each group uses its own terminology and taxonomy. Data on the same object exists in many places. Finally, data does not flow back to those that need it such as designers to inform life extension decisions.

Figure 1:
Examples of tools and processes for semantic interoperability



There are two steps in creating this digital thread. The first step, as shown in Figure 1, is mapping existing data to machine-readable, semantic data. The second step is to make this machine-readable data interoperable and amenable to queries and reasoning. We call this overall process Technical Language Processing.

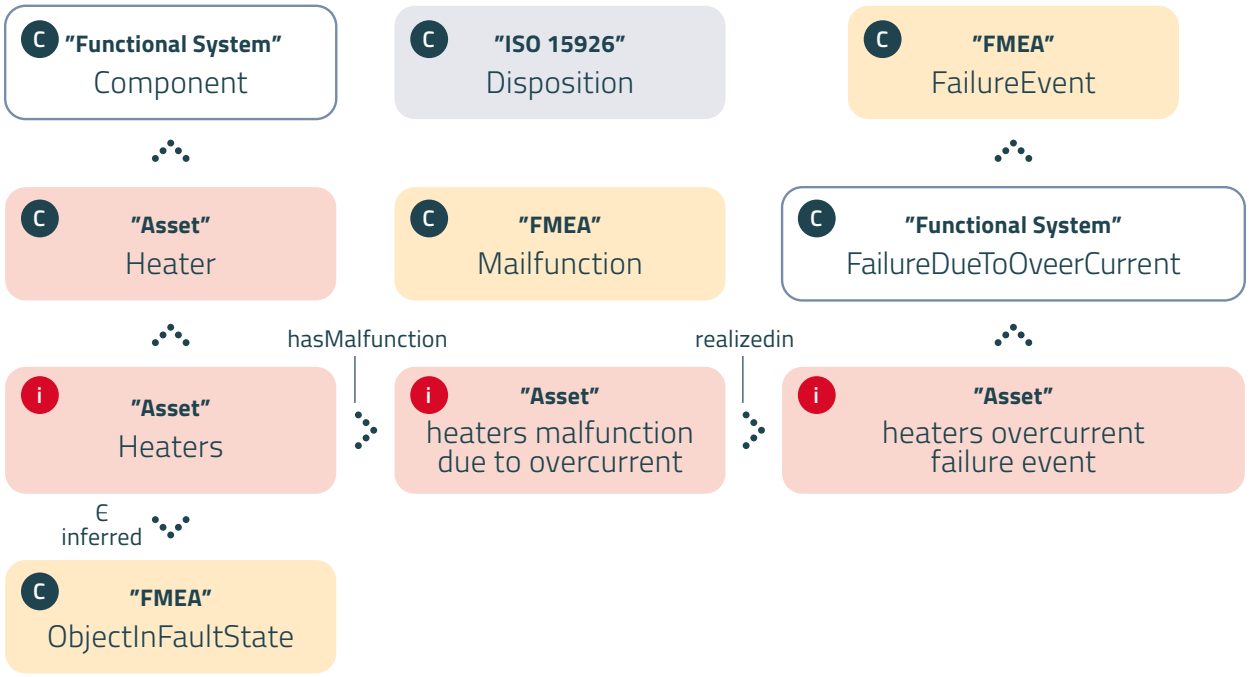
SIRIUS, with its expertise in ontologies and ontology patterns, is at the forefront of work to turn this unstructured data into semantically interoperable knowledge. We are seeing an uptick in interest in ontologies from industrial companies. Early in 2019, DNV GL, the Industrial Ontology Foundry (IOF) and SIRIUS hosted the International Industrial Ontologies Workshop. This event attracted 140 delegates from 43 companies, 24 research institutions and 18 countries. Many attendees went on to join the IOF and contribute to its ongoing work of co-creating a set of open, reference ontologies for manufacturing

and engineering. Early in 2020 we expect to see a new ISO/IEC Standard 21838-1 Requirements for Top-level ontologies. Associated standards for specific ontologies (e.g. BFO and DOLCE) are mature. Members of the IOF have collaborated on developing concepts and first order logic definitions, aligned to the BFO ontology, for key terms used in manufacturing, production, logistics and maintenance. These are planned to be available in March 2020.

We are working with industry on developing use cases and building skills. Johan Klüwer, Martin G. Skjæveland, Daniel Lupp and Leif Harald Karlsen have been collaborating with the Siri-for-Maintenance team at the University of Western Australia. Together, they have developed an ontology that allows failure mode codes in maintenance work orders to be checked against failure modes in a failure modes and effects analysis (FMEA). This tool identifies if failures occurred that were not envisaged by designers or those responsible for maintenance strategy. Doing this manually is a laborious and time-consuming for engineers. A key part of the end-to-end process is being able to ingest the data, often from Excel spreadsheets, into the ontology. This is done using the Reasonable Ontology Templates (OTTR) developed by Martin and his team at SIRIUS. A workshop demonstrating these developments was presented at the prestigious ISWC conference in New Zealand in October 2019.

One of the major challenges for Technical Language Processing is access to industry data. Industrial companies have generated a vast corpus of semi-structured and unstructured data, for example, equipment manuals, instructions and procedures, corporate memos, inspection and safety records, failure investigations, and operational reports. If this corporate data were available and connectable, semantic web techniques could be applied. Confidentiality requirements mean that the data is guarded by the groups that generate and own it, protected by copyright and hidden behind firewalls in corporate systems. This means that there are very few openly available sets of training data based on corporate engineering data. We need this data to build named entity recognition models, support the development of relevant dictionaries and fine-tuned word embeddings. During her visit to Oslo, Professor Hodkiewicz made a presentation to SIRIUS partners about the Data Risk Assessment Tool (DRAT). This freely available web tool allows corporates to risk assess data sets and select controls in a transparent and repeatable way. It is now being used by several resources companies in Australia to manage data release for research partners such as the ARC Centre for Transforming Maintenance through Data Science.

Figure 2:
Fragment of an ontology for Failure Modes and Effects analysis



The UWA Siri-for-Maintenance team and SIRIUS are now working on the ontological and NLP representation of concepts associated with the FMEA risk assessment process. This involves extracting data from Excel tables in a way that maintains and captures the knowledge in the table structure and its engineering implications.

SIRIUS, with its expertise in ontologies and ontology patterns, is at the forefront of work to turn unstructured data into semantically interoperable knowledge.


Melinda Hodkiewicz

Professor **Melinda Hodkiewicz** is instrumental in developing and implementing best practice methods in asset management nationally and internationally. She chaired the Standards Australia committee responsible for the ISO 55001 Asset Management Standard and in 2016 received the MESA Medal, a lifetime achievement award for services to the asset management community. In 2015, Professor Hodkiewicz became the BHP Billiton Fellow for Engineering for Remote Operations at UWA. In 2016, she joined Australia's National Offshore Petroleum Safety and Environmental Management Authority Advisory Board (NOPSEMA). In 2018, Professor Hodkiewicz was awarded a Visiting Fellowship at the Alan Turing Institute in London with the Data Centric Engineering group for the "Siri for Maintenance" project and contributed to the successful Australian Research Council and industry-funded bid to establish a new A$8.8M Centre for Transforming Maintenance through Data Science. In 2019 she was made a Fellow of the Australian Academy of Science and Engineering.

Fabricio Rodrigues

Project
Integrating Ontological Primitives for Analysis and Modelling of Events
Fabrício Henrique Rodrigues is a PhD student from UFRGS (Universidade Federal do Rio Grande do Sul - Porto Alegre - Brazil), working on a research group focused on knowledge representation with applications in the oil industry.

His research concerns the ontology-based conceptual modelling of events. He is staying at UiO for a year, collaborating with the GeoAssistant team in the conceptual modelling of geological processes related to the formation of petroleum reservoirs.

Collaboration in Brazil

SIRIUS' activity in Brazil built on the work in 2018. Once again, SIRIUS had a key role in SINOS, Norway's common academic engagement with the energy sector in Brazil. David Cameron led the program committee and managed the project funding from the Norwegian Ministry of Petroleum and Energy. In 2019, the conference was hosted by Petrobras at their Petrobras University site in Rio de Janeiro. The conference's theme was energy transformations and, for the first time, brought together sessions on petroleum, renewables and industrial social studies. The conference was the best attended ever, with 200 registered participants, 66 from Norway. Plenary addresses were given by senior managers in Petrobras, Equinor and Aker Solutions. Morten Dæhlen, the Dean of the Faculty of Mathematics and Natural Sciences held a plenary address on digitalisation and data science. We collaborated with UiO Energy and its director, Vebjørn Bakken, to organize sessions on solar energy and paths towards carbon neutrality. Through this conference we have established a collaboration with the TIK Centre for Technology, Innovation and Culture. A doctoral student from TIK will be working alongside our planned activities in Brazil. Thomas Østerlie, the leader of our Digital Industrial Transformation program, held a talk in the conference's session on transformation of the oil & gas supply chain.

Our collaboration with the Federal University of Rio Grande do Sul (UFRGS) has broadened and deepened. Unfortunately, it was difficult to recruit students for the exchange positions for which we received funding from DIKU in 2017. However, we have exchanged doctoral students, and a master's student from the Department of Geosciences spent three months in Porto Alegre in the autumn of 2019. Vinicius Gracioli was worked on his PhD project from in 2018 and 2018 and Fabricio Rodrigues has been working here from September 2019. Farhad Nooralahzadeh spent six months in Porto Alegre in 2018 and early 2019.

In the 2018 report we noted that we submitted one of five proposals received in the autumn 2018 joint FINEP-PETROMAKS call. This proposal was for a project called PeTWIN: Whole-field digital twins for production optimization and management. PeTWIN aim is to define a research-based vision and best practices for implementing sustainable, usable and maintainable digital twins for field management. The work addresses the substantial IT challenges posed by digital twins through the application and development of knowledge representation techniques and data analytics. The focus of the project is to build the knowledge and competence needed to build the next generation of digital twins for field digitalisation. This project was awarded in mid-2018 and we plan to start work

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 nine 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.

Universidade Petrobras was the venue for the 2019 November Conference.
David Cameron leads a plenary discussion with Morten Dæhlen, Dean of Mathematics and Natural Sciences and senior managers from Petrobras, Equinor, Aker Solutions and NTNU.

in early 2020. This project is an ambitious collaboration between the University of Oslo, UFRGS, Equinor, Shell and Petrobras. Our aim is to use this project as a focal point for all activities in the digital twins beacon program.

European Collaboration

The centre is an active participant in the European digitalisation 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 further in 2020 in a role that seeks to bridge the common interests between BDVA and A.SPIRE.

The MELODIC Horizon 2020 Project (ICT-731664 https://melodic.cloud) ended with a successful final review at the end of 2019. This project to develop a multi-cloud management platform with the same name. The MELODIC platform enables data-intensive applications to run within defined security, cost, and performance boundaries seamlessly on geographically distributed and federated cloud infrastructures.

Recruitment, Education and Equal Opportunity

Theses defended in 2019

Rule-Based Stream Reasoning

Alessandro Ronca

Alessandro Ronca

In recent years, there has been an increasing interest in extending stream processing engines with rule-based temporal reasoning capabilities. To ensure correctness, such systems must be able to output results over the partial data received so far as if the entire (infinite) stream had been available; furthermore, these results must be streamed out as soon as the relevant data is received, thus incurring the minimum possible latency; finally, due to memory limitations, systems can only keep a limited history of previous facts in memory to perform further computations. These requirements pose significant theoretical and practical challenges since temporal rules can derive new information and propagate it both towards past and future time points; as a result, streamed answers can depend on data that has not yet been received, as well as on data that arrived far in the past. Towards developing a solid foundation for practical rule-based stream reasoning, we propose and study in this thesis a suite of decision problems that can be exploited by stream reasoning algorithms to tackle the afore-mentioned challenges, and provide tight complexity bounds for a core temporal extension of Datalog. All of the problems we consider can be solved at design time (under reasonable assumptions), prior to the processing of any data. Solving these problems enables the use of reasoning algorithms that process the input streams incrementally using a sliding window, while at the same time supporting an expressive rule-based knowledge representation language and minimising both latency and memory consumption.

Higher-Level View of Ontological Modeling

Daniel Lupp

Daniel Lupp

Daniel Lupp successfully defended his PhD Thesis *A Higher-Level View of Ontological Modeling: Rule-based approaches for data transformation, modeling, and maintenance* 7 May 2019.

Constructing and maintaining a high-quality ontology are by no means simple tasks; indeed, many of the specification and maintenance tasks are currently performed manually and thus prone to human error. This work provides interfaces to ontologies and ontology-based data access systems that allow for users to see the big picture rather than needing to understand complex logical constructions. It provides a solid theoretical basis for sophisticated ontology design and maintenance, supporting the encoding of design choices that previously were left implicit as well as algorithms for the detection and removal of unwanted redundancies. It provides a data transformation framework that supports automatic, robust handling of exceptions and incomplete data, thus cutting down on manual maintenance tasks. Furthermore, a new formalism for creating ontologies (called OTTR) which supports the definition and instantiation of recurring patterns is introduced. This formalism and its extensions support creating ontologies quickly and in bulk while simultaneously ensuring uniformity of modelling following the Don't-Repeat-Yourself (DRY) principle.

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 & 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 2021.

Mentor program kick-off October 2019.

DataScience@UiO

In 2017 SIRIUS, together with the BigInsight SFI, applied to the Faculty of 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.

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 Petrobas 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. In 2019 a student from the Institute of Geosciences completed his thesis in Porto Alegre.

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 & 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, *Forskerlinjen I*,

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/). This pilot will be continued for a third year, (the academic year 2020/2021), and will also offer an additional program, Forskerlinjen II, which is a master program available for those who have completed Forskerlinjen I.

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 centre identity by exchanging expertise, value, 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 and was a great success with both the mentors and the mentees experiencing it as both rewarding and useful.

The program brings together SIRIUS PhD students and junior researchers with mentors from our partner companies. The program focuses on building a common centre identity, by forming a shared understanding between the industrial and academic partners; this creates a foundation for further innovation in SIRIUS. It also serves as a common arena that contributes to career development and opportunities for the participants beyond SIRIUS. This is done through regular meetings between mentor/mentee pairs, seminars and business visits for the mentees.

The SIRIUS Mentoring Program started up again in 2019 with ten new mentees from SIRIUS and seven new mentors and three mentors participating for the second time. The kick-off of the program was held on the 21st and 22nd October. The program will continue for a year and will have the closing seminar in September 2020. Ingrid Chieh Yu is the leader of the program and has developed this program together with AFF who has 65 years of experience in mentoring, coaching and leadership development.

The mentor program brings together SIRIUS PhD students and junior researchers with mentors from our partner companies. The program focuses on building a common centre identity, by forming a shared understanding between the industrial and academic partners; this creates a foundation for further innovation in SIRIUS.

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 coordinator for collaboration environments, then as a Site Manger / 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.

Governance

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 2019: one in May, hosted by Schlumberger in Asker and the second one in November hosted by Computas at their new offices in the centre of Oslo. The practice of our partners hosting the General Assembly meetings generates closer contact between the centre and the hosting partner.

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 2020.

Executive Committee of the General Assembly

The Executive Committee of the General Assembly held regular meetings in 2019. The executive committee acts on behalf of the General Assembly to exercise regular oversight over plans and progress for the centre.

Knut Sebastian Tungland

Knut Sebastian Tungland is a Senior Advisor in Equinor. He works within the digitalisation unit of the CIO office and intersection between software and the energy industry, with software products, as a consultant and with Equinor.

This Committee in 2019 consisted of:

- Knut Sebastian Tungland, Equinor, Chairman.
- Jostein Fonneland, Schlumberger, with Per Eivind Solum, Schlumberger, as deputy.
- Frode Myren and Øystein Haaland, IBM, with Nicholas Peels, then Ivan Mudron, OsiSoft, as deputy.
- Einar Rustad, Numascale, with Hugo Kohmann, Dolphin Interconnect Solutions, as deputy.
- Arild Waaler, University of Oslo, with Ingrid Chieh Yu, University of Oslo, as deputy.

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 2019 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

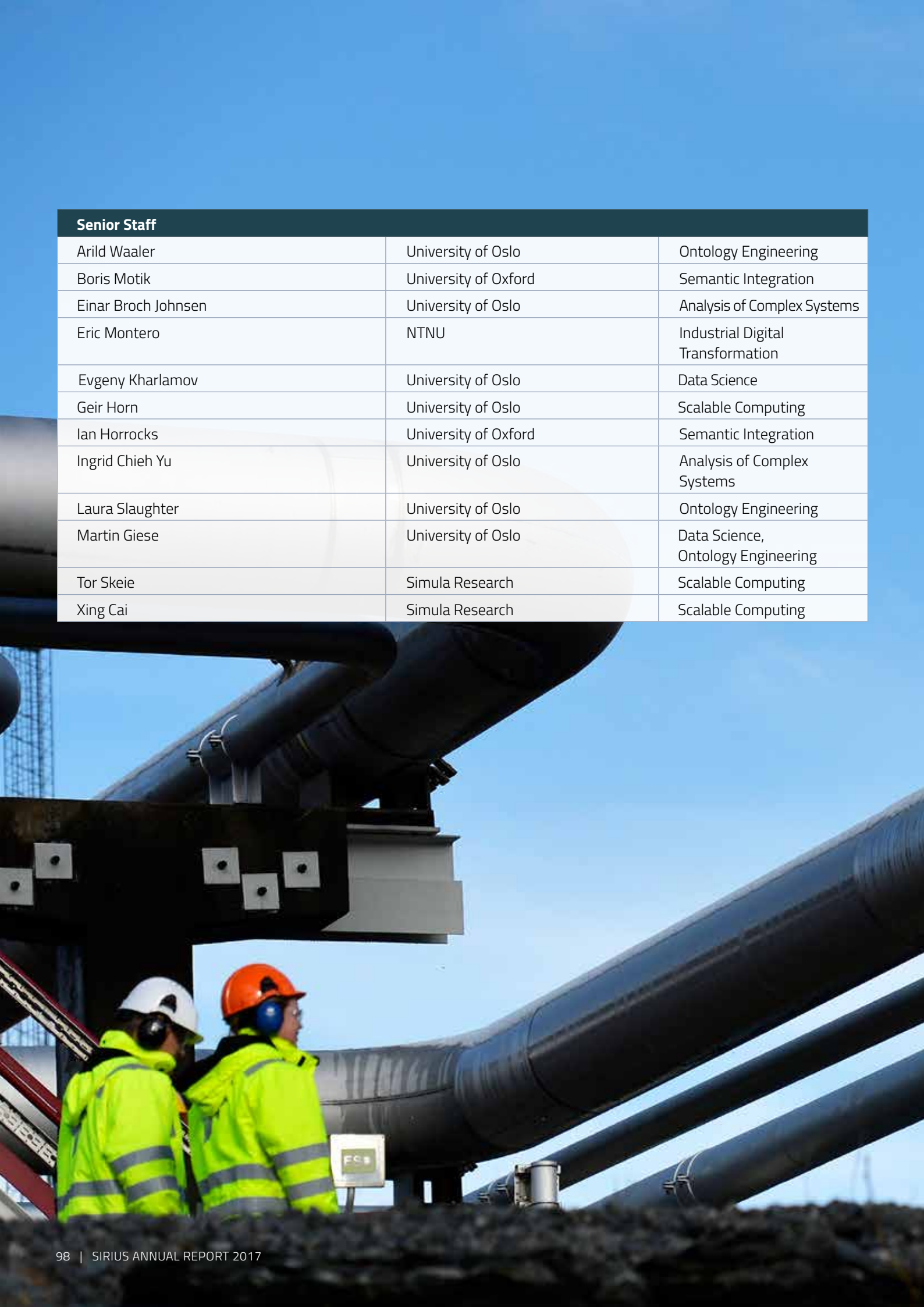
The Operations Board (or Centre Management team) is responsible for day-to-day operation of the centre and the definition of work plan that implements the decisions of the Strategy Board. It is chaired by the Operations Manager and consists of the Centre Leader, Administration Manager, Mentor and Education Coordinator, Work Package Leaders and Research Program Leaders.

Hammerfest LNG. Photo: Harald Pettersen, Statoil

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	Silvia Lizeth Tapia Tarifa	University of Oslo
Ontology Engineering	Martin G. Skjæveland	University of Oslo
Data Science	Basil Ell	University of Oslo
Semantic Integration	Dag Hovland	University of Oslo
Scalable Computing	Geir Horn	University of Oslo
Industrial Digital Transformation	Thomas Østerlie	NTNU

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



Research, Technical and Administrative Personnel		
Adnan Latif	University of Oslo	Subsurface Work Package
Basil Ell	University of Oslo/Universität Bielefeld	Data Science
Chi Mai Nguyen	University of Oslo	Analysis of Complex Systems
Crystal Chang Din	University of Oslo	Analysis of Complex Systems
Dag Hovland	University of Oslo	Semantic Integration
Daniel Lupp	University of Oslo	Ontology Engineering
David Cameron	University of Oslo	Centre Coordinator
David Tena Cucala	University of Oxford	Semantic Integration
Elena Parmiggiani	NTNU	Industrial Digital Transformation
Ernesto Jimenez-Ruiz	University of Oslo	Data Science
Inger-Lise Reang	University of Oslo	Administration Manager
Irina Pene	University of Oslo	Ontology Engineering
Jan Tore Lønning	University of Oslo	Data Science
Jens Otten	University of Oslo	Semantic Integration
Jiaoyan Chen	University of Oxford	Data Science
Jieying Chen	University of Oslo	Ontology Engineering
Keith Lewis	University of Oslo	
Leif Harald Karlsen	University of Oslo	Ontology Engineering
Martin G. Skjæveland	University of Oslo	Ontology Engineering
Michael Heeremans	University of Oslo	
Oliver Stahl	University of Oslo	SIRIUS Laboratory
Rudolf Schlatte	University of Oslo	Analysis of Complex Systems
Silvia Lizeth Tapia Tarifa	University of Oslo	Analysis of Complex Systems
Shuwen Liu	University of Oxford	Semantic Integration
Thomas Østerlie	NTNU	Industrial Digital Transformation
Violet Ka I Pun	University of Oslo	Analysis of Complex Systems

Kårstø. Photo: Harald Pettersen, Statoil

Publications in 2019

Ahrendt, Wolfgang; Tapia Tarifa, Silvia Lizeth. *Integrated Formal Methods* - 15th International Conference, IFM 2019, Bergen, Norway, December 2-6, 2019, Proceedings. Springer 2019 (ISBN 978-3-030-34967-7) 551 s. Lecture Notes in Computer Science (LNCS)(11918) UiO.

Ajileye, Temitope, Boris Motik, and Ian Horrocks. 2019. *Datalog Materialisation in DistributedRDF Stores with Dynamic Data Exchange*. In International Semantic Web Conference, 21–37.

Bezirgiannis, Nikolaos; de Boer, Frank; Johnsen, Einar Broch; **Pun, Ka I; Tapia Tarifa, Silvia Lizeth**. *Implementing SOS with Active Objects: A Case Study of a Multicore Memory System*. Lecture Notes in Computer Science (LNCS) 2019 ;Volum 11424. s.332-350 HVL UiO.

Bijo, Shiji; Johnsen, Einar Broch; Pun, Ka I; Tapia Tarifa, Silvia Lizeth. *A formal model of data access for multicore architectures with multilevel caches*. Science of Computer Programming 2019 ;Volum 179. s.24-53 HVL UiO.

Chen, Jiaoyan, Ernesto Jiménez-Ruiz, and Ian Horrocks. 2019. *Canonicalizing Knowledge Base Literals*. In International Semantic Web Conference, 110–27.

Chen, Jiaoyan, Ernesto Jiménez-Ruiz, Ian Horrocks, and Charles A. Sutton. 2019. *Learning Semantic Annotations for Tabular Data*. In Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, 2088–94.

Chen, Jiaoyan, Ernesto Jimenez-Ruiz, Ian Horrocks, and Charles Sutton. 2019. *ColNet: Embedding the Semantics of Web Tables for Column Type Predictio*n. AAAI 2019 : Thirty-Third AAAI Conference on Artificial Intelligence 33 (1): 29–36.

Cucala, David Tena, Bernardo Cuenca Grau, and Ian Horrocks. 2019. *15 Years of Consequence-Based Reasoning*. Description Logic, Theory Combination, and All That, 573–87.

de Boer, Frank; Giachino, Elena; de Gouw, Stijn; Hähnle, Reiner; **Johnsen, Einar Broch; Laneve, Cosimo; Pun, Ka I; Zavattaro, Gianluigi**. *Analysis of SLA Compliance in the Cloud -- An Automated, Model-based Approach*v. Electronic Proceedings in Theoretical Computer Science 2019 ;Volum 302. s.1-15 HVL UiO.

Fernandez-Reyes, Kiko; Clarke, Dave; Henrio, Ludovic; Johnsen, Einar Broch; Wrigstad, Tobias. *Godot: All the benefits of implicit and explicit futures*. Leibniz International Proceedings in Informatics 2019 ;Volum 134. UiO.

Gkolfi, Anastasia; Din, Crystal Chang; Johnsen, Einar Broch; Kristensen, Lars Michael; Steffen, Martin; Yu, Ingrid Chieh. *Translating active objects into colored Petri nets for communication analysis*. Science of Computer Programming 2019 ;Volum 181. s.1-26 HVL UiO.

Harrow, Ian; Balakrishnan, Rama; Jiménez-Ruiz, Ernesto; Jupp, Simon; Lomax, Jane; Reed, Jane; Romacker, Martin; Senger; Christian; Splendiani, Andrea; Wilson, Jabe; Woollard, Peter. *Ontology mapping for semantically enabled applications*. Drug discovery today (2019).

Holter, Ole Magnus; Myklebust, Erik Bryhn; Chen, Jiaoyan; Jiménez-Ruiz, Ernesto. *Embedding OWL ontologies with OWL2Vec*. CEUR Workshop Proceedings. Vol. 2456. 2019.

Hu, Pan, Boris Motik, and Ian Horrocks. 2019. *Modular Materialisation of Datalog Programs*. AAAI 2019 : Thirty-Third AAAI Conference on Artificial Intelligence 33 (1): 2859–66.

Hu, Pan, Jacopo Urbani, Boris Motik, and Ian Horrocks. 2019. *Datalog Reasoning over Compressed RDF Knowledge Bases*. In Proceedings of the 28th ACM International Conference on Information and Knowledge Management, 2065–68.

Kamburjan, Eduard; Din, Crystal Chang; Hahnle, Reiner; **Johnsen, Einar Broch**. *Asynchronous Cooperative Contracts for Cooperative Scheduling*. Lecture Notes in Computer Science (LNCS) 2019, Volume 11724 LNCS, 4866.

Kharlamov, Evgeny; Kotidis, Yannis; Mailis, Theofilos; Neuenstadt, Christian; Nikolaou, Charalampos; Özcep, Özgür; Svingos, Christoforos; Zheleznyakov, Dmitriy; Ioannidis, Yannis; Lamparter, Steffen; Möller, Ralf; Waaler, Arild. *An ontology-mediated analytics-aware approach to support monitoring and diagnostics of static and streaming data*. Journal of Web Semantics 2019 ;Volum 56. s.30-55 UiO.

Klungre, Vidar N; Soylu, Ahmet; Jiménez-Ruiz, Ernesto; Kharlamov, Evgeny; Giese, Martin. *Query Extension Suggestions for Visual Query Systems through Ontology Projection and Indexing*. New Generation Computing 37.4 (2019): 361-392.

Li, Huanyu; Dragisic, Zlatan; Faria, Daniel; Ivanova, Valentina; Jiménez-Ruiz, Ernesto; Lambrix, Patrick; Pesquita, Catia. *User validation in ontology alignment: functional assessment and impact*. The Knowledge Engineering Review 34 (2019).

Monteiro, Eric; Parmiggiani, Elena. *Synthetic Knowing: The Politics of the Internet of Things*. MIS Quarterly 2019 ;Volum 43.(1) s.167-184 NTNU.

Nooralahzadeh, Farhad; Lønning, Jan Tore; Øvrelid, Lilja. *Reinforcement-based denoising of distantly supervised NER with partial annotation*. I: Proceedings of the 2nd Workshop on Deep Learning Approaches for Low-Resource NLP (DeepLo 2019). Association for Computational Linguistics 2019 ISBN 978-1-950737-789. s.225-234 UiO.

Parmiggiani, Elena; Monteiro, Eric. *Digitized Coral Reefs. I: Digitalsts*: A Field Guide for Science & Technology Studies. Princeton University Press 2019 ISBN 978-0-691-18707-5. s.300-325 NTNU.

Rasouli, Peyman; Yu, Ingrid Chieh. *Meaningful Data Sampling for a Faithful Local Explanation Method*. Lecture Notes in Computer Science (LNCS) 2019 s.28-38 UiO.

Sanchez-Graillet, Olivia; Cimiano, Philipp; Witte, Christian; Ell, Basil. *C-TrO: An Ontology for Summarization and Aggregation of the level of Evidence in Clinical Trials*. Proceedings of the Workshop Ontologies and Data in Life Sciences (ODLS 2019).

Vega-Gorgojo, Guillermo; Slaughter, Laura; Zernichow, Bjørn Marius von; Nikolov, Nikolay; Roman, Dumitru. *Linked Data Exploration With RDF Surveyor*. IEEE Access 2019 ;Volum 7. s.172199-172213 SINTEF UiO.

Wałęga, Przemysław A, Mark Kaminski, and Bernardo Cuenca Grau. 2019. *Reasoning over Streaming Data in Metric Temporal Datalog*. AAAI 2019 : Thirty-Third AAAI Conference on Artificial Intelligence 33 (1): 3092–99.

Presentations in 2019

Cameron, David B.. *Digital Twin Research* – University of Oslo. FutureDistributed.org [Podcast] 2019-11-11

Cameron, David B.. *SIRIUS Digital Twins, Semantics and Petrobras*. 2nd Semantic Technology Meeting; 2019-11-15 UiO

Cameron, David B.. *SIRIUS: Innovation-driven research? Research-driven innovation? Or both together?*. Bygge bro over teknologiens Valley of Death; Tekna seminar, Oslo.

Cameron, David B.; Waaler, Arild; Abel, Mara. *Digital Twins as a Platform For Artificial Intelligence in the Petroleum Supply Chain*. 1 Seminário Inteligência Artificial Aplicada à Indústria do Petróleo; 2019-11-06 - 2019-11-08

Cameron, David B.; Waaler, Arild; Skjæveland, Martin G; Gjerver, Anders; Hansen, Christian Mahesh. *The Whole Plant Digital Twin: What Can Semantic Technologies Contribute?*. Advances in Process Automation and Control 2019; 2019-11-18 2019-11-20

Cameron, David B.; Waaler, Arild; Tungland, Knut Sebastian; Nøst, Elisabeth. *Using the Industry Collaboration Canvas in the Mid-Term Review*. Experience in applying the framework in the oil and gas industry. 2019 University-Industry Interaction Conference; 2019-06-18 - 2019-06-20

Din, Crystal Chang; Karlsen, Leif Harald; Pene, Irina; Stahl, Oliver; Yu, Ingrid Chieh; Østerlie, Thomas. *Geological Multi-scenario Reasoning*. NIK: Norsk Informatikkonferanse 2019 NTNU UiO.

Jiménez-Ruiz, Ernesto; Hassanzadeh, Oktie; Efthymiou, Vasilis; Chen, Jiaoyan; Srinivas, Kavitha. *SemTab 2019: Resources to Benchmark Tabular Data to Knowledge Graph Matching Systems*. ESWC 2020.

Kharlamov, Evgeny. *Knowledge Graphs and Semantic Technologies for Industry* 4.0.. JIST2019: The 9th Joint International Semantic Technology Conference; 2019-11-25 - 2019-11-27 UiO.

Lécué, Freddy; Chen, Jiaoyan; Pan, Jeff Z.; Chen, Huajun. *Augmenting Transfer Learning with Semantic Reasoning*. IJCAI 2019.

Mumtaz, Summaya; Giese, Martin. *Frequency-Based vs. Knowledge-Based Similarity Measures for Categorical Data*. AAAI-Make2020.

Myklebust, Erik Bryhn; Jiménez-Ruiz, Ernesto; Chen, Jiaoyan; Wolf, Raoul; Tollefsen, Knut Erik. *Knowledge Graph Embedding for Ecotoxicological Effect Prediction*. ISWC 2019.

Skjæveland, Martin G; Karlsen, Leif Harald; Lupp, Daniel P.. I*OF Workshops 2019 Tutorial: Ontology patterns in practice*. International Industrial Ontologies Foundry Workshop; 2019-02-04 UiO.

Skjæveland, Martin G; Lupp, Daniel P.. *Practical and Scalable Pattern-based Ontology Engineering with Reasonable Ontology Templates (OTTR)* - Half-day tutorial. Extended Semantic Web Conference 2019; 2019-06-07 - 2019-06-07 UiO.

Skjæveland, Martin G; Lupp, Daniel P.; Karlsen, Leif Harald; Hodkiewicz, M. R.. *Scalable Construction of Sustainable Knowledge Bases* - Full day tutorial. International Semantic Web Conference 2019; 2019-10-26 UiO.

Thune, Andreas. *Mesh partitioning in the presence of strong coefficient heterogeneity*. OPM Summit 2019; 2019-01-24 - 2019-01-25.

Thune, Andreas; Cai, Xing. *Balancing the numerical and parallel performance for reservoir simulations*. SIAM Conference on Computational Science and Engineering (CSE19); 2019-02-25 - 2019-03-01 UiO.

Glossary

ABS	A simulation and analysis tool for complex computer and organizational systems.
AI	Artificial Intelligence
CDSS	Clinical Decision Support System
EHR	Electronic Health Records
EMA	The Execution, Modelling and Analysis strand in SIRIUS.
FMEA	Failure Modes and Effects Analysis
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
MMD	Material Master Data
NIVA	Norwegian Institute for Water Research
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.
OSDU	Open Subsurface Data Universe
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.
SCD	Scope-Condition-Demand: pattern used to define rules for requirements so the requirements can be represented in a structured, machine readable form.
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.
XAI	Explainable Medical Artificial Intelligence

Dvalin M40 load-out. Photo: Øyvind Sætre, Aibel


aibel
6507/7
HEIDRUN
EQUINOR

Annual Accounts

Costs

All figures in 1000 NOK	2015	2016	2017	2018	2019
Personnel and indirect costs	539	5188	10087	12378	20922
Purchase of research services	-	600	2113	3415	6166
Equipment	-	31	122		
Other operational costs	62	9505	10424	14131	17202
Total Sum	**601**	**15324**	**22746**	**29924**	**44290**

Funding

All figures in 1000 NOK	2015	2016	2017	2018	2019
Research Council	-	4168	8180	12430	16961
University of Oslo	601	2141	5593	6387	7721
Public partners	-	144	813	107	1137
Private partners	-	8510	7910	11000	17700
International partners		361	250		771
Total Sum	**601**	**15324**	**22746**	**29924**	**44290**

Photo credits
Aibel
Davi Costa
iStock
Shutterstock
Statoil/Equinor
Unsplash

Editor:
Bente Halsnes

Layout
FÆRD AS

SIRIUS

ANNUAL REPORT 2019

Sirius-labs.no