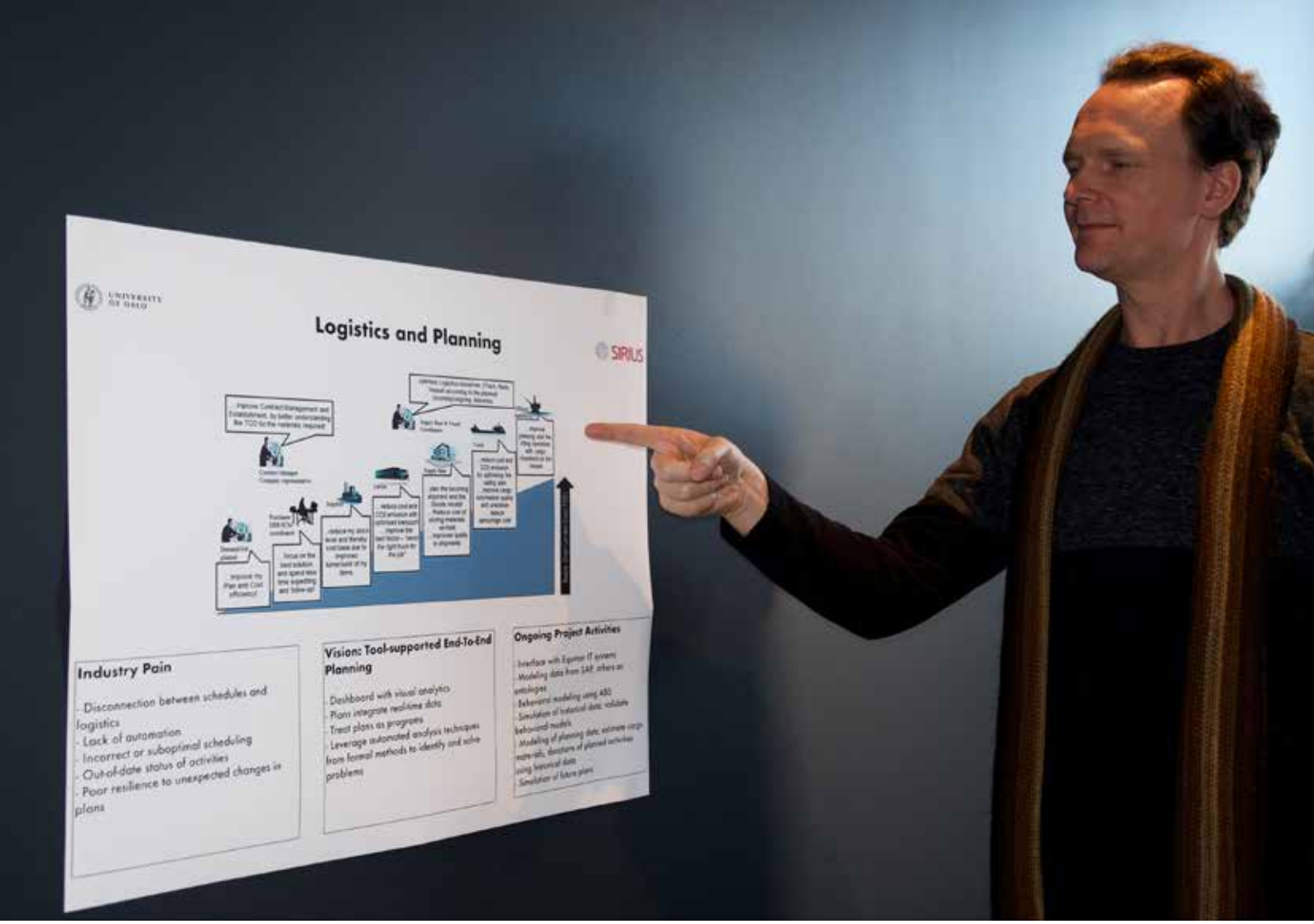
SIRIUS ANNUAL REPORT 2020

Centre for Research-based Innovation
The Research Council of Norway

UiO : University of Oslo

Contents


Logistics and Planning
SIRIUS
Industry Pain
- Disconnection between schedules and logistics
- Lack of automation
- Incorrect or suboptimal scheduling
- Out-of-date status of activities
- Poor resilience to unexpected changes in plans
Vision: Tool-supported End-To-End Planning
- Dashboard with visual analytics
- Plans integrate real-time data
- Treat plans as programs
- Leverage automated analysis techniques from formal methods to identify and solve problems
Ongoing Project Activities
- Interface with Equinor IT systems
- Modeling data from SAP, others as ontologies
- Behavioral modeling using ABS
- Simulation of industrial data: validate behavioral models
- Modeling of planning data: estimate cargo materials, durations of planned activities using historical data
- Simulation of future plans

We realized the need to strengthen our research programs and started to take actions to formulate robust programs.

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

Executive Summary

Global society faced a significant challenge in 2020. The COVID pandemic had a negative impact on many businesses. Academia and University Research also faced its challenges, particularly when home working and online tuition became the norm.

With the support of our researchers and centre partners we took on the challenge to adapt to the new ways of working. This resulted in a continuation of research activities. However, even though a significant effort was made to maintain a sense of community, there was the inevitable impact on the centre's ability to work efficiently.

Faced with these challenges, SIRIUS continued to perform. During the year, we closed the Geological Assistant project with the delivery of a prototype, and we completed the research within the BIGMED Personalized Healthcare project. Most importantly, we realized the need to strengthen our research programs and started to take actions to formulate more robust programs.

The continued participation of SIRIUS researchers in Joint Industry Projects and SIRIUS research demonstration projects (such as the Geological Assistant) has not only helped to build capability, but also highlighted areas where we needed to strengthen our research.

The READI Joint Industry Project developed a new and promising tool for asset information modelling. This Information Modelling Framework provides a standard way to designate an item in an engineering design, the project also developed a Reference Designation System for oil and gas facilities based on the ISO/IEC 81346 standard. Complementary to this, the KONKRAFT digital design basis project developed digital representation of the starting point of any field development project. Asset Information modelling is an area of research that will be strengthened within SIRIUS during 2021.

The Reasonable Ontology Templates project continued its development in 2020, improving the usability of its framework and the support of its pilot users in our partner companies. The Reasonable Ontology Templates project is already set up to deliver technology and methods to asset modelling in field development projects through 2021 and beyond.


Arild Waaler

The Analysis of Complex Systems research program has delivered results for lightweight systematic techniques for model exploration, techniques combining executable models with semantic technologies, analysis techniques for active object languages, and analysis techniques for feature models. These core capabilities have the potential for demonstration in the remaining years of the SIRIUS centre. We see a significant opportunity to demonstrate our complex systems research in operations and logistics planning. In combination with asset modelling techniques and ontology engineering our complex systems research has great potential to support the PeTWIN Project and potentially also the first subsurface application of asset modelling with the Open Group OSDU forum.

After the conclusion of contractual agreements, we are excited to have started the PeTWIN collaborative project on digital twins for field management, together with Equinor, Shell, Petrobras and the Federal University of Rio Grand do Sul. This project is financed by the Research Council of Norway and FINEP, the Brazilian federal innovation organization.

We were also honoured to be participants in the award

of two European H2020 projects, ONTOCOMMONS and Eur3ka. The ONTOCOMMONS project is a Coordination and Support Action, which has the aim of building up a shared semantic foundation for applications in manufacturing and materials science. Eur3ka (EUropean Vital Medical Supplies and Equipment Resilient and Reliable Repurposing Manufacturing as a Service NetworK for Fast Pandemic Reaction) is financed under the European Union's pandemic response program. It will build and demonstrate an IT framework that will enable manufacturing companies to repurpose production quickly and effectively.

Following the positive experience of our participation in the Open Group OSDU Forum, we expanded our global presence by winning funding to establish the DSYNE INTPART network. This network will build competence and courses in the field of digital requirements in engineering. The network brings together leading groups in System Engineering and Computer Science from Norway, Brazil and the United States.

SIRIUS researchers have in 2020 been very active in the first round of the Research Council's competition for new Centres of Excellence with two submitted proposals. One of them, led by Einar Broch Johnsen, has its basis in the Analysis of Complex Systems research program with contributions from other programs. The other is an initiative of our collaborators in the Big Insight SFI with significant contributions from our Domain-Adapted Data Science program. Whether or not these two initiatives are in the end funded, the proposal work in itself contributes massively to the international culture and ambition of our SIRIUS researchers.

Reflecting on the wider societal challenges of 2020 and the resulting constraints on close physical presence, I feel humbled by the resilience of our researchers and wish to extend my sincere appreciation for the continued support of our partners and the Norwegian Research Council.

SIRIUS is running until the end of 2023 with continued focus on strengthening the research programs, increasing our capabilities and ambition. In a challenging year, I believe that together, we have built a foundation to take the centre forward.

In 2020 several important technical developments occurred.

Highlights of 2020

The Analysis of Complex Systems research program develops and applies techniques that support the representation and analysis of correct behaviours and interactions of complex systems, together with analysis related to time-dependent systems, resource management, constraint solving, optimisation and prediction. In the past year, this research program has delivered results for lightweight systematic techniques for model exploration, techniques combining executable models with semantic technologies, analysis techniques for active object languages, and analysis techniques for feature models.

The Reasonable Ontology Templates (OTTR) project has continued its developments in 2020. While the focus in the beginning of the project has been on the core foundations of the framework, including establishing the formal theoretical foundations, developing a suitable syntax and implementing support for the basic functionality of the framework, the focus in this year has been on improving the usability of the framework. This includes developing a methodology for developing template libraries, developing best practices for template library publishing, publishing core set of templates that should be useful to all ontology engineering projects, and improving the public programming interface of the project's reference implementation. All these developments are available via the project's website: **http://ottr.xyz**.

In 2020 several important technical developments occurred in the Semantic Integration research program: The PAGODa reasoner improved its querying capabilities and the algorithms used by RDFox for distributed triple storage and querying were significantly improved. Also, a modal sequent calculus was developed for use in the READi project. A team was started to focus on Ontology-Based data Access. The team includes members from several research programs, and from the newly joined companies Ontopic and Oxford Semantic Technologies. The team aims at being a gateway between demonstration projects using OBDA into the research programs.

A prototype domain specific language SMOL has been deployed which supports high-level ways to program the orchestration of simulation units which implement the FMI standard. SMOL seamlessly integrates semantic technology with programming constructs for interacting with simulators. SMOL has a formally defined semantics and a prototype runtime implemented in Kotlin, a dialect of Java. SMOL addresses the composition problem for simulators with different domain models through lifting semantic technologies to enable correct configuration and orchestration of connections inside the digital twin. By embracing the semantic technologies, formal methods can contribute to the development of provably correct digital twins.

The start-up of The PeTWIN collaborative project on digital twins for field management started up together with Equinor, Shell, Petrobras and the Federal University of Rio Grand do Sul. This project is financed by the Research Council of Norway and FINEP, the Brazilian federal innovation organization.

The DSYNE INTPART network held its kick-off meeting. This network will build competence and courses in the field of digital requirements in engineering. The network brings together leading groups in System Engineering and Computer Science from Norway, Brazil and the United States. We are also collaborating with industrial end-users, including Shell, Equinor, TechnipFMC and DNV.

Two European H2020 projects, ONTOCOMMONS and Eur3ka, were awarded The ONTOCOMMONS project is a Coordination and Support Activity, which has the aim of building up a shared semantic foundation for applications in manufacturing and materials science. The University of Oslo and SINTEF are the Norwegian partners in this work, and Aibel is participating in an application case study. Eur3ka (EUropean Vital Medical Supplies and Equipment Resilient and Reliable Repurposing Manufacturing as a Service NetworK for Fast PAndemic Reaction) is financed under the European Union's pandemic response program. It will build and demonstrate an IT framework that will enable manufacturing companies to repurpose production quickly and effectively.

SIRIUS provided the project contract and collaboration for the KONKRAFT digital design project. This project involved Equinor, AkerBP, Lundin Energy, Aker Solutions (with Aize and Cognite), Aibel and TechnipFMC.

SIRIUS personnel have worked in the READI Joint Industry Project. This work is coordinated by DNV and has involved other SIRIUS partners: Computas, Equinor, Aker Solutions, Aibel and TechnipFMC. This project has worked on the formulation digital requirements that can replace current document-based requirements and standards. This work is directly relevant to the DSYNE network, and we plan to use DSYNE as a platform to internationalize this work. READI has also developed a new and promising tool for asset information modelling. This Information Modelling Framework provides a standard way to designate an item in an engineering design. READI has developed a Reference Designation System for oil and gas facilities based on the ISO/IEC81346 standard.

Work on GeoAssistant continued in 2020, where the main focus was to extend the methodology developed in 2019, based on a synthetic use case, to a complex real-world case study from the Johan Sverdrup field. Development of this prototype completed in 2020.

The Cross-Domains Applications work package in SIRIUS contains projects that transfer technology between the oil and gas sector and other important economic sectors. In 2020, we completed research within the BIGMED Personalized Healthcare project and are in the finishing stages of a collaborative project with NIVA, the Norwegian Institute for Water Research. Both of these projects involved work within the research areas of semantic integration and ontology engineering.

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 concurrently (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 supercomputers 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 (e.g, time analysis, resource analysis, performance analysis, etc.). An effective way to understand and analyse such systems is by using formal methods, which are logical-based techniques to abstractly represent or model the behaviour and interaction of systems. This program researches 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.

In the past year, our research program has deliver results for: (1) lightweight systematic techniques for model exploration with the aim to reduce the search space of all possible final states that can be reached during execution, (2) techniques combining executable models with semantic technologies to add semantic meaning to the changing state for and during an execution, (3) analysis techniques for active object languages, and (4) analysis techniques for feature models. We now detail some of these contributions.

Partial Order Reduction for Guarded Command languages

Testing whether an object with cooperatively scheduled tasks locally exhibits deterministic behavior (where the final state of the object, after execution, is independent of the local scheduling of tasks) is not a trivial problem. This contribution studies how the cooperatively scheduled tasks of a single object can be abstracted in terms of a guarded command language (GCL) with shared state (a slight simple variation on Dijkstra's original language). **This contribution proposes a method which combines GCL with Partial Order Reduction (POR)**. POR is a technique to reduce the size of the search space when exploring the different executions of a concurrent program by exploiting the commutativity of concurrently executed independent transition steps, which result in the same state when executed in different orders. The proposed method can answer questions such as whether all permutations of an execution trace are equivalent, and therefore pruned via POR techniques. More details about this method can be found in Recent Developments in the Design and Implementation of Programming Languages. Results from this method are currently being integrated into ABS.

ABS is a method and language for modelling, analyzing and simulating distributed timed, resource-aware systems. In addition to supporting modelling of functional behavior and distributed algorithms and systems, ABS 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 datatypes and 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 was designed to be easy to read and use by industrial programmers. It is an open source research project that is used in teaching and research, including industrial innovation research at SIRIUS.

For the end user, the main features and application areas of ABS are:

- **Discrete-Event Simulation of timed, resource-aware systems**
- **Custom visualization of live simulation data**
- **Data export of simulation results into other tools**
- **Model exploration using formal analysis tools**

Online Resources

The ABS homepage is at **https://abs-models.org**; the current language manual can always be found at **https://abs-models.org/manual/**. A categorized list of major publications is at **https://abs-models.org/publications/**.

Rudolf Schlatte

Silvia Lizeth Tapia Ta

ABS is an open source project developed at **https://github.com/abstools/abstools**.
The latest release is accessible at **https://github.com/abstools/abstools/releases/latest**.

Symbolic Execution meets Partial Order Reduction

Symbolic execution is a technique for software analysis. It is especially used for testing, but also for debugging and verification. In the context of executable models, one single symbolic execution may correspond to a large, possibly infinite, class of normal executions with concrete values. Symbolic execution has mainly been applied to sequential languages, where different executions can be captured by different Boolean conditions, when following an execution trace. Concurrent languages give rise to non-determinism during execution (given by different scheduling decisions), which is another source of search space explosion. **To address the state space explosion resulting from the different possible non-deterministic scheduling decisions, This contribution proposes a combination of symbolic execution and POR.** More details about this combination can be found in the book Deductive Software Verification: Future Perspectives.

Combining Rewriting logic with Semantic Technology

Rewriting logic (RL) is a logical framework in which other logics can be represented and a framework for transition systems, in which many different models of concurrency, distributed algorithms, programming languages, and different systems can be naturally represented, executed and analyzed as rewrite theories, which include a set of rewriting rules expressing state transitions. Semantic Technologies covers different techniques to attach semantic meaning to data, as a formal and conceptual description of a relevant domain. **This contribution proposes a framework which combines rewriting rules for geological processes and semantic technology for the geology domain.** More details about this combination can be found in the geological multi-scenario reasoning project at SIRIUS.

Semantic Micro Object Language (SMOL)

This contribution develops a Semantic Micro Object Language (SMOL), which allows the programmer to use established query and data modeling techniques directly in his program, thus benefiting from semantic technology support for debugging, domain modeling and data access. The integration of semantic technologies directly into a programming language is used to ensure the correct usage of semantic meaning in the program through type systems and other tools. Through the clean separation between data modeling and programming, SMOL integrates smoothly with industry-strength frameworks and builds on the expertise already present for existing semantic technologies, such as SPARQL, OWL or RDF. We plan to use SMOL to build orchestrators with semantic meaning for co-simulation in Digital Twins.

Additional contributions

Our research program has also proposed an analysis technique for feature models that enables to create and adapt a feature model evolution plan while preventing undesired impacts on its consistency, and analysis techniques for active object languages to reproduce a specific run for non-deterministic active object systems, to identify restrictions on active object programs which are sufficient to guarantee a deterministic behaviour, to specify behavioral contracts for cooperative scheduling and a formal model of resource consumption and scaling for containerized microservices managed by Kubernetes.

Participation in new projects

Our research program is one of the beneficiaries of the MSCA-ITN REMARO, which aims to deliver expertise in reliable AI for marine robotics, and is one of the partners of the RCN SJM, which aims to increase the quality of digital services by combining formal methods with user journeys and process mining.

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 and usable by both computers and humans and should ideally represent the concepts and relationships in a manner to which domain experts are accustomed. This way users and systems may explore and extract implicit information from data through the help of automated 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 objects. This makes maintaining the ontology a real issue.

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

To do this, a pattern-based approach for engineering ontologies is employed, developing tools and methods that 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.

OTTR is a language and framework for representing and instantiating recurring patterns for engineering ontologies. This allows building and interfacing with the ontology at a higher level of abstraction than what is possible using the current standard ontology language OWL. This includes:

- Building ontologies and knowledge bases by instantiating templates;
- presenting, transferring and visualising the knowledge base as a set of template instances at different levels of abstraction; and
- securing and improving the quality and sustainability of the knowledge base via structural and semantic analysis of the templates used to construct the knowledge base.

We believe the benefits of using the OTTR framework for engineering ontologies are many:

- **Better abstractions:**
  The use of templates as building blocks for knowledge bases makes knowledge bases easier to understand and work with since modelling constructs are expressed using patterns at a suitable abstraction level rather than the low-level RDF triples or OWL axioms.

Martin G. Skjæveland

- **Uniform modelling:** Knowledge bases built using a well-designed template library makes its modelling more uniform, which greatly improves its ease of use and maintenance.
- **Modular, encapsulated patterns:** The fact that templates can be defined by referring to other templates follows the don't repeat yourself (DRY) principle, and leads to templates that are easier to maintain and, in many cases, easier to understand.
- **Separation of design and content:** Instantiable template naturally establishes a clear separation between the design of the knowledge base—represented by templates, and the bulk content of the knowledge base—represented by template instances. The templates and the template instances can then typically be managed by different competencies and tools: templates are managed by ontology experts using specialised tooling such as ontology editors and reasoners, while templates instances are created by domain experts using tools they are familiar with, e.g., spreadsheets.
- **Open standards support:** OTTR is designed to work smoothly with relevant core W3C standards, such as, RDF, OWL, SPARQL, XML, XSD, SAWSDL, XSLT. (For the XML standards family only rudimentary prototypes have been tested.)
- **Publish, share and reuse:** Templates can be published and shared using common web publishing practices, such as Best Practices for Publishing Linked Data.
- **Tool support for maintenance:** The formal basis of OTTR template, i.e., RDF graphs and OWL ontologies, allow templates to be analysed and quality assessed using OWL reasoners, but also using well-known ideas from software engineering, such as identifying redundancies and suggesting refactoring for poorly integrated templates.

All results produced by the project is available the website **http://ottr.xyz.** This includes:

- **Lutra**, our open-source reference implementation of working with OTTR templates to instantiate and expand instances to RDF graphs and OWL ontologies.
- The **core OTTR template library**, a open library of shared, reusable templates for representing basic RDF and OWL modelling patterns.
- Screencasts, demonstrations, tutorials, and an interactive primer
- Academic publications and posters

Scalable Computing

The Scalable Computing (SC) research program is about making data access and processing fasterto 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 using hardware accelerators.

The Scalable Computing has the following areas of intervention:

1. Better and Flexible Execution Platforms:
The focus is on making advanced execution platforms available to all SIRIUS partners through open interfaces that can be used remotely allowing researchers without direct access to the computers to use and experiment with different hardware configurations for their applications.

The methodology on the HPC side is largely experimentation with different computing platforms built on technology from SIRIUS hardware partners and to evaluate the performance of these platforms for real applications of the SIRIUS application partners. The work will therefore involve experimental software design and hardware architectures for scalable computing ranging from accelerators to numerical methods Stochastic combinatorial optimization is the methodology used for managing applications across different Cloud providers to allocate the application components where they give the best utility for the application owner, and to reconfigure in response to changing application execution context.

Activities:

- The NUMAScale computer that used to be part of the academic HPC infrastructure has been decommissioned and moved to the Department of Informatics as a SIRIUS controlled experimental computational platform.
- The main machine and the storage for this computer has been upgraded
- Work is ongoing to link this computer with NREC to enable its use as Cloud HPC platform to demonstrate how HPC applications can benefit from using Cloud computing
- A new and flexible HPC cluster architecture based on PCIe has been established and is under testing.
- Selection has started for the hardware for a future exascale experimental platform.

2. Scalable Application Support:
On one side this will continue the support to open source for better numerical computations for reservoir simulations, and on the other side it will continue the development of Cross-Cloud and Multi-Cloud application management middleware.

Scalable Computing successfully delivered the Horizon 2020 project MELODIC2 in 2020. The result of the MELODIC project is a multi-Cloud application management platform. This is available as open source from the main European open-source community OW23, or as a supported and installed package on standard commercial terms4. There are already several paying customers of MELODIC, and it currently has support for managing applications across all the big Cloud providers, Amazon Web Services (AWS5). Google Cloud6, and the Azure Cloud7 from Microsoft. Additionally, there is support for some smaller European Cloud providers, and the open-source Cloud infrastructure management platform OpenStack8, which is used by most academic Cloud installations worldwide, among them the he Norwegian Research and Education Cloud (NREC9).

Activities:

- The research activities on numerical methods for reservoir simulations and the associated code optimizations have continued
- The work on installing the Cloud management software Open Stack on the moved NUMAScale computer has started
- The Cloud management software from the Horizon 2020 MELODIC project has been successfully demonstrated for several applications and will be installed on the NUMAScale machine once Open Stack is running.
- PhD project starting on utility based autonomic computing

3. SIRIUS Application Execution:
To evaluate the research delivered under the previous objectives, demanding real world applications from the SIRIUS partners may be tested using the software and the hardware available.

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.

Activities:

- Interactions with Equinor on the reservoir simulations through two ongoing PhD projects
- Identification of SIRIUS partner applications in need for scalability and the projects of Scalable Computing have been presented to the industrial partners on multiple occasions.

Cross-Cloud application management

The calable Computing researches were awarded another Horizon 2020 project called MORPHEMIC[10] to extend the successful MELODIC platform. MELODIC supports application management through the Application Programming Interface (API) offered by the various Cloud providers and it is thereby able to deply Cloud computing instances of the application's components in the form of virtual machines (VMs), containers, and serverless functions. However, artificial intelligence (AI) applications may benefit from using specialized hardware when training the algorithms. These are accelerators like Graphical Processing Units (GPU), Field Programmable Gate Arrays (FPGA) or Tensor Processing Units (TPUs) tailored for TensorFlow[11] processing. Hence, an application component may come as a standard Central Processing Unit (CPU) artefact, which is currentl being deployed by MELODIC, or as artefacts compiled for one or more of the hardware accelerators. Furthermore, the acclerated version will only be beneficial during the training of the AI components, and more costly to use in the Cloud than standard CPUs. It is therefore a need to switch between different component artefacts depending on the application's need, and MORPHEMIC adds this support.

Håkon K. Stensland

Xing Cai

Geir Horn

The optimization in MELODIC is largely reactive based on measured changes in the managed application's execution context. Aquiring Cloud resources may unfortunately take several minutes, and the execution context may therefore manage to change significantly before the reconfigured

application is up and running on the new resources. MORPHEMIC aims at remedying this by *proactive adaptation* based on real time forcasting for time series of the measurements of the running application's execution context, and perform the optimization and reconfiguration early so that the Cloud resources are available when they are needed by the application.

MORPHEMIC started in 2020 just as the Covid-19 restrictions came into effect. The project has still managed to survey the state of the art to identify forecasting methods for which extensive experiments have been carried out, despite the travel restrictions hampering collaboration and research progress. The forecasting module is now being integrated with MELODIC and the first applications proactively optimized will be demonstrated during 2021. The initial steps toward making utility function modelling easier for the MORPHEMIC users have also been successfully completed. These results will be enhanced and integrated with MORPHEMIC's optimization architecture during the remaining two years of the project.

Smart Scalable PCI Express

I/O resources like NVMe[12], GPUs, FPGAs are today installed in many modern servers and computers. Today, these resources are only available to applications running on the same server without sub-optimal networks and software. The Smart Scalable PCI Express (PCIe) project's goal is to solve the network inefficiency and enable servers connected by PCIe Gen4 networks to access remote I/O resources and achieve the same performance as if the I/O devices were local.

In 2020, the project reached several important milestones: We have improved Dolphin's existing SmartIO software technology and solutions and enable sharing and migration of PCIe Gen4 devices. Work is also underway with extending the existing SISCI API to allow sharing of non single root input/output virtualization (SR-IOV) NVMe devices and direct transfers between NVMe and memory specified by users, e.g. main memory or GPU memory. A prototype and experimental testing of the new PCIe NVMe driver (DIS_NVMe) allows several systems to share Non-SR-IOV NVMe devices across PCIe Gen4 in early

In 2020, the project reached several important milestone

2021. Work is also underway to develop a shared disk and distributed file system optimized for shared NVMe disks.

Advanced HW accelerators for HPC

Compute Express Link (CXL) is one of the proposed next-generation high-speed interconnect buses designed to connect CPUs, memory, and accelerators in servers, datacentres, and computers.

CXL version 2.0 was released in November 2020 and uses the same physical and link layer as PCI Express 5.0. The CXL standard is also backed by all the large hardware vendors such as Intel, AMD, ARM, Nvidia, etc.

This project has started investigating the potential of the CXL bus, both as a next-generation replacement for PCI Express and the possibility of using it as a next-generation chip for NUMAScale to connect accelerators, memory, and CPUs in an HPC environment. Activities in this project in 2020 have been preliminary discussions about the possibility of the CXL standard for multi-host communication or vendor-specific extensions for adding the functionality in future versions. This work will continue in 2021.

HPC Suport for reservoir simulation

The demonstration project that has seen substantial contributions from the Scalable Computing team is the Digital Field and Reservoir Management project, where our 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, the researchers investigate how to improve the various algorithmic and implemental aspects of such computations, so that reservoir simulations can be carried out faster and become capable of efficiently using more compute nodes.

Continuing in the spirit of open science, SIRIUS reserchers during 2020 collaborated actively with Equinor researchers in further improving the performance and capabilities of the reservoir flow simulator inside the Open Porous Media (OPM[13]) initiative. Specifically, the topic of concurrent, point-to-point MPI message exchanges on heterogeneous interconnect topologies was investigated. Extensive research has been carried out to quantatively model the cost of such heterogeneously competing communication tasks. The expected impact on parallel performance improvement will not be limited to reseservoir simulations, but applicable to a wide range of parallel scientific computations. Moreover, SIRIUS researchers have assisted the OPM community in providing more flexibilities in the parallel numerical treatment of the physical wells in oil reservoirs, aiming at further computational performance boost. During the second half of 2020, a new SIRIUS PhD candidate started in another 3-year collaborative project with Equnior, with the purpose of improving the performance and capabilities of ensemble simulations of oil resevoirs. Parallel programming techniques will be combined with advanced numerical strategies, as well as machine learning, to secure both high performance and better reliability.

2. https://h2020.melodic.cloud/
3. https://gitlab.ow2.org/melodic
4. https://melodic.cloud/
5. https://aws.amazon.com/
6. https://cloud.google.com/
7. https://azure.microsoft.com/en-us/
8. https://www.openstack.org/
9. https://www.nrec.no/
10. https://www.morphemic.cloud/
12. https://www.tensorflow.org/
12. Non-Volatile Memory host controller Express interface specification (NVMe), see https://en.wikipedia.org/wiki/NVM_Express
13. https://opm-project.org

Domain-Adapted Data Science

Imagine a child who, either in person or in pictures, has seen the usual animals that live on Norwegian farms, such as sheep and horses, but has never seen a giraffe, neither in person nor in pictures. If the child has sufficient language skills, then you can tell them, before going to a zoo, that a giraffe is an animal that looks like a horse but with a very long neck. Then, at the zoo, most likely the child will correctly identify a giraffe as a giraffe with ease.

Basil Ell

What this example shows is that humans can combine what they have learned from experience (in our example, what they have seen before) with declarative statements (the description of the similarities and differences between giraffes and horses). Machines are not yet good at that.

Within artificial intelligence research one can identify two major strands that differ in how knowledge is represented and how it is processed:

a. Knowledge is represented symbolically, for example in the form of ontologies, and is processed by emantic reasoning.

b. Knowledge is represented sub-symbolically, for example in terms of vectors, and is used for machine learning.

While data used for machine learning in general needs to be "big", data representing symbolically represented knowledge is usually very small and concise. Traditionally, there is little overlap between these two strands. Each strand has its strengths and weaknesses and to a large extent complements the other. And as with the child and the giraffe, these two ways of thinking seem to coexist in a tightly coupled way in our minds.

Applying hybrid approaches combining symbolic and sub-symbolic knowledge can therefore lead to better results with less training data. Additionally, the symbolic knowledge may lead to more explainable algorithms. Within our research program, we develop novel hybrid approaches that identify and exploit these capabilities. The goal is a general methodology for how data science tasks can be enhanced through the combined use of symbolic and sub-symbolic knowledge.

One term that we use to refer to a particular class of hybrid approaches is what we refer to as domain-adapted approaches. Very often in machine learning and data science tasks, data in the form of textual documents, images, or tables is processed, where making use of domain knowledge, for example in the form of an ontology, can improve the results.

Within our research program, we have developed hybrid approaches, gained evidence for the benefits of hybrid approaches, and work towards developing novel hybrid approaches:

- Erik Bryhn Myklebust, Ernesto Jimenez-Ruiz, Jiaoyan Chen and colleagues have shown in the context of ecotoxicological effect prediction that the accuracy of predictions can be improved when domain knowledge is incorporated into the prediction model. [1]

- Egor V. Kostylev and colleagues established fundamental connections between Graph Neural Networks (GNNs), a modern structure-aware machine learning architecture, and classic logic-based knowledge representation formalisms, such as first-order and description logics. In particular, they have shown that plain GNNs can learn exactly those first-order formulas that are in description logic ALCQ. [2]

- In the context of a task relevant for the oil and gas industry, namely reservoir analogue identification, Summaya Mumtaz and Martin Giese have shown that a similarity measure based on the combination of domain knowledge (in the form of a taxonomy) with classical frequency-based features leads to significantly better results. [3]

Monogrammist M S, and Hans Adam. Eyn seltzam vnd Wunderbarlichen Thier/ Der gleichen von Uns vor nie gesehen worden : u. Dis thier wird Surnappa genant Vnd ist von der erden an mit Sampt dem kopff/ höher als fünff mañ hoch. Nürnberg: durch Hans Adam, 1559. Print. **https://www.e-manus cripta.ch/zuzneb/doi/10.7891/e-manuscripta-92218**

- In the context of the natural language processing task of named entity recognition, Farhad Nooralahzadeh, Lilja Øvrelid and colleagues have developed an approach that considers domain knowledge to generate annotated training data in a distantly supervised manner while significantly reducing the noise that typically occurs in distant supervision and established a new state-of-the-art on four benchmark datasets taken from different domains and different languages. [4]

- In the context of classification, based on a use case that is relevant to the oil and gas industry, namely that of excess inventory reduction, Daniel Bakkelund has developed theory and methodology for improved classification of interchangeable equipment, by integrating equipment structure awareness into classical methods for unsupervised machine learning.

- Jiaoyan Chen, Ernesto Jimenez-Ruiz, Ole Magnus Holter, Ian Horrocks and colleagues have developed an ontology embedding framework named OWL2Vec* that can embed symbolic knowledge in an OWL ontology into a vector space, so that the information can be consumed by machine learning algorithms. OWL2Vec* can be directly applied to ontology completion tasks such as subsumption prediction as well as to help address machine learning challenges, such as sample shortage, by injecting symbolic knowledge. [5,6]

- Current Explainable Artificial Intelligence (XAI) techniques rely on the data to analyse and explain the behaviour of models. However, a human-comprehensible explanation of machine learning models can only be obtained when domain knowledge is exploited to bridge between the model and human concepts. Peyman Rasouli and Ingrid Chieh Yu aim to integrate domain knowledge with unstructured data to provide feasible, comprehensive, and faithful explanations.

- Gong Cheng, Evgeny Kharlamov and colleagues have shown how to summarise symbolic domain knowledge with the help of machine learning, by first embedding such knowledge into a low-dimensional vector space and then applying reinforcement learning over the results of the embedding [7]. In another work, they have shown how the most relevant fragment of domain knowledge can be identified via a neural network on top of the knowledge embedded into a low dimensional vector space. [8]

- Basil Ell is developing approaches to align symbolic data (i.e., ontologies) with sub-symbolic data (e.g., texts or tables). The alignment enables labelled training data

to be generated via distant supervision for approaches such as information extraction for ontology population or natural language generation. Having symbolic and sub-symbolic data aligned means to obtain hybrid data that can be processed by hybrid approaches.

The full potential of hybrid approaches has yet to unfold. There are several ways in which the gap can be bridged and a hybrid approach can be realised. For example, in a high-level perspective, given symbolically represented knowledge A and sub-symbolically represented knowledge B, one can transform A into sub-symbolically represented knowledge A' and combine it with B and then apply classical machine learning. Embedding of knowledge graphs into vector spaces is an example of this approach. In contrast, one can also extract symbolic representations from unstructured data (e.g., from text via Information Extraction, Ontology Learning, Semantic Parsing, or Knowledge Base Population) and combine it with symbolic representations, on which one can then perform semantic reasoning. Besides these two extreme approaches (having everything symbolically represented vs. having everything sub-symbolically represented), there are many other approaches, such as, making use of symbolically represented knowledge to pre-process (e.g., clean or otherwise improve) data that can then be used as training data for a machine learning algorithm or, as another example, a neural network learns when it needs to perform numerical reasoning (e.g., addition, subtraction, sorting and counting) and then externalises counting to a unit that performs numerical reasoning symbolically.

Our ambition in this research program is to develop novel hybrid approaches that enable domain adaptation so that results of data science activities can be improved and become more explainable. In 2021 we will start the Domain-adapted Data Science Pipeline project. The goals of the project are to

- identify tasks that benefit from domain adaptation,
- develop a roadmap for further research,
- develop a conceptual framework to describe and compare assemblies of domain-adapted data science tasks,
- create a survey of approaches,
- unify existing approaches (ours and beyond) into a methodology,
- acquire funding to further intensify our work,
- intensify the exchange and collaboration with our partners.

References

[1] Prediction of Adverse Biological Effects of Chemicals Using Knowledge Graph Embeddings. Erik Bryhn Myklebust, Ernesto Jimenez-Ruiz, Jiaoyan Chen, Raoul Wolf, and Knut Erik Tollefsen. Semantic Web Journal (under review)

[2] The logical expressiveness of graph neural networks. Pablo Barceló, Egor V. Kostylev, Mikael Monet, Jorge Pérez, Juan Reutter, and Juan Pablo Silva. International Conference on Learning Representations. 2019.

[3] Frequency-Based vs. Knowledge-Based Similarity Measures for Categorical Data. Summaya Mumtaz and Martin Giese. AAAI Spring Symposium: Combining Machine Learning with Knowledge Engineering (1). 2020.

[4] Reinforcement-based denoising of distantly supervised NER with partial annotation. Farhad Nooralahzadeh, Jan Tore Lønning, and Lilja Øvrelid. Association for Computational Linguistics, 2019.

[5] OWL2Vec*: Embedding of OWL ontologies. Jiaoyan Chen, Pan Hu, Ernesto Jimenez-Ruiz, Ole Magnus Holter, Denvar Antonyrajah, and Ian Horrocks (2020). Machine Learning. (to appear)

[6] https://github.com/KRR-Oxford/OWL2Vec-Star

[7] Entity Summarization with User Feedback. Qingxia Liu, Yue Chen, Gong Cheng, Evgeny Kharlamov, Junyou Li, and Yuzhong Qu. ESWC 2020: 376-392

[8] Neural Entity Summarization with Joint Encoding and Weak Supervision. Junyou Li, Gong Cheng, Qingxia Liu, Wen Zhang, Evgeny Kharlamov, Kalpa Gunaratna, Huajun Chen. IJCAI 2020: 1644-1650

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 demonstration 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.

Digitisation 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 Application Programming Interfaces (APIs), which may or may not adhere to various standards. 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.

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. The Ontop project is hosted by the Free University of Bolzano, and is also commercially supported by the company Ontopic, which became SIRIUS partner in 2020. SIRIUS is continuing to contribute to the development of this tool. Ontop can expose the content of arbitrary relational databases as knowledge graphs. These graphs are virtual, which means that data remains in the data sources instead of being moved to another database. At query time, Ontop translates SPARQL queries expressed over the knowledge graphs into SQL queries executed by the relational data sources. The recent releases of Ontop have improved its compliance with relevant W3C recommendations and provides good performance in query answering. It supports almost all the features of SPARQL 1.1, GeoSPARQL, R2RML, OWL2 QL and SPARQL entailment regime, and the SPARQL 1.1 HTTP Protocol. In 2020 support for aggregates in queries were added. Our plans are to extend Ontop further with mechanisms that support 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-

> **We focused** our work, and made significant progress, on query answering over unrestricted OWL 2 ontologies.

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

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

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

CQ answering

We focused our work, and made significant progress, on query answering over unrestricted OWL 2 ontologies. We proposed a few improvements to PAGOdA, an OWL 2 ontology reasoner for conjunctive query answering with a unique modular approach, first designed in 2015. We extensively documented our results and showed that the introduction of our improvements makes the difference between feasibility and unfeasibility on some harder queries.

Dag Hovland

Part of the effort went into an extensive analysis of sound ontology approximations into tractable fragments w.r.t. fact entailment and CQ answering. Finally, we i mproved the integration with RDFox by making use of its latest features.

Distributed reasoning

Over the past 12 months we have developed new partitioning schemes to improve reasoning performance over distributed RDF graphs. The strategy used to partition the input data can determine not just how fast can Datalog materialisation can be computed, but if it is feasible at all within certain hardware and network constraints. We have adapted two streaming partitioning schemes originally developed for general undirected graphs, and showed how they can significantly improve reasoning performance, but without unrealistic demands on the memory of the servers used for partitioning.

KG construction, curation, etc.

In 2020, we have made significant progress in Knowledge Graph (KG) construction, integration and refinement, with several results published, accepted, or submitted. Regarding KG construction, we jointly organized SemTab: Semantic Web Challenge on Tabular Data to Knowledge Graph Matching in ISWC'20, and the resources from SemTab in 2019, such as evaluation data and metrics, have been published in ESWC'20 [1]. Regarding KG integration, we developed LogMap-ML, an extension of LogMap with machine learning, which has achieved promising results for aligning a health lifestyle ontology (HeLis) and a food ontology (FoodOn) for SamSung Research UK [2]. Jiaoyan Chen also collaborates with Tencent Technology for matching KGs that are largely composed of facts: he acts as an equivalent first author for a work that benchmarks embedding-based KG entity alignment [3] and acts a co-author for a new KG entity alignment system (submitted to IJCAI'21) that combines probability reasoning and semantic embedding for state-of-the-art performance. Regarding KG refinement (or curation), we developed a

In 2020, we have made significant progress in Knowledge Graph (KG) construction, integration and refinement, with several results published, accepted, or submitted.

new ontology embedding method named OWL2Vec* for two ontology completion tasks – class subsumption and membership prediction. We have code implementation, case studies on HeLis, FoodOn and the Gene Ontology, and a paper accepted by Machine Learning [4]. We also studied the refinement of KGs largely composed of facts: (1) correcting erroneous facts, with one paper accepted in WWW'20 [5] and one paper submitted to Semantic Web Journal [7]; and (2) predicting relational facts (a.k.a. link prediction) under sample shortage, i.e., zero-shot KG completion, for which an IJCAI survey track paper was just accepted [7], and Jiaoyan Chen acted as the second author for OntoZSL [8].

Modal Logics for Engineering

Motivated by the need to model the concepts of "verification" and"requirement" in the READI project, we developed a multi-modal logic with the two modalities for "verified" and "required". The standard Kripke structures are used to specify the semantics of these two new modal operators, whereas the classical semantics of first-order logic is used for the standard classical operators and quantifiers. We have developed a modal sequent calculus for this multi-modal logic as a basis for an automated proof search. The calculus was implemented and the resulting tool can be used to automate formal reasoning overspecifications containing the new "verified" and "required" operators.

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*.

References

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

[2] Augmenting Ontology Alignment by Semantic Embedding and Distant Supervision, Jiaoyan Chen, Ernesto Jimenez-Ruiz, Ian Horrocks, DenvarAntonyrajah, Ali Hadian, Jaehun Lee. ESWC, 2021. (accepted)

[3] An Industry Evaluation of Embedding-based Entity Alignment, Ziheng Zhang, Jiaoyan Chen, Xi Chen, Hualuo Liu, Yuejia Xiang, Bo Liu, Yefeng Zheng, COLING, 2020.

[4] OWL2Vec*: Embedding of OWL Ontologies, Jiaoyan Chen, Pan Hu, Ernesto Jimenez-Ruiz, Ole Magnus Holter, DenvarAntonyrajah, Ian Horrocks, Machine Learning, Springer, 2021. (accepted)

[5] Correcting Knowledge Base Assertions, Jiaoyan Chen, Xi Chen, Ian Horrocks, Ernesto Jiménez-Ruiz and Erik B. Myklebust. The Web Conference (WWW), 2020.

[6] Correcting Assertions and Alignments of Large Scale Knowledge Bases, Jiaoyan Chen, Ernesto Jiménez-Ruiz, Xi Chen, Ian Horrocks, and Erik B. Myklebus. (under review)

[7] Knowledge-aware Zero-shot Learning: Survey and perspective, Jiaoyan Chen, Yuxia Geng, Zhuo Chen, Ian Horrocks, Jeff Z. Pan and Huajun Chen. IJCAI Survey Track, 2021. (accepted)

[8] OntoZSL: Ontology-enhanced Zero-shot Learning, Yuxia Geng, Jiaoyan Chen, Zhuo Chen, Jeff Z Pan, Zhiquan Ye, Zonggang Yuan, Yantao Jia, Huajun Chen. The Web Conference (WWW), 2021. (accepted)

[9] Temitope Ajileye, Boris Motik, and Ian Horrocks. 2019. "Datalog Materialisation in Distributed RDF Stores with Dynamic Data Exchange." In ISWC.

[10] Temitope Ajileye, Boris Motik, Ian Horrocks. 2021. "Streaming Partitioning of RDF graphs for Datalog Reasoning." In ESWC. (accepted)

[11] Federico Igne, Stefano Germano and Ian Horrocks. 2021. "Computing CQ lower-bounds over OWL 2 through approximation to RSA." In ISWC. (under review)

[12] Elem Güzel Kalaycı, Irlan Grangel Gon ´alez, Felix Lösch, Guohui Xiao, Anees ul Mehdi, Evgeny Kharlamov, and Diego Calvanese. Semantic integration of Bosch manufacturing data using virtual knowledge graphs. In ISWC, 2020.

[13] Guohui Xiao, Davide Lanti, Roman Kontchakov, Sarah Komla-Ebri, Elem Güzel-Kalayci, Linfang Ding, Julien Corman, Benjamin Cogrel, Diego Calvanese, and Elena Botoeva. The virtual knowledge graph system Ontop. In ISWC. 2020,

[14] Linfang Ding, Guohui Xiao, Diego Calvanese, and Liqiu Meng. A framework uniting ontology- based geodata integration and geovisual analytics. ISPRS Int. J. Geo-Inf., 9(8):474, 2020.

Industrial Digital Transformation

Digital transformation is not easy. While new digital technologies offer the potential to transform and improve organizational practices, both experience and research shows that many digitalization efforts fall short of meeting these expectations.

The research program on Digital Industrial Transformation approaches these challenges from the perspective of implementation and adoption of digital technologies. In particular, we hold that in planning and organizing for digital transformation, companies need to identify and cultivate the organizational preconditions necessary for realizing the potential of digital technologies.

We have pursued this through close collaboration with selected industrial partners throughout the lifetime of SIRIUS. While our academic goal is to be at the forefront of theory development on the dynamics and mechanisms of digital transformation, our approach is to work closely with partner companies to support them in planning and organizing for digital transformation.

Our research serves both an academic purpose (i.e. publications, education), as well as providing the centre's computer scientists with researchable challenges to be translated into working prototypes. These prototypes can, in turn be tested and evaluated by user organizations. Most notable of here is the Geological Assistant. With basis in Equinor's observation that the G&G tools did not "speak geologists's language", we worked closely with Equinor and Schlumberger to elaborate how the existing tool-set provided geologists limited functionality to support their narratively-oriented mode of reasoning.

We have pursued this agenda through multiple venues the past year, where SIRIUS is part of the research program participants' broader engagement with digital transformation of the oil and gas industry. We want to highlight three activities here.

Prototyping support for subsurface analytics. We have, together with the research program on Semantic Integration, collaborated with a group of data scientists in Equinor. This has been early-stage research, aimed at identifying researchable challenges that aligns with the corporations' increased focus on data-driven decisions through Big Data analytics.

Thomas Østerlie

Digital delivery of infrastructure projects. We are engaged with the offshore construction industry's efforts towards digital project delivery at multiple arenas. Our role here has predominantly been to work actively with industrial partners and offer advice on how to plan and organize the transition towards digital project delivery. Mina Haghshenas (who is now finishing off her PhD) has been involved in the READI project, whereas Thomas Østerlie has been engaged with the NORSOK Z-TI expert group. Østerlie is developing the Digital Transitions Framework (DTS) based on these activities. The DTS seeks to operationalise existing scientific knowledge on implementation, adoption and use of digital technologies into actionable methods and recommendations for planning and implementing large-scale digital transformation projects.

Executive masters course in remote operations. Elena Parmiggiani has developed a new executive masters course in remote operations for the energy sector. The course teaches participants to identify and critically evaluate technical and organizational challenges with planning and conducting remote operations. The course is delivered as part of NTNU Videre's Executive Master's program in technology management and digital transformation.

Analysis of Digital Twins

Background

The digital twin is a vision for a technology, originally conceived for NASA's space program, enabling industry to significantly improve the life-cycle management of physical assets. A digital twin is typically a system which collects data about a physical asset (such as a plant or a reservoir), continuously revises this data set through, e.g., updates reflecting changes to the asset's structure and sensor data reflecting the physical asset's state and uses this data to monitor and make predictions about the physical asset. The digital twin can be thought of as a three-layered structure: the data sources, an information layer, and an insight layer. Industrial focus is today mainly on collecting data into shared, and increasingly structured, data sets which we think of as the information layer, and on providing dashboard-like insights into the system.

This project will focus on analysis support for digital twins, by building or combining tools which can leverage the information layer into insights. The purpose of these tools can be to reproduce and explain past events, to explore alternatives for decision making, to prepare for incidents or to optimize production. A central goal for this project is to combine semantics, behavioral and conceptual modelling techniques, and analysis methods in the context of digital twins.

Einar Broch Johnsen

Methodological background for the work is an integration of ontology-based conceptual modelling techniques, formal methods, and data-driven techniques for system analysis.

The work has synergies with PeTWIN and the Digital Twins demonstration.

Objectives:

- Understand the design space for coupling behavioral and conceptual models.
- Develop methods that combine structured information with behavioral analyses.
- Understand how conceptual modelling can be used to integrate analyses results.
- Develop experience with semantics foundations for co-simulation.
- Develop methods for decision making with digital twins.

Activities:

- Develop a formal theory of coupled behavioral and conceptual models.
- Develop prototype tool for programming with semantics.

References

[1] Eduard Kamburjan, Rudi Schlatte, Einar Broch Johnsen, Silvia Lizeth Tapia Tarifa: Designing distributed control with hybrid active objects. In: ISoLA, Springer (2020) In press.

[2] Eduard Kamburjan, Vidar Norstein Klungre, Rudolf Schlatte, Einar Broch Johnsen, Martin Giese: Programming and Debugging with Semantically Lifted States. In: ESWC, Springer (2021). In press.

[3] Eduard Kamburjan, Vidar Norstein Klungre, S. Lizeth Tapia Tarifa, Rudolf Schlatte, Einar Broch Johnsen, Martin Giese, and David Cameron: Emerging Challenges in Integrated Semantic and Formal Methods for Digital Twins. Position paper in progress 2021, available on request.

- Develop methods for coupling simulators by means of semantics and constraint representations.
- Develop methods for exploring semantics to express "what-if" scenarios in multi-model simulation and analysis
- Collaborate with and disseminate results through PeTwin

Progress in 2020

A prototype domain specific language SMOL has been developed which supports high-level ways to program the orchestration of simulation units which implement the FMI standard. SMOL seamlessly integrates semantic technology with programming constructs for interacting with simulators. SMOL has a formally defined semantics and a prototype runtime implemented in Kotlin, a dialect of Java. SMOL addresses the composition problem for simulators with different domain models through lifting semantic technologies to enable correct configuration and orchestration of connections inside the digital twin. By embracing the semantic technologies, formal methods can contribute to the development of provably correct digital twins.

Publications

- The paper "Designing distributed control with hybrid active objects" explores different ideas for distributed orchestration of simulation units appeared at ISoLA 2020 [1].
- The first results on SMOL will be presented at ESWC 2021 [2].
- A position paper on our approach is under development [3]. We are happy for comments and suggestions and use cases that can help bring these ideas forward.

Digital Requirements, Integrated Asset Modelling and Standards

During 2020, the SIRIUS centre has built up the core of substantial activity around asset information modelling and its use as an enabler for digital transformation of engineering projects which can lead to more effective digital twins in the future. We have been working with partners in all parts of the industry supply chain to develop our research programs for digital twin and asset modelling. Our aim is to use a combination to projects funded by the Research Council of Norway, the European Union and joint industry projects to finance a robust centre of competence hat can support the implementation of transformative projects in the industry.

Highlights of the programs in 2020 are:

- The start-up of the PeTWIN collaborative project on digital twins for field management, together with Equinor, Shell, Petrobras and the Federal University of Rio Grand do Sul. This project is financed by the Research Council of Norway and FINEP, the Brazilian federal innovation organization. This is described in more detail below.

- The start-up of the DSYNE INTPART network. This network will build competence and courses in the field of digital requirements in engineering. The network brings together leading groups in System Engineering and Computer Science from Norway, Brazil and the United States. We are also collaborating with industrial end-users, including Shell, Equinor, TechnipFMC and DNV.

- The award of two European H2020 projects, ONTOCOMMONS and Eur3ka. The ONTOCOMMONS project is a Coordination and Support Activity, which has the aim of building up a shared semantic foundation for applications in manufacturing and materials science. The University of Oslo and SINTEF are the Norwegian partners in this work, and Aibel is participating in an application case study for the work. Eur3ka (EUropean Vital Medical Supplies and Equipment Resilient and Reliable Repurposing Manufacturing as a Service NetworK for Fast PAndemic Reaction) is financed under the European Union's pandemic response program. It will build and demonstrate an IT framework that will enable manufacturing companies to repurpose production quickly and effectively.

- SIRIUS provided the project contract and collaboration for the KONKRAFT digital design project. This project involved Equinor, AkerBP, Lundin Energy, Aker Solutions (with Aize and Cognite), Aibel and TechnipFMC. This is described in more detail below.

- SIRIUS personnel have worked in the READI Joint Industry Project. This work is coordinated by DNV and has involved other SIRIUS partners: Computas, Equinor, Aker Solutions, Aibel and TechnipFMC. The READI project has worked on the formulation digital requirements that can replace current document-based requirements and standards, a work that is directly relevant to the DSYNE network, We plan to use DSYNE as a platform to inter-nationalize this work. READI has also developed a new and promising concept for asset information modelling.

Pragmatic Use of Standards

Digital twins are built by integrating data from many sources. The data has many formats and the names used to identify the data are inconsistent. This makes building and maintaining digital twins difficult and expensive. Standards provide ways for industry to agree on how they structure and share data. Unfortunately, there are many different, competing standards available and it is difficult to navigate this jungle. The oil and gas industry has worked in many years to develop the ISO15926 family of standards,

Machines that work faster and more accurate than humans should take over repetitive manual tasks and deliver faster, cheaper and with higher quality.

whereas the manufacturing and aerospace industries developed the ISO10303 standards. Joint projects in the industry have also worked on candidate standards. IOGP, the International Oil and Gas Producers' association, run two joint industry projects, JIPs. JIP36 is developing the CFIHOS (Capital Facilities Information Handover Specification) standards, which provide data models for engineering oil and gas facilities. JIP33 prepares standard specification sheets for equipment types. Finally, DEXPI, which is developed by the process industries, describes detailed data models for all the pieces of equipment that can be found on Piping and Instrumentation diagrams.

In Norway, the READI Joint Industry project has been working on transforming standards and requirements from a text- and document-based form to a digital, data-driven form. This project has been a driver for adoption and development of SIRIUS' tools for ontology engineering and domain-adapted data science for several years. As work proceeded, it became apparent that digital requirements needed to be accompanied by a flexible asset model, that defined the systems and equipment that the requirements referred to.

Existing asset models are built using a hierarchical break-down of a system into sub-systems and components. The breakdown is done using a single criterion, usually related to the function of the systems. Thus, the NORSOK tagging system breaks a facility into main systems, sub-systems and equipment types and uses a numerical code to indicate an item's place in the hierarchy. The ISA95 hierarchy breaks the facility down in a similar way. Functional locations in an asset management program like SAP PM or IBM Maximo can be organized by either function or location, but not both. This makes it complex to manage requirements and interdisciplinary work, as the asset model is not rich enough to capture the various aspects of requirements and engineering.

These limitations have been addressed by the ISO/IEC81346 standard. This standard provides a way to break down a complex system in several ways, so that there are

Arild Waaler

David Cameron

parallel, but interlinked hierarchical breakdown structures. The framework for building these is called a Reference Designation System (RDS). The main principle in this modelling is that every element in an asset model can be understood as a system that can be viewed from different perspectives:

- The *function* aspect: what the system is intended to do, i.e. its purpose.
- The *product* aspect: what the system is constructed of. This allows a product type breakdown into parts.
- The *location* aspect: the space that the system occupies.

Building an RDS-based asset model needs an agreed set of codes and system types. This is produced by industry experts for a specific type of industry. The standard has codes for power plants (ISO/IEC81346-10) and the building industry (ISO/IEC81346-12). The READI project has developed its own RDS for the oil and gas industry. You can get a copy of the draft for this standard at **https://readi-jip.org/reference-designation-system-for-oil-and-gas/**.

The RDS is one component in a complex set of technologies that need to work together to enable new ways of working on capital projects and building digital twins. Information structured using the RDS is shared and made available by

technologies adopted by the Industrie 4.0 program, notably OPC UA and the Asset Administration Shell. SIRIUS is working with is operating company and engineering company partners to develop and apply our knowledge of semantic technologies to making these technologies work together in the best possible way.

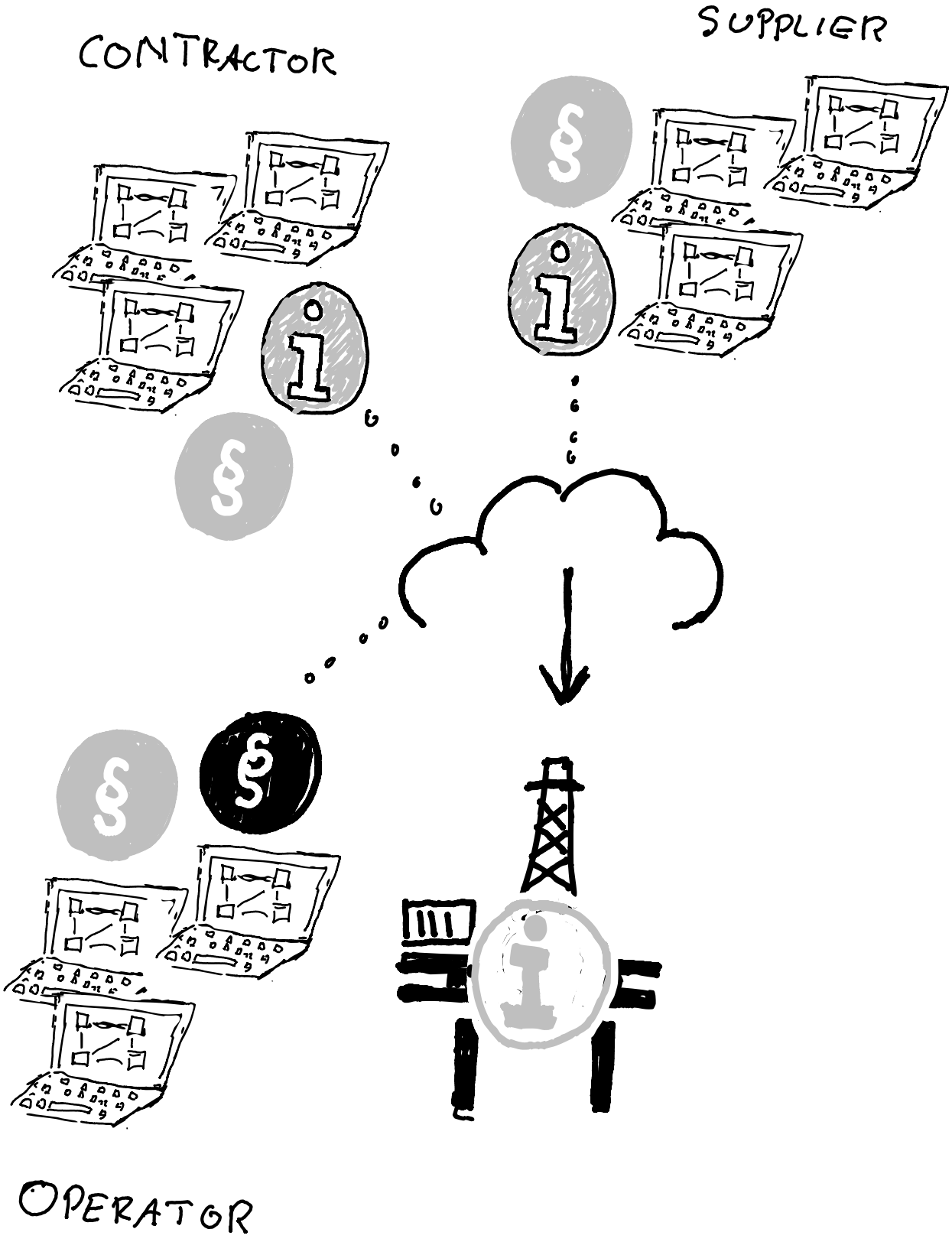
Asset modelling as a key to manage fragmented information

Industry is in the midst of a digital transformation. This comes with promises and expectations of improved work processes. Machines that work faster and more accurate than humans should take over repetitive manual tasks and deliver faster, cheaper and with higher quality. There is even hope that radically new workflows can emerge from novel use of artificial intelligence. SIRIUS research on asset modelling targets the key barrier for achieving these goals: the fragmentation of facility asset information.

The research aims at bridging the system of system based concepts that underlie the ISO/IEC81346 standard with the ontology based classification approach of ISO 15926. This is the basic appoach of the READI Information Modelling Framwork, which is based on four guiding principles:

- The (Oil & Gas) disciplines should own the role of specifying of the asset information model. This calls for a simple and standards-based method that discipline experts can understand, learn and apply with only little help from experts on the method.
- The method should give incremental value. The baseline is that the landscape of asset information is extremely fragmented. The method should be able structure fragments of information in isolation and to incrementally combine such fragments into more comprehensive models that give added value.
- Reusability and interoperability should be achieved through shared resources and open protocols. This means that the method should promote application-independent formats for library resources, and representation of specifications on open formats.
- The formalization of the method should use logic-based rigorous semantics. This level of precision enables the use of reasoning techniques to achieve a high level of automation and advanced tool support, thereby increasing the quality and value of the model.

Our research to support asset modelling is in particular applying methods from the ontology engineering, semantic integration and industrial digital transformation programs. Researchers from these programs have, in collaboration with partner representatives, been active in the development of the READI Information Modelling Framework. You can download information about this work from the READI web page **https://readi-jip.org/asset-information-modelling-framework/.**


CONTRACTOR
SUPPLIER
OPERATOR

Geological Assistant

Ambitions: 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.

SIRIUS Researchers: 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 geology.

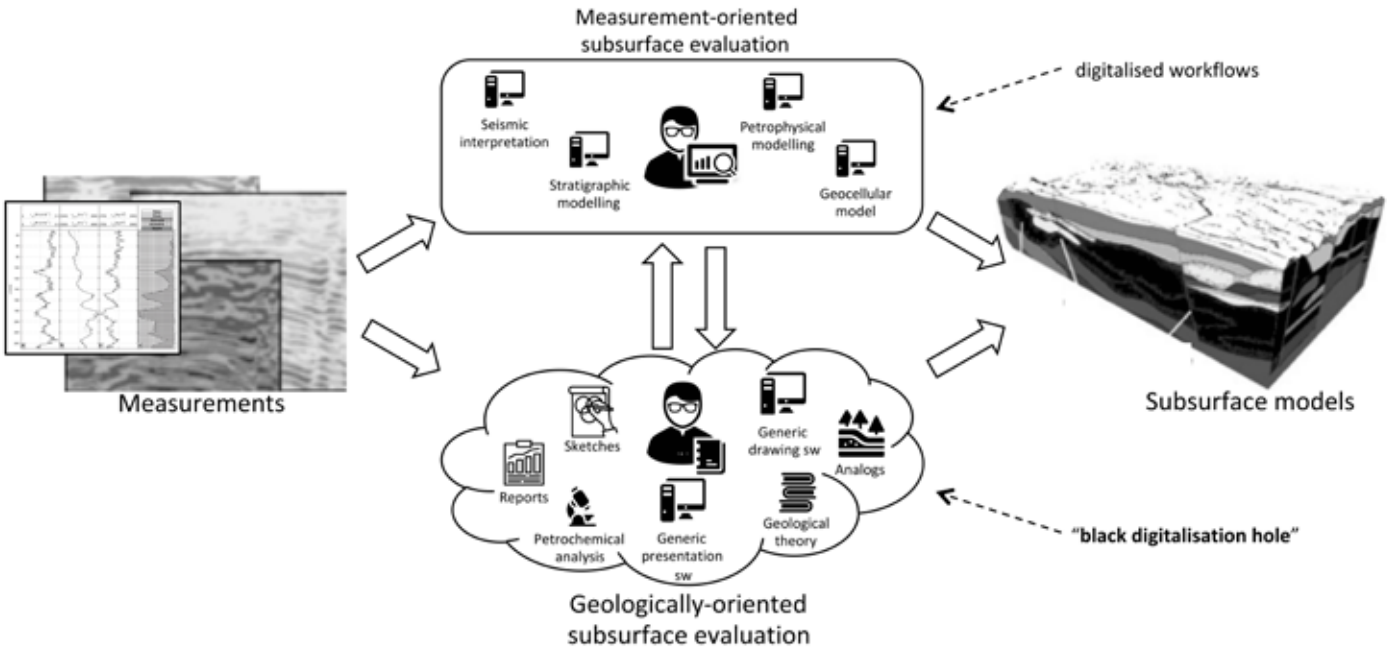
Partners: Equinor and Schlumberger

Timeline: Started- Nov 2016, Finishing- Nov 2023

Publications:

- Best Paper award at the 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.
- Book chapter: "Subsurface Evaluation through Multi-ScenarioReasoning" A detailed chapter is written for a book planned to be published by Springer in 2021 with the title "Interactive Data Exploration of the Subsurface - Realtime data processing and visualization techniques of the solid earth." This chapter is authored by Ingrid Chieh Yu, Irina Pene, Crystal Chang Din, Leif Harald Karlsen, Chi Mai Nguyen, Oliver Stahl, Adnan Latif.
- Demo at SIS globale forum GeoAssistant methodology (and tool) was demonstrated at the Schlumberger's SIS global forum

Recent advancements in computation, network and storage have led to numerous opportunities to improve the subsurface evaluation workflows. Further, the volatility and uncertainty in the oil and gas industry have forced exploration and production companies to find improved and cost-effective solutions by automating the subsurface evaluation workflows.



When it comes to digitalization, traditionally, the focus has been on purely data-driven workflows. Although geological reasoning is the most crucial factor that defines exploration success, little attention has been given to exploit digitalization opportunities in reasoning-based evaluation. In geological reasoning workflows, explorationist still rely on ad hoc manual work practices and tools and use pen and paper along with computer drawing and presentation tools to develop and communicate multiple hypothetical geological scenarios of the prospect. This leaves them with little to no efficient means to make the fullest use of state-of-the-art digital technologies to communicate and systematically compare and assess different hypothetical geological scenarios before deciding which scenario to pursue when assessing exploration prospects [Black Digitalization hole].

In GeoAssistant project, we experimented with logic-based techniques for subsurface modeling, with focus on how the inherent complexity in geology such as spatial and temporal aspects can be formally captured and reasoned about using the strength of different formalizations. In particular, we demonstrated the use of abstraction and how formal modelling gives a precise and human-readable representation of domain knowledge. Further, we developed a mechanism to bring together the various models in a novel tool-based approach that constructs multi-scenarios to support geologically oriented subsurface evaluation. In this work, we combined techniques from knowledge representation with formal methods (mathematical approaches that support the rigorous specification, design and verification of computer systems) to the exploration domain.

Adnan Latif

This logic-based technology enables explorationists to express interpretive uncertainty as discrete scenarios with branches of potential alternative interpretations. With this approach, common-sense explorationist domain knowledge and rules of thumb are explicitly represented in the tools together with collected O&G data.

Cross-Domain Applications

The goal of Cross-domain applications is to demonstrate technology transfer between the oil&gas sector to other important sectors. The primary focus has been on Personalized Medicine and Environmental Applications. This year, we wrapped up our work on two demonstration Projects: BIGMED and NIVA.

BIGMED

Precision Medicine is the customization of treatments and disease preventions based on the patient's individual characteristics, variations in genes, and lifestyle. BIGMED has been an ambitious innovation project that focused on the implementation of precision medicine in Norway. Through participation in BIGMED, the SIRIUS researchers in semantic integration and ontology engineering have gained experience in working with medical ontologies, including the Human Phenotype Ontology and Disease Ontology, and in addition, gained knowledge of the complexities of integrating healthcare data from Electronic Health Records. We completed our research in BIGMED this past year with a focus on the capture of clinical phenotypes to improve communication between healthcare personnel and lab technicians.

Laura Slaughter

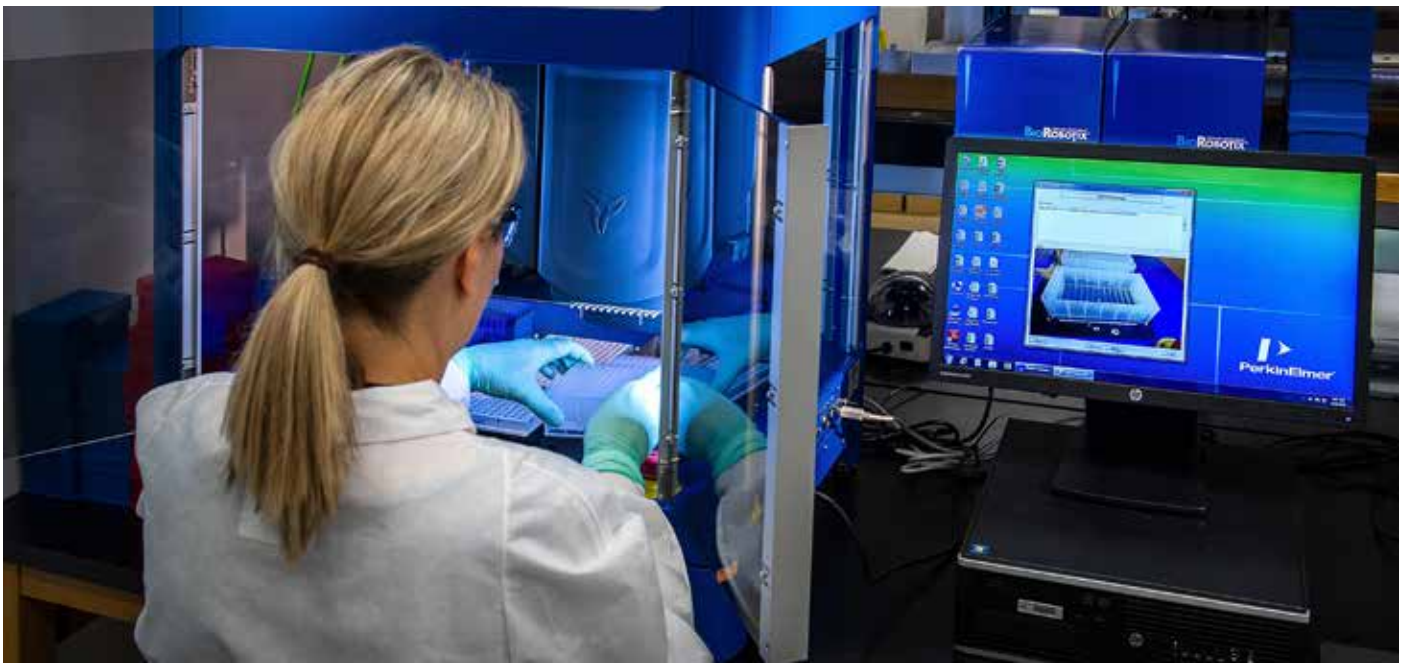
SIRIUS has also conducted scoping work within the ontology engineering research group to look at patterns in healthcare ontologies. This work involved healthcare registries and the OTTR team.

NIVA

SIRIUS collaborated with NIVA, The Norwegian Institute for Water Research, to explore how semantic integration and ontologies can improve the quality of decisions made in environmental toxicology. Scalable data access is a challenge in many environmental applications. SIRIUS PhD candidate, Erik Bryhn Myklebust worked at NIVA with the focus on machine learning on graph structures. He focused on environmental risk assessments. This year, he is putting the finishing touches on his dissertation "Ecotoxicological Effect Prediction using a Tailored Knowledge Graph".

The Norwegian Biodiversity Information Centre

SIRIUS researchers in the ontology engineering group are working with The Norwegian Biodiversity Information Centre. In this project, SIRIUS provides the Biodiversity Centre with an overview of the ontology engineering process. We are working with them to explore tools to be used to handle controlled vocabulary as well as domain ontologies related to the Traits Bank. We identify areas where ontology alignment and mapping tasks are necessary. The purpose is to facilitate data sharing and data integration for users of Biodiversity data.

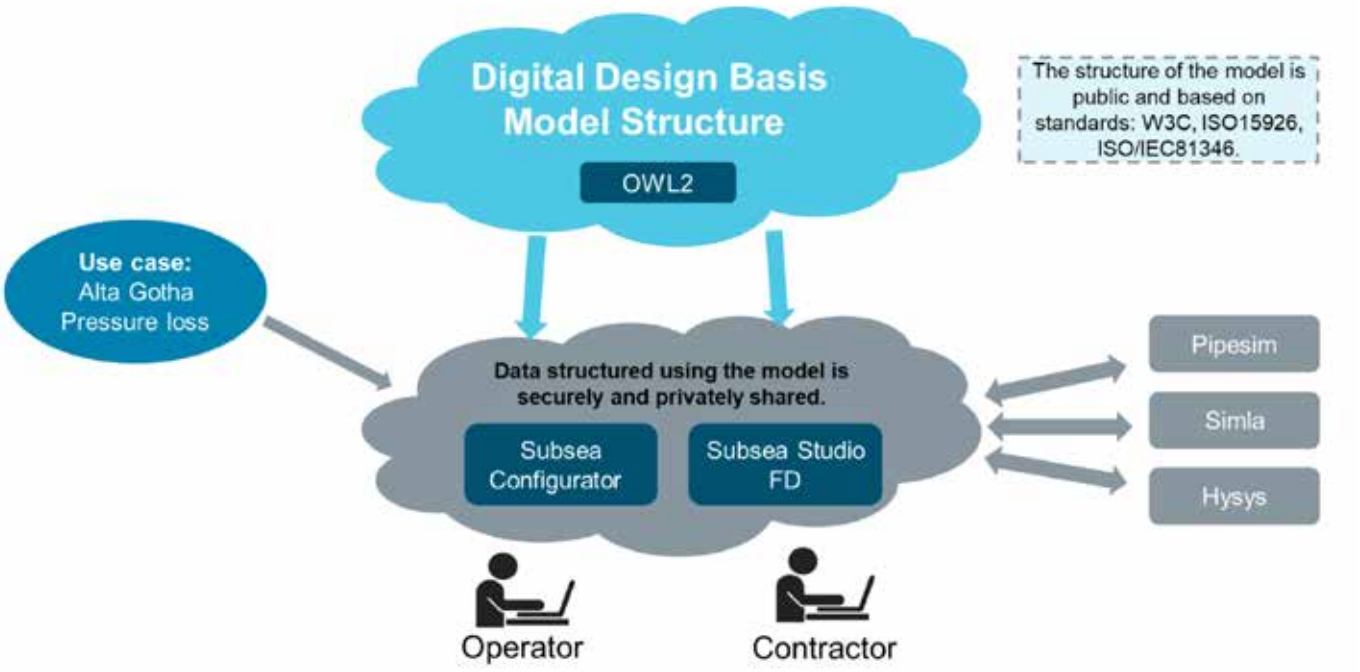


Digital Design Basis

In the 2017 annual report, we discussed how SIRIUS was working to support the findings of the KONKRAFT report on competiveness on the Norwegian continental shelf. This meant that SIRIUS workers have been part of a 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 natural conditions, 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 contained representatives of Aker Solutions, TechnipFMC, Aibel, Equinor, Lundin Energy and AkerBP. This resulted in a project that started in 2020 to build a prototype data model and demonstrate its use for sharing the design basis. The partners in this project were Lundin, AkerBP, TechnipFMC, Aker Solutions and Aibel. During the year, Aker Solution's information technology organization was reorganized. The project now also involves participants from Aize and Cognite.

The aim is to build a common data model that organizes the data that an operator shares with their suppliers in early-phase development. Instead of sending documents and spreadsheets, the operating company can prepare a set of data that uses the common model. This is then sent to or shared with the supplier. The data will be imported into the suppliers' own engineering tools and simulation programs.

The concept is shown in the following figure. The data model is written using the OWL2 standard for semantic modelling and provides a framework into which the operator can place their data. The modelling also uses the new Part 14 of the ISO15926 standard. The model describes the main systems in the proposed concept, using the ISO/IEC81346 system breakdown proposed in READI project. This breakdown enables the various requirements and pieces of information in a design basis document to be linked to systems in the concept.



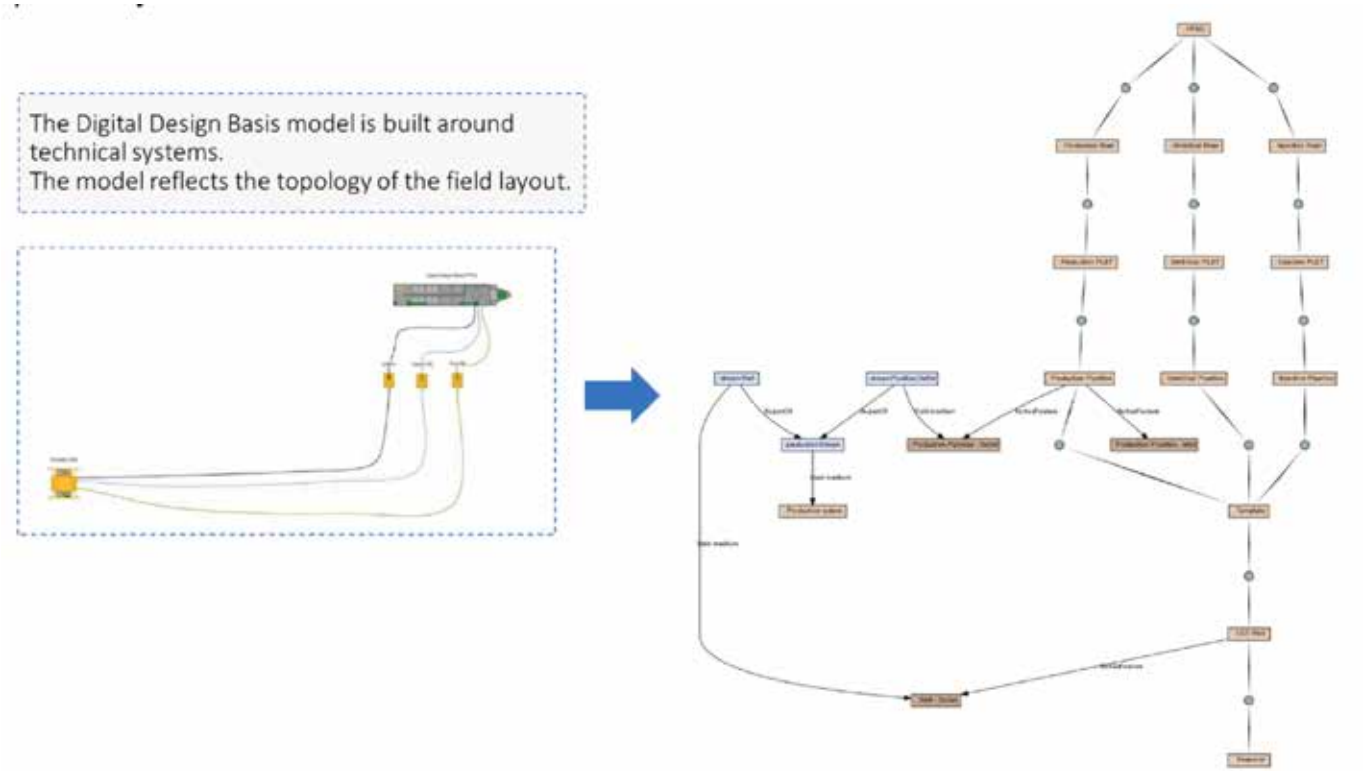
At the time of writing (April 2021) the project has built the first version of the common data model and used it to demonstrate a flow of data about Lundin's Alta Gotha field to Aker Solutions and TechnipFMC. Each of the suppliers were able to load this data into their internal engineering tool and use it to set up a calculation of pipeline pressure loss.

The work is organized as a SIRIUS innovation project. The project partners wanted a simple, easy-to-negotiate project model with an agreed management of intellectual property. SIRIUS' contract provided just such a model.

Results from this phase are promising. The model structure is built on existing industry standards and can be a starting point for standardizing data transmission at later phases of the development. Our ambition for further work is to deepen the model so that it supports FEED studies on the chosen concepts. In this way, we will be able to track requirements and design decisions from early phase, through FEED and engineering and eventually into operations. Here we will be able to exploit synergies with other projects in the Digital Twins and Digital Field Development demonstration, such as PeTWIN and DSYNE.

However, just as a tool for early phase development, the DDB offers savings in labour and improvements in the quality of data. The development uses SIRIUS skills and technology, such as the OTTR ontology templates to simplify the development of the model. The model is open sourced and will be maintained and made available through the SIRIUS Laboratory.

This project has validated the SIRIUS innovation model. The centre has provided a neutral place for collaboration between companies across the business ecosystems. Our contract models provided a quick and effective way of setting up formal collaboration. Finally, our tools and methods have been used in the work and the project has provided feedback to the SIRIUS laboratory.



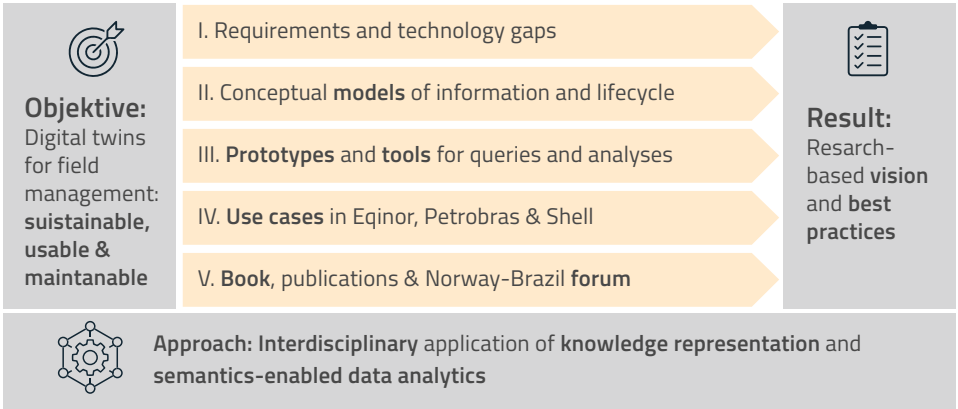
PeTWIN

Whole-field digital twins for production optimization and management.

In the 2019 report we were able to announce that we had been awarded large Norwegian-Brazilian research project. This is a knowledge-building project for the petroleum industry, 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.

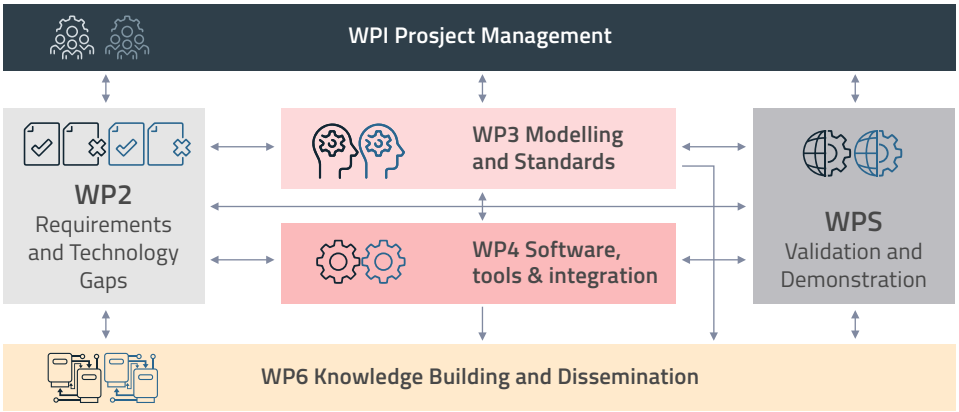
Contract negotiations were delayed by the COVID-19 pandemic. This meant that we were finally able to kick off the work in September 2020. The objectives of the project remain unchanged. Researchers in Norway and Brazil will be working together with Petrobras, Shell and Equinor to find ways of making digital twins more reliable, scalable, useful and cheaper.

The project, called PeTWIN, aims to provide a research-based foundation for digital twins used in the management of complex oilfields.



The project is working with digital twins that look at the entire field value chain, from the reservoir to off-loading. Digital twins that support this perspective need integration of data and applications from the sub-surface, petroleum technology and facilities domains. For this reason, we will be working with several use cases, some related to digital twins for subsurface and reservoir management, others looking at subsea production and facilities. We will be developing common, standards-based conceptual models that meet the needs of these use cases. Our ambition is to show that all three of the partner operator companies can use the common models successfully.

The project has recruited an interdisciplinary team of workers with skills in geosciences, semantic technology, formal methods, software engineering, cloud computing, requirements analysis and automation engineering. The work is still in its inception phase.


WPI Prosject Management
WP2
Requirements and Technology Gaps
WP3 Modelling and Standards
WP4 Software, tools & integration
WPS
Validation and Demonstration
WP6 Knowledge Building and Dissemination

PeTWIN

SIRIUS' Partners

SIRIUS draws together a consortium of leading industrial organisations across the oil & gas value chain, including operators (Equinor), service companies (Schlumberger, Aibel, Aker Solutions, Robert Bosch GmbH, DNV GL and TechnipFMC) and IT companies (Computas, Dolphin Interconnect Solutions, Evry, IBM, Kadme, Numascale, OSISoft and SAP).

The Spring 2020 General Assembly for SIRIUS was held as a videoconference in May 2020. This meeting approved a program in which Small and Medium Enterprises could become SIRIUS partners. As a result, SIRIUS welcomed Billington Process Technology, ONTOPIC, Oxford Semantic Technologies and Prediktor as partners from 1st July 2020.

Partnership for Small and Medium Enterprises (SMEs)

SIRIUS partners, with one exception, have been large firms with well-defined technology development and research organisations. They have the economic resources and capacity to fulfil the obligations of partnership in SIRIUS. Small and medium-sized enterprises (SMEs), however, lack this capacity. They are, however, valuable actors in the oil & gas supply chain. They also develop and sell tools that can contribute to the SIRIUS laboratory. By working with these companies, we can find new paths to commercialization for SIRIUS laboratory results.

During the last three years of the centre, we will be seeking to demonstrate how our research can be used practically in industrial companies. We also wish to demonstrate that our methods and tools can work with and enhance IT and business processes related to sub-surface data management, digital twins, digital field development, integrated digital planning and cross-domain applications.

Therefore, SIRIUS has opened for SMEs to join the centre with a financial contribution that is better fitted to their capacity. A SME can be a partner if they supply products and services that contribute to SIRIUS' research programs and demonstration initiatives. Their contribution to the centre is that they make their product and expertise available to the SIRIUS researchers and projects. In this way, we get access to the forefront of commercial tools. This means that our research can address industrial challenges, without the need to "reinvent the wheel" by building academic versions of commercially available functionality. Our SME partners, in return, can implement our research results to improve their products, thereby providing our researchers with a new path to innovation.

Two partners who deliver leading tools in semantic technology

SIRIUS believes that semantic technologies have the potential to revolutionize how we access and reason about industrial data. For this to happen, we need to have robust, scalable and commercially supported software. The OPTIQUE project demonstrated the power of ontology-based data access using pre-commercial software from universities, including the Free University of Bolzano in Italy and the University of Oxford in the United Kingdom. Since the end of the OPTIQUE project, both universities have created start-ups to commercialize and support the software that was used in OPTIQUE. Bringing these companies in as partners is good for both of them, SIRIUS research and other SIRIUS partners. The SIRIUS laboratory gets access to commercially supported software for use in laboratory and innovation projects. The start-ups get access to industrial

projects and can qualify their products against realistic projects with SIRIUS partners. Experience from the SIRIUS projects is used to improve the products and feed back into further research at the parent universities.

Oxford Semantic Technologies

Oxford Semantic Technologies is an Oxford University spin-out, established in 2017. Oxford Semantic Technologies develop RDFox, a highly scalable standards-based, in-memory RDF triple store and semantic reasoning engine. It is the first knowledge graph built from the ground up with semantic reasoning in mind. Many advanced features, including patented modern computing techniques, underpin RDFox's ability to deliver responses to complex queries on the fly. RDFox is unmatched in power, speed and reasoning capabilities.

Ontopic

ONTOPIC s.r.l. is a young innovative SME, which was founded in April 2019, as the first spin-off of the Free University of Bozen-Bolzano. Ontopic provides consultancy services and develops software solutions for data and information integration.

The key expertise of Ontopic is techniques and technologies for data access, management, and integration based on Virtual Knowledge Graphs (VKGs). Ontopic develops and constantly improves software tools that support the VKG technology, thus lowering the cost of typical data integration projects. In this way, companies and organizations can readily exploit the value of their data assets and make such data available for Business Intelligence and applications based on Machine Learning.

In the past, the founders of Ontopic collaborated with the University of Oslo, Equinor, and DNV GL within the large EU Project OPTIQUE, working on data integration in the oil and gas domain using the VKG technology. After OPTIQUE, this technology became mature and Ontopic was founded to commercialize it and promote its use in industry. SIRIUS, as a research center, has been tightly connected to OPTIQUE, and therefore it is a natural development for Ontopic to join it. Membership in SIRIUS gives us a unique opportunity to reach out to large industrial partners and, more in general, to several companies for which data is a critical asset that needs to be managed efficiently and effectively, and that could strongly profit from our technology.

Two partners at the forefront of digital twin technologies

The other two partners in SIRIUS are Norwegian specialist companies that both have a long history of delivering highly technical products and services to the oil and gas industry. Their experience and products are important in implementing digital twins for oil and gas and the process industries. The two companies work together already to provide commercial on-line systems and digital twins.

Billington Process Technology

Billington Process Technology (BPT) is a supplier of simulation and engineering design services and software. They are based in Bærum, just west of Oslo, and were founded in 1998. They deliver process engineering and simulation services to the oil and gas industry and deliver a set of process engineering applications that work together with commercial tools for process simulation. BPT is a leading centre of competence in process simulation and computer-aided process engineering. Dynamic and steady-state process simulation is a core component in a digital twin for oil and gas. This means that their software and knowledge will be a valuable contribution to SIRIUS' digital twin demonstration.

Prediktor

Prediktor AS, based in Fredrikstad, Norway, is an integrator of operational data from facilities. They have supplied a semantic OPC gateway to Equinor. The semantic node set for OPC UA that was developed has just been open sourced by Equinor.

Prediktor's Apis product is a component based real-time industrial software platform used in more than 1000 installations, mostly in mission critical areas within Oil & Gas, Maritime and Manufacturing industries. Apis supports a variety of communication protocols such as OPC, OPC UA, Modbus and WITS. Typical use of Apis is as:
Data Historian
Data Acquisition Hub
Protocol converter
IOT Gateway
OPC UA-based information modelling framework

Apis has extensive support for information models based on open standards from OPC Foundation. The overall objective of information modelling is to provide a standardized and unified way of accessing operational data from different assets. Major oil & gas operators are examples of customers using our products as an integral part of their digitization initiatives.

Prediktor is a leading implementer of semantic modelling of real-time data, especially using OPC UA. They will be contributing to work in the Ontology Engineering, Semantic Integration and Digital Twins programmes.

In SIRIUS the mentioned companies work with researchers from the University of Oslo, NTNU, the University of Oxford, Simula Research Laboratories and SINTEF.

Defended PhDs

David Tena Culcana
Consequence-Based Reasoning for the Description Logic SROIQ

ABSTRACT: Consequence-based (CB) reasoners combine ideas from resolution and (hyper) tableau calculi to solve the problem of ontology classification in Description Logics (DLs). Existing CB reasoners, however, are only capable of handling DLs without nominals (such as SRIQ), or DLs without disjunction (such as Horn-SROIQ). In this thesis, we present a novel CB calculus for concept subsumption in the DL ALCHOIQ+, a logic that extends ALC with role hierarchies, inverse roles, number restrictions, and nominals. To the best of our knowledge, ours is the first CB calculus for an NExpTime-complete DL. By using standard transformations, our calculus extends to SROIQ, which covers all of OWL 2 DL except for datatypes. A key feature of our calculus is its pay-as-you-go behaviour: the calculus is worst-case optimal for many well-known fragments of ALCHOIQ, including ELH, Horn-ALCHOIQ , and ALCHIQ.

Furthermore, the calculus is worst-case time exponential for the full logic ALCHOIQ+, except for cases where nominals, inverse roles, and number restrictions interact simultaneously, which are very rare in practice. Our calculus can also decide DL reasoning problems other than subsumption, such as ontology classification, instance retrieval, and realisation. We have implemented our calculus as an extension of the reasoner Sequoia, which previously supported ontology classification for SRIQ. Our implementation includes novel optimisation techniques that overcome some of the performance limitations affecting the previous version. We have carried out an empirical evaluation of our implementation which shows that the performance of Sequoia is competitive with other state-of-the-art systems. Our results also show that Sequoia nicely complements tableau-based reasoners, as it can more easily classify some ontologies. Thus, the calculus and implementation presented in this thesis provide an important addition to the repertoire of reasoning techniques and practical systems for expressive DLs.

Pan Hu
Efficient Computation and Maintenance of Datalog Materialisations

ABSTRACT: Datalog is a prominent knowledge representation language whose popularity is mainly due to its ability to express recursive definitions. Datalog applications typically require efficient reasoning over datalog programs and facts. To support this, datalog systems often materialise all consequences of a datalog program so that all queries can later be evaluated directly over the materialisation. This style of reasoning is usually complemented with an incremental maintenance algorithm that updates the materialisation whenever the input facts change. Existing solutions to the incremental maintenance problem either handle only nonrecursive rules or contain steps that evaluate rules "backwards" by matching their heads to a fact and evaluating the partially instantiated rule bodies as queries.

We show that this can be a considerable source of overhead even on very small updates, and we present two hybrid approaches that reduce or even eliminate "backward" rule evaluation while still handling arbitrary datalog programs. Moreover, we observe that for both materialisation and incremental maintenance, certain (combinations of) rules can be handled much more efficiently using custom algorithms. To integrate such algorithms into a general reasoning approach that can handle arbitrary rules, we propose a modular framework for computing and maintaining a materialisation. We split a datalog program into modules that can be handled using specialised algorithms, and we handle the remaining rules using the seminaive evaluation strategy as in most existing solutions.

We also present two algorithms for computing the transitive and the symmetric–transitive closure of a relation that can be used within our framework. Finally, we show empirically that the proposed solutions are usually significantly faster than existing approaches, sometimes by orders of magnitude.

Farhad Nooralahzadeh

Low-Resource Adaptation of Neural NLP Models

MAIN RESEARCH FINDINGS: Real-world applications of natural language processing (NLP) are challenging. NLP models rely heavily on supervised machine learning and require large amounts of annotated data. These resources are often based on language data available in large quantities, such as English newswire. However, in real-world applications of NLP, the textual resources vary across several dimensions, such as language, dialect, topic, and genre. It is challenging to find annotated data of sufficient amount and quality.

The objective of this thesis is to investigate methods for dealing with such low-resource scenarios in information extraction and natural language understanding. To this end, we study distant supervision and sequential transfer learning in various low-resource settings. We develop and adapt neural NLP models to explore a number of research questions concerning NLP tasks with minimal or no training data.

Vidar Klungre

Adaptive Query Extension Suggestions for Ontology-Based Visual Query Systems

MAIN RESEARCH FINDINGS: Many modern search systems allow users to narrow down their search by applying one or more search filters to the available data. This works well until the combination of filters becomes too restrictive, and no results remain. When this happens, the user needs to backtrack and remove the filter they just added, which is very counter-productive. In order to prevent this from happening, the search system should detect and disable too restrictive filters before the user activates any of them.

Current solutions are only able to do this in some special situations, but the solution presented in this thesis is more flexible and works in any situation.

International Activity and Dissemination

Remote International Activity in the Pandemic

SIRIUS' international activity was reduced by the pandemic in 2020. All communication moved to videoconferences and virtual events. We also found that projects were delayed through the effects of the pandemic on decision making in our partner countries. All exchanges and staff visits have been postponed. At the time of writing, early 2021, we do not expect normal international collaboration to resume before the first quarter of 2022.

Brazil

Despite the pandemic, SIRIUS' collaboration with leading industrial and academic partners deepened in 2020. Our collaboration with the Department of Informatics at the Federal University of Rio Grande do Sul (UFRGS) in Porto Allegre continued and was cemented through the start-up of two collaborative projects.

The PeTWIN project, which had been awarded in 2019, was finally able to start in the fourth quarter of 2020. This project brings together our researchers with a large and skilled team at UFRGS. The project kick-off was held in October 2020.

We also started DSYNE, which is an INTPART-funded network for research and teaching collaboration. This project brings together our collaborators at UFRGS and workers at the Federal University of Espirito Santo. The

network also includes the University of South-Eastern Norway and two universities in the United States: Stevens Institute of Technology and the University of Houston. The aim of this network is build a research and teaching agenda around the use of digitalized requirements in the energy, oil & gas and process industries.

As in previous years, David Cameron from SIRIUS was project manager for the annual November Conference on Norwegian-Brazilian Energy Research. This year, the conference was purely virtual, with a plenary session held on 9th November 2020 and parallel sessions held between the end of October and beginning of December. We also worked with NTNU to organize the parallel session on Digitalization and Autonomy on 10th November. The plenary session was addressed by Tina Bru, the Norwegian Minister for Petroleum and Energy, and senior representatives for the Brazilian government, Equinor and Petrobras. This year, the conference moved once more from a petroleum theme to a broader, energy transformation theme. Feedback from the conference was positive, and we were pleased that we could continue our collaboration using this virtual format. As in previous years, the conference was expertly organized by the staff of Innovation Norway in Rio de Janeiro.

European Projects

The MELODIC project (**https://melodic.cloud**, Grant Agreement 731664) concluded successfully in January 2020 and delivered a product, namely multi-cloud optimization platform and automatic deployment solution: This allows deployment of an application to different cloud providers without changing configuration. The system uses a fully cloud agnostic approach. The selection of cloud providers and cloud resources is fully optimized with regard to prices, performance, reliability and other factors. After the initial deployment the application is continuously monitored to check the fulfilment of business goals fulfilment and reconfigure if needed to maintain optimal operation.

The MELODIC project was followed up by the MORPHEMIC project (**https://www.morphemic.cloud**, Grant Agreement 871643) this extends MELODIC to support live application reconfiguration. This project started in January 2020 and is planned to end in 2022. It is coordinated by the University of Oslo, with Geir Horn as coordinator.

SIRIUS also participated in two successful proposals that were funded in 2020, where we will be developing and applying our expertise in semantic technologies. The first of these is the ONTOCOMMONS co-ordination and support activity (Grant Agreement 958371, **https://ontocommons.eu**). This project aims to lay the foundations for standardising the format of data and ensure interoperability so that it will be easier to use and distribute data in the manufacturing sector and the material science field.

The second project was Eur3ka: EUropean Vital Medical Supplies and Equipment Resilient and Reliable Repurposing Manufacturing as a Service NetworK for Fast PAndemic Reaction (Grant Number 101016175). Here we will be applying our semantic technologies to build a digital framework that will ensure that we can respond to future pandemics faster and better by re-purposing manufacturing to make necessary equipment, like masks and ventilators. This project started in December 2020.

European Private-Public Partnerships

SIRIUS has continued to lead the University of Oslo's participation in the A.SPIRE and BDVA private-public partnerships. Effort in 2020 in these partnerships was devoted to the transition from Horizon 2020 to Horizon Europe. This has resulted in a rebranding of both of these partnerships. BDVA has become DAIRO: Data, Artificial Intelligence and Robotics, while A.SPIRE has built its Processes for the Planet agenda.

SIRIUS workers have contributed to the A.SPIRE working group of Digitalization and BDVA's working group on Smart Manufacturing.

Nordic Interoperability Cooperation

SIRIUS has become involved with an informal collaboration between the Nordic countries around the topic of interoperability of facility and engineering data. This collaboration brings together researchers and companies in Finland, Sweden and now, Norway. It is built around two organizations for industrial interoperability: THTH in Finland and SEIIA in Sweden. The purpose of the work is to build consortia for externally funded projects and build up adoption of interoperability standards in Nordic industries.

Dissemination

SIRIUS' dissemination activities also moved into the Zoom-sphere in 2020. As a result of the pandemic, we needed to provide regular events that brought researchers and partners together. Regular SIRIUS lunch seminars are organized by a committee of junior researchers. These have proved to be a valuable way of communicating our research to our industrial partners and the other researchers in the centre.

List of Staff

Role	Name	Affiliation
Adnan Latif	Senior Advisor	University of Oslo
Andreas Thune	PhD	Simula
Arild Waaler	Centre Director/Professor	University of Oslo
Baifan Zhou	Researcher	University of Oslo
Basil Ell	Researcher	University of Oslo
Boris Motik	Professor	University of Oxford
Chi Mai Nguyen	Postdoc	University of Oslo
Chinmayi Prabhu Prasad Baramashetru	PhD	University of Oslo
Crystal Chang Din	Researcher	University of Oslo
Dag Hovland	Researcher	University of Oslo
Daniel Bakkelund	PhD	University of Oslo
Daniel Lupp	Postdoc	University of Oslo
David Cameron	Centre Coordinator	University of Oslo
David Tena Cucala	PhD	University of Oxford
Dirk Hesse	Ass.Professor	University of Oslo
Eduard Kamburjan	Postdoc	University of Oslo
Egor Kostylev	Ass.Professor	University of Oslo
Einar Broch Johnsen	Professor	University of Oslo
Elena Parmiggiani	Ass.Professor	NTNU
Eric Monteiro	Professor	NTNU
Erik B. Myklebust	PhD	NIVA
Erik Hide Sæternes	PhD	Simula
Ernesto Jimenez-Ruiz	Researcher	University of Oslo
Evgeny Kharlamov	Ass.Professor	University of Oslo
Fabricio Rodrigues	PhD	UFRGS
Farhad Nooralahzadeh	PhD	University of Oslo
Federico Igne	PhD	University of Oxford
Geir Horn	Senior Advisor	University of Oslo
Gianluca Turin	PhD	University of Oslo
Ian Horrocks	Professor	University of Oxford

Role	Name	Affiliation
Inger-Lise Reang	Admin	University of Oslo
Ingrid Chieh Yu	Deputy Director/Ass.Professor	University of Oslo
Irina Pene	PhD	University of Oslo
Jan Tore Lønning	Professor	University of Oslo
Jens Otten	Postdoc	University of Oslo
Jiaoyan Chen	Senior Researcher	University of Oxford
Jieying Chen	Postdoc	University of Oslo
Ka I Pun	Ass.Professor	University of Oslo
Keith Lewis	Senior Advisor	University of Oslo
Laura Slaughter	Ass.Professor	University of Oslo
Leif Harald Karlsen	Senior Lecturer	University of Oslo
Marta Rozanska	PhD	University of Oslo
Martin G. Skjæveland	Researcher	University of Oslo
Martin Giese	Professor	University of Oslo
Maunya Doroudi Moghadam	Research Coordinator	University of Oslo
Michael Heeremans	Senior Engineer	University of Oslo
Mina Haghshenas	PhD	NTNU
Ole Magnus Holter	PhD	University of Oslo
Oliver Stahl	Technical Programmer	University of Oslo
Peyman Rasouli	PhD	University of Oslo
Ratan Bahadur Thapa	PhD	University of Oslo
Rustam Mehmandarov	PhD	University of Oslo
Rudolf Schlatte	Senior Researcher	University of Oslo
Shuwen Liu	PhD	University of Oxford
Sigurd Kittilsen	PhD	University of Oslo
Silvia Lizeth Tapia Tarifa	Researcher	University of Oslo
Stefano Germano	Researcher	University of Oxford
Summaya Mumtaz	PhD	University of Oslo
Temitope Ajileye	PhD	University of Oxford
Thomas Østerlie	Researcher	NTNU
Tor Skeie	Professor	Simula
Vidar Norstein Klungre	Postdoc	University of Oslo
Xing Cai	Professor	Simula
Yuanwei Qu	PhD	University of Oslo

Publications in 2020

Cameron, David B. Digitale tvillinger. Hva er de? Hvordan kan de brukes?. NTVA Stavanger Fagseminar; 2020-03-11 UiO

Chen, Jiaoyan; Chen, Xi; Horrocks, Ian; Myklebust, Erik Bryhn; Jimenez-Ruiz, Ernesto. Correcting Knowledge Base Assertions. I: WWW '20: Proceedings of The Web Conference 2020. Association for Computing Machinery (ACM) 2020 ISBN 978-1-4503-7023-3. s.1537-1547 NIVA UiO

Cheng, Gong; Gunaratna, Kalpa; Kharlamov, Evgeny. Entity Summarization in Knowledge Graphs: Algorithms, Evaluation, and Applications. I: WWW '20: Proceedings of The Web Conference 2020. Association for Computing Machinery (ACM) 2020 ISBN 978-1-4503-7023-3. s.301-302 UiO

Cutrona, Vincenzo; Bianchi, Federico; Jimenez-Ruiz, Ernesto; Palmonari, Matteo. Tough Tables: Carefully Evaluating Entity Linking for Tabular Data. Lecture Notes in Computer Science (LNCS) 2020 ;Volum 12507. s.328-343 UiO

de Boer, Frank; Bonsangue, Marcello; Johnsen, Einar Broch; Pun, Ka I; Tapia Tarifa, Silvia Lizeth; Tveito, Lars. SymPaths: Symbolic Execution Meets Partial Order Reduction. Lecture Notes in Computer Science (LNCS) 2020 ;Volum 12345. s.313-338 HVL UiO

de Boer, Frank; Johnsen, Einar Broch; Pun, Ka I; Tapia Tarifa, Silvia Lizeth. From SOS to asynchronously communicating actors. Lecture Notes in Computer Science (LNCS) 2020 ;Volum 12226. s.269-275 HVL UiO

Faria, Daniel; Ferrara, Alfio; Jimenez-Ruiz, Ernesto; Montanelli, Stefano; Pesquita, Catia. Crowd-assessing quality in uncertain data linking datasets. Knowledge engineering review (Print) 2020 ;Volum 35. UiO

Forssell, Henrik; Kharlamov, Evgeny; Thorstensen, Evgenij. On equivalence and cores for incomplete databases in open and closed worlds. Leibniz International Proceedings in Informatics 2020 ;Volum 155. s.1-21 USN UiO

Ghzouli, Razan; Berger, Thorsten; Johnsen, Einar Broch; Dragule, Swaib; Wsowski, Andrzej. Behavior Trees in Action: A Study of Robotics Applications. I: Proceedings of the ACM SIGPLAN International Conference on Software Language Engineering (SLE 2020). Association for Computing Machinery (ACM) 2020 ISBN 978-1-4503-8176-5. s.196-209 UiO

Gkolfi, Anastasia; Johnsen, Einar Broch; Kristensen, Lars Michael; Yu, Ingrid Chieh. Model checking starvation for resource-aware active objects with coloured petri nets. CEUR Workshop Proceedings 2020 ;Volum 2651. s.68-85 HVL UiO

Haghshenas, Mina; Østerlie, Thomas. Coordinating innovation in digital infrastructure: The case of transforming offshore project delivery. I: Digital Business Transformation. 2020 ISBN 978-3-030-47355-6. NTNU

Haghshenas, Mina; Østerlie, Thomas. Navigating towards a digital ecosystem: The case study of offshore infrastructure industry. I: Proceedings of the 11th Scandinavian Conference on Information Systems (SCIS2020). Association for Information Systems (AIS) 2020 ISBN 978-91-88947-82-6. NTNU

Henrio, Ludovic; Johnsen, Einar Broch; Pun, Ka I. Active Objects with Deterministic Behavior. Lecture Notes in Computer Science (LNCS) 2020 ;Volum 12546. s.181-198 HVL UiO

Hoff, Adrian; Nieke, Michael; Seidl, Christoph; Sæther, Eirik; Motzfeldt, Ida; Din, Crystal Chang; Yu, Ingrid Chieh; Schaefer, Ina. Consistency-Preserving Evolution Planning on Feature Models. I: SPLC'20: 24th ACM International Systems and Software Product Line Conference, Montreal, Quebec, Canada, October 19-23, 2020, Volume A. Association for Computing Machinery (ACM) 2020 ISBN 978-1-4503- 7569-6. s.8:1-8:12 UiO

Jimenez-Ruiz, Ernesto; Agibetov, Asan; Chen, Jiaoyan; Samwald, Matthias; Cross, Valerie. Dividing the ontology alignment task with semantic embeddings and logic-based modules. Frontiers in Artificial Intelligence and Applications 2020 ;Volum 325. s.784-791 UiO

Jimenez-Ruiz, Ernesto; Hassanzadeh, Oktie; Efthymiou, Vasilis; Chen, Jiaoyan; Srinivas, Kavitha. SemTab 2019: Resources to Benchmark Tabular Data to Knowledge Graph Matching Systems. Lecture Notes in Computer Science (LNCS) 2020 ;Volum 12123. s.514-530 UiO

Johnsen, Einar Broch; Steffen, Martin; Stumpf, Johanna Beate. Assumption-Commitment Types for Resource Management in Virtually Timed Ambients. Lecture Notes in Computer Science (LNCS) 2020 ;Volum 12476. s.103-121 UiO

Kalayci, Elem Guzel; Grangel-González, Irlán; Loesch, Felix; Xiao, Guohui; Mehdi, Anees; Kharlamov, Evgeny; Calvanese, Diego. Serving Bosch Production Data as Virtual KGs. CEUR Workshop Proceedings 2020 ;Volum 2721.(589) UiO

Kalayci, Elem Güzel; Grangel-González,, Irlán; Lösch, Felix; Xiao, Guohui; ul-Mehdi, Anees; Kharlamov, Evgeny; Calvanese, Diego. Semantic Integration of Bosch Manufacturing Data Using Virtual Knowledge Graphs. Lecture Notes in Computer Science (LNCS) 2020 ;Volum 12507. s.464-481 UiO

Li, Junyou; Cheng, Gong; Liu, Qingxia; Zhang, Wen; Kharlamov, Evgeny; Gunaratna, Kalpa; Chen, Huajun. Neural Entity Summarization with Joint Encoding and Weak Supervision. I: Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020. 2020 ISBN 978-0-9992411-6-5. s.1644-1650 UiO

Li, Shuxing; Huang, Zixian; Cheng, Gong; Kharlamov, Evgeny; Gunaratna, Kalpa. Enriching Documents with Compact, Representative, Relevant Knowledge Graphs. I: Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020. 2020 ISBN 978-0-9992411-6-5. UiO

Lienhardt, Michael; Damiani, Ferruccio; Johnsen, Einar Broch; Mauro, Jacopo. Lazy product discovery in huge configuration spaces. Proceedings - International Conference on Software Engineering 2020 s.1509-1521 UiO

Lin, Jia-Chun; Lee, Ming-Chang; Yu, Ingrid Chieh; Johnsen, Einar Broch. A Configurable and Executable Model of Spark Streaming on Apache YARN. International Journal of Grid and Utility Computing (IJGUC) 2020 ;Volum 11.(2) s.185-195 NTNU UiO

Liu, Qingxia; Chen, Yue; Cheng, Gong; Kharlamov, Evgeny; Li, Junyou; Qu, Yuzhong. Entity Summarization with User Feedback. Lecture Notes in Computer Science (LNCS) 2020 s.376-392 UiO

Lupp, Daniel P.; Hodkiewicz, Melinda; Skjæveland, Martin G. Template Libraries for Industrial Asset Maintenance: A Methodology for Scalable and Maintainable Ontologies. CEUR Workshop Proceedings 2020 ;Volum 2757. s.49-64 UiO

Mezzina, Claudio Antares; Schlatte, Rudolf; Glück, Robert; Haulund, Tue; Hoey, James; Cservenka, Martin Holm; Lanese, Ivan; Mogensen, Torben Æ.; Siljak, Harun; Schultz, Ulrik Pagh; Ulidowski, Irek. Software and Reversible Systems: A Survey of Recent Activities. I: Reversible Computation: Extending Horizons of Computing - Selected Results of the COST Action IC1405. Springer 2020 ISBN 978-3-030-47360-0. s.41-59 UiO

Mumtaz, Summaya; Pene, Irina; Latif, Adnan; Giese, Martin. Data-based support for petroleum prospect evaluation. Earth Science Informatics 2020 UiO

Oliver, Benjamin; Otten, Jens. Equality Preprocessing in Connection Calculi. CEUR Workshop Proceedings 2020 ;Volum 2752. s.76-92 UiO

Rasouli, Peyman; Yu, Ingrid Chieh. EXPLAN: Explaining Black-box Classifiers using Adaptive Neighborhood Generation. Proceedings of ... International Joint Conference on Neural Networks 2020 UiO

Shi, Yuxuan; Cheng, Gong; Kharlamov, Evgeny. Keyword Search over Knowledge Graphs via Static and Dynamic Hub Labelings. I: WWW '20: Proceedings of The Web Conference 2020. Association for Computing Machinery (ACM) 2020 ISBN 978-1-4503-7023-3. s.235-245 UiO

Svetashova, Yulia; Zhou, Baifan; Pychynski, Tim; Schmidt, Stefanie; Sure-Vetter, York; Mikut, Ralf; Kharlamov, Evgeny. Ontology-Enhanced Machine Learning: A Bosch Use Case of Welding Quality Monitoring. Lecture Notes in Computer Science (LNCS) 2020 s.531-550 UiO

Svetashova, Yulia; Zhou, Baifan; Schmid, Stefan; Pychynski, Tim; Kharlamov, Evgeny. SemML: Reusable ML for Condition Monitoring in Discrete Manufacturing. CEUR Workshop Proceedings 2020 ;Volum 2721. s.214-218 UiO

Tveito, Lars; Johnsen, Einar Broch; Schlatte, Rudolf. Global Reproducibility Through Local Control for Distributed Active Objects. I: Fundamental Approaches to Software Engineering 23rd International Conference, FASE 2020, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2020, Dublin, Ireland, April 25–30, 2020, Proceedings. Springer 2020 ISBN 978-3-030-45233-9. s.140-160 UiO

Yu, Ingrid Chieh; Cameron, David B.; Vølstad, Ann; Larsen, Åshild Hanne. Industrial mentoring for junior researchers: An enabler for personal development and innovation. 2020 University-Industry Interaction Conference; 2020-06-08 - 2020-06-11 UiO

Zhou, Baifan; Svetashova, Yulia; Pychynski, Tim; Baimuratov, Ildar; Soylu, Ahmet; Kharlamov, Evgeny. SemFE: Facilitating ML Pipeline Development with Semantics. I: CIKM 20 : The 29th ACM international conference on information and knowledge management : Virtual Event Ireland, October 19-23, 2020.. ACM Publications 2020 ISBN 9781450368599. s.3489-3492 NTNU SINTEF UiO

Zhou, Baifan; Svetashova, Yulia; Pychynski, Tim; Kharlamov, Evgeny. Semantic ML for Manufacturing Monitoring at Bosch. CEUR Workshop Proceedings 2020 ;Volum 2721. UiO

Annual Accounts

Costs

All figures in 1000 NOK	2015	2016	2017	2018	2019	2020
Personnel and indirect costs	539	5188	10087	12378	20922	21783
Purchase of research services	-	600	2113	3415	6166	6356
Equipment	-	31	122	-	-	-
Other operational costs	62	9505	10424	14131	17202	17124
Total Sum	**601**	**15324**	**22746**	**29924**	**44290**	**45263**

Funding

All figures in 1000 NOK	2015	2016	2017	2018	2019	2020
Research Council	-	4168	8180	12430	16961	17805
University of Oslo	601	2141	5593	6387	7721	11758
Public partners	-	144	813	107	1137	1350
Private partners	-	8510	7910	11000	17700	14350
International partners	-	361	250	-	771	-
Total Sum	**601**	**15324**	**22746**	**29924**	**44290**	**45263**

Photo credits
Aibel
David Costa
iStock
Shutterstock
Statoil/Equinor
Unsplash

Editor:
Maunya Doroudi Moghadam

Layout
FÆRD AS

SIRIUS

ANNUAL REPORT 2020

Sirius-labs.no