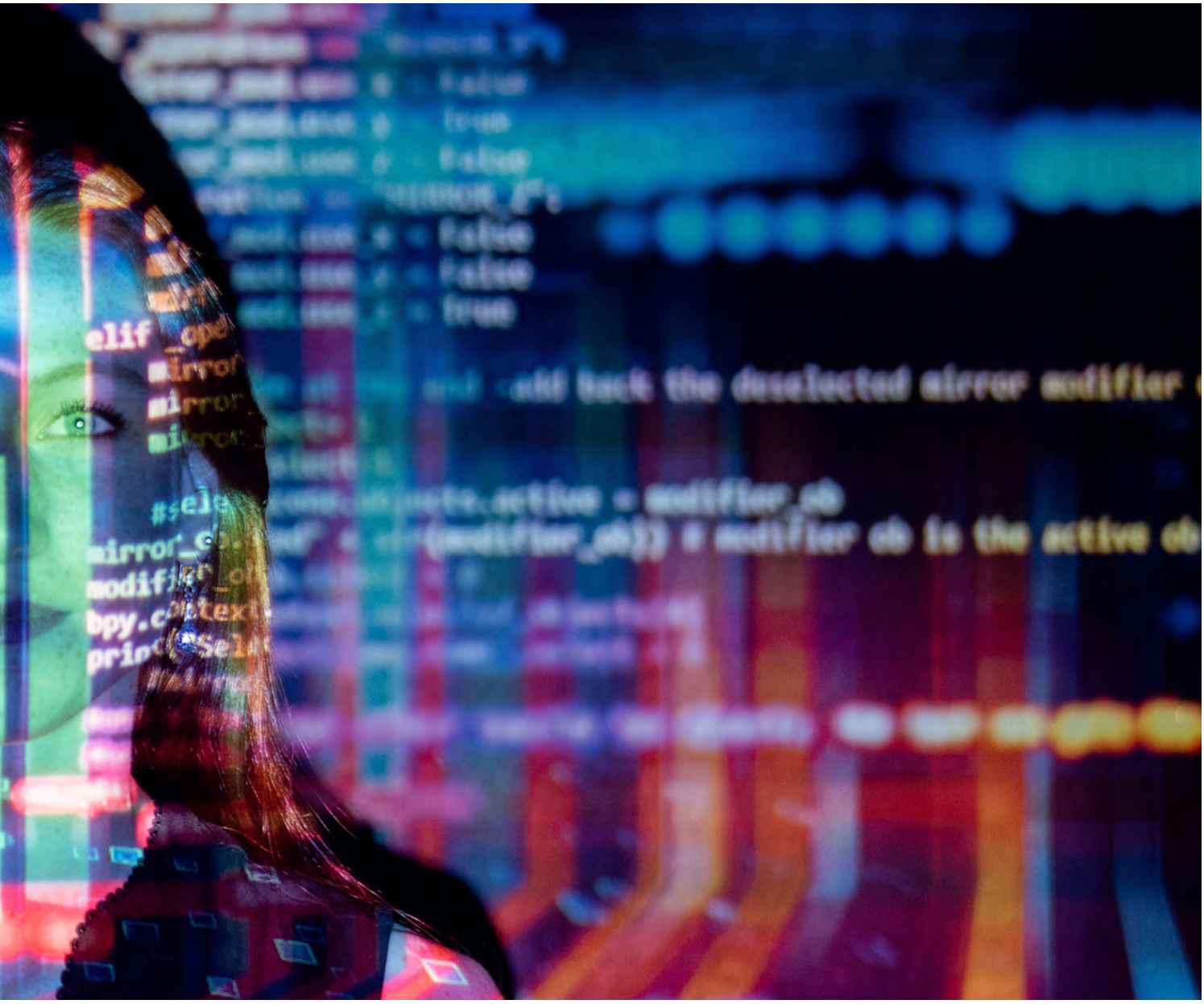
SIRIUS ANNUAL REPORT 2022



Centre for Research-based Innovation
The Research Council of Norway

UNIVERSITY OF OSLO

Contents

Executive summary and the way forward

I am pleased to present the 2022 annual report for SIRIUS.

Arild Waaler

As we enter our last year of operation, SIRIUS research projects enter their last phase. We can look back at a year where we begin to see real impact of our research. We celebrated the graduation of 3 PhD students in 2022. Their achievements reflect our commitment to fostering the next generation of researchers and equipping them with the skills and knowledge to drive research and innovation in their respective fields.

SIRIUS researchers were instrumental in preparing a proposal for Integreat, a 10 years Centre of Excellence aiming for inventing novel machine learning driven by knowledge. We received a positive funding decision from the Research Council of Norway in 2022. Integreat will start its operation back to back with the ending of SIRIUS, thus securing a smooth transition of the SIRIUS research program on data science into a new centre environment.

The method for asset information modelling, IMF, developed by SIRIUS partners and researchers, is starting to create significant impact. IMF will be published as a DNV Recommended Practice in 2023 and steps towards ISO standardization are made.

A primary focuses this year has been disseminating our research findings and promoting further collaboration across academia and industry. We organized a successful scientific event in the spring featuring keynote speakers, plenary sessions, and detailed technical workshops. This event attracted more than 100 attendees from academia and industry. Further, we also established dissemination webpages from our research programs and beacon projects to document our results. Our portfolio of European projects is also steadily increasing.

These results demonstrate our center's dedication to providing innovative solutions that address real-world challenges and contribute to the broader scientific community. Our innovation work and methodology for defining innovation projects have received widespread recognition and adoption, with the dScience centre at UiO adopting our methods as a model for its innovation work.

In addition, SIRIUS has organized several events with industry to review the produced research and explore opportunities for continued collaboration, even after SIRIUS ends. These events included a joint academic and industry workshop in November, which received a positive response and resulted in many industrial partners committing to work together beyond the center's formal conclusion.

We have also been working on developing assets with long-term value, such as the SIRIUS Subsurface Lab, where we used a big open dataset from Equinor to create a sand-box for experimenting with subsurface data. Assets like this will continue to provide value to future researchers and industry partners.

To evaluate and document our learning from running SIRIUS, we have conducted a detailed survey involving industrial participants and academic researchers. By sharing these experiences and insights, we hope to contribute to the continuous improvement and success of similar initiatives in the scientific community.

As we approach 2023, the final year for SIRIUS, our dedication to research and innovation remains steadfast. We will continue to build upon the strong foundation we have established over the years, with a special focus on projects that can lead to further academic-industry collaboration. Our commitment to fostering partnerships and driving innovation will persist beyond the lifetime of SIRIUS as an SFI as we work together to create a lasting impact in the scientific community and beyond.

In conclusion, I want to thank all SIRIUS researchers, partners and management team for their support and collaboration throughout the year. We look forward to building on these achievements and continuing to drive innovation and excellence in the final year of SIRIUS.


Pitch by Jieying Chen


Value of research @ SIRIUS
SIRIUS

Highlights of 2022

In 2022, SIRIUS remained committed to collaborating with the centre's research programs and beacon projects together with our partners. Our focus has been on strengthening our teams of researchers to enhance the organization's capability and robustness in delivering high-quality research. Our fundamental research has progressed in line with the plan, and three PhDs have graduated from the centre this year.

Highlights of the centre's activities:

SIRIUS researchers were actively contributing to the establishment of Integreat, a new Centre of Excellence that will start its operation in 2023. We received the positive funding decision from the Research Council of Norway in 2022. Targeting knowledge driven machine learning, Integreat will continue the line of research of our research program on domain-adapted data science.

The center has made progress in its SME program designed to support small and medium-sized businesses. In 2022, Bouvet AS joined our list of partners, further strengthening our program.

In 2022, we successfully implemented OBDA-based database access to Equinor's Volvo data set, which serves as a valuable resource for developing our research methods. It can also be utilized as a collaborative tool for research that requires subsurface data, making it an important resource.

UiO (via SIRIUS) has been an active partner in the Open Group's OSDU, an international collaboration aimed at creating a unified platform for the subsurface data used by oil companies. As part of our involvement in this initiative, we are currently working on extending ontology-based data access (OBDA) as a plug-in to the OSDU platform. This will further enhance the platform's capabilities, enabling efficient access and utilization of subsurface data.

The PeTWIN project is now fully staffed and actively contributing to SIRIUS' ambitions in digital twins. This project is a collaboration between UiO, UFRGS, Shell, Petrobras, and Equinor and is financed by PETROMAKS and FINEP. Through PeTWIN, we aim to develop new tools and methods for the digital twins in the oil and gas industry. The expertise and resources of our partners, combined with our research capabilities, have resulted in significant progress towards achieving this goal. We remain committed to this innovative project and look forward to its continued success.

The second phase of the Joint Industry Project (JIP), READI, was completed in 2021. This project, led by DNV, in collaboration with Equinor, Aibel, Aker Solutions, TechnipFMC, and Computas. As a partner in the READI project, SIRIUS researchers developed methods that have been instrumental in its success. These methods have since been incorporated into other key projects, including the IMF and NOAKA projects.

SIRIUS researchers have been actively contributing to the development of the Facility Asset Information Modeling Framework (IMF). This framework combines system-oriented techniques with semantic technology and methodology to model industrial facilities. IMF has been designed to take existing standards as its starting point, with the aim of becoming a standard in its own. In 2023, DNV plans to publish a Recommended Practice document for IMF. It is an ambition that IMF will later become a standard within ISO/IEC.

SIRIUS has actively participated in the development of ISO/TR 15926 Part 14, a top-level industry information ontology that works seamlessly with W3C-standardized semantic technology. In 2023, we will be participating in the working group that will elevate ISO/TR 15926 Part 14 from a technical report to a full-fledged industry standard.

SIRIUS has actively participated in the DEXPI project (Data EXchange in the Process Industry), a working group in ProcessNet, a joint initiative from DCHEMA and CDI - GVC. In 2022, our researchers contributed to the DEXPI project by writing the new DEXPI standard for modelling process design in the process industry.

Adnan Latif

SIRIUS is the coordinator of the INTPART project, DSYNE. This groundbreaking initiative brings together key industry players, including Aibel, DNV, Equinor, and TechnipFMC, and two universities in Brazil and two in the USA, to focus on requirements management in field development projects.

In 2022, we organized a workshop in Rio de Janeiro, bringing together participants from all partners to exchange knowledge and insights. The workshop provided an excellent platform for discussing and refining project objectives and developing strategies to achieve them.

Rodrigo Calhau from the Federal University of Espírito Santo has been working at UIO and USN for six months since September 2022, looking at semantic system modelling.

SIRIUS' innovation work and methodology for defining innovation have received widespread recognition and adoption. This has been well-documented and proven highly effective in driving research and Innovation projects. The dScience centre at UiO has adopted our methods as a model for its innovation work.

International collaboration

Highlights:

- Partnership in the centre of the University of Oxford and active participation by Ernesto Jiménez-Ruiz from the City University London.
- Continued participation of Birkbeck, University of London researchers in the work around ontology-based data access to the petroleum databases.
- Membership and participation in the EU's Big Data Value Association PPP and A.SPIRE PPP as part of participation in Horizon 2020.
- Collaboration with UFRGS in Porto Alegre, Brazil. This is part of a collaboration within Digital Geosciences, funded partly by DIKU and DIKU/CAPES:
- Continuation of the PETROMAKS/FINEP project PeTWIN.
- Participation in the SINOS collaboration with Brazil.
- Continued participation in the DEXPI initiative.
- Participation and coordination of DSYNE with partners in Brazil and the USA

Horizon 2020 and European Cooperation

- H2020 project Melodic (731664) was expanded with one more project; Morphemic (871643). This project contributes to the infrastructure and content of our program within HPC.
- SIRIUS, by Einar Broch Johnsen, is partner in REMARO. This is a Marie Curie project that looks at the use of formal methods to ensure that autonomous robotics is safe and reliable.
- SIRIUS submitted eight research proposals for H2020 this year: One of them, Plooto, was accepted. Plooto aims to deliver a Circular and Resilient Information System (CRIS) to support manufacturers in their green, digital and circular transition.
- EU Project RE4DY, which was accepted in 2021, was started in 2022. This project focuses on building data-driven digital value networks 4.0 to maintain competitive advantage through digital continuity and superior data spaces across all product and process life cycle phases.
- SIRIUS is a partner and WP leader in ONTOCOMMONS, an H2020 Coordination and Support Action, for the use of semantics in materials technology, design, and manufacturing.
- SIRIUS is a partner in a H2020-SC1-PHE-CORONAVIRUS-2020-2-NMBP project Eur3ka, which will use ICT to build flexible responses to pandemics within the industry.
- Positioning work towards the EU has been done through membership in BDVA (Big Data Value Association) and A.SPIRE PPP for the process industries.
- SIRIUS participates in a Nordic Interoperability Cooperation, where we collaborate with universities and companies in Sweden (Luleå) and Finland (Tampere) about establishing EU projects in this important subject area.

How SIRIUS is Organized

SIRIUS' projects are organised into either business-related work packages - beacon projects - or technically oriented research programs. Each researcher works in a research program and may be involved in one or more beacon projects.

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

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

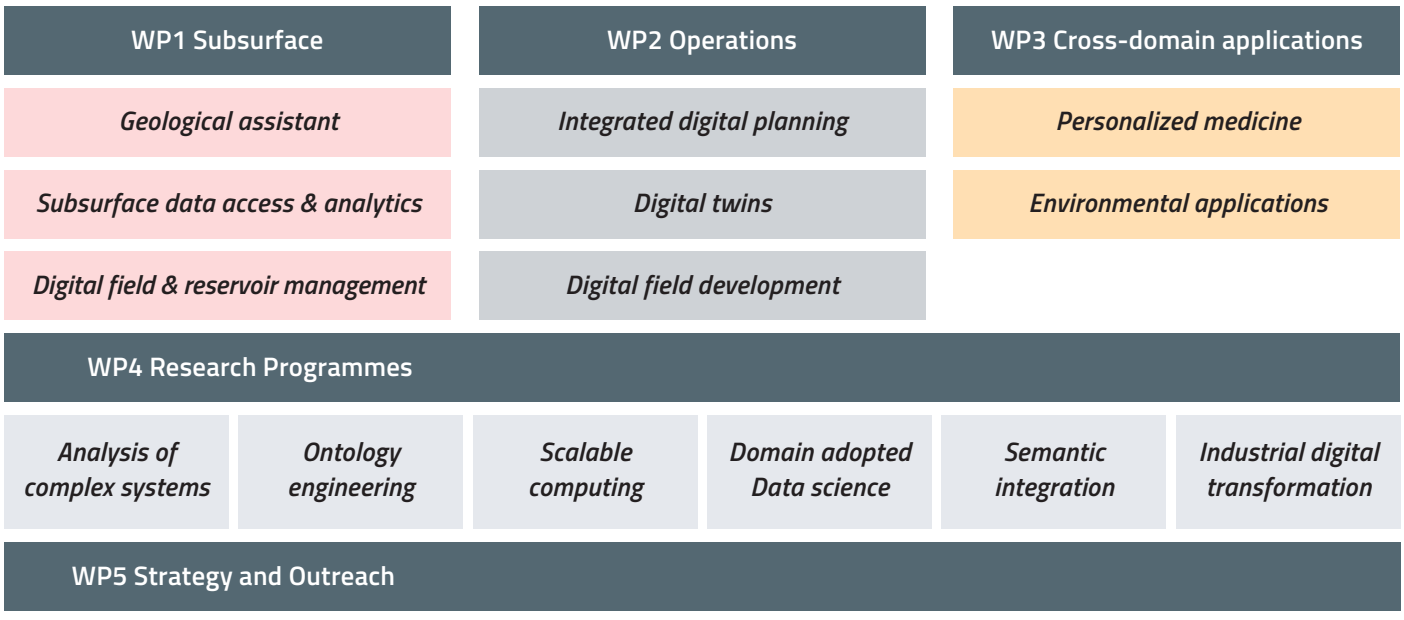
SIRIUS has five work packages:

- Subsurface: beacon projects in subsurface applications.

  Operations: beacon projects in the design, construction, operation, maintenance and decommissioning of complex industrial facilities.

- Cross-domain Applications: applications in areas outside and beyond the natural resources industries, notably health and environmental applications.

- Research Programs: SIRIUS' computer science results that can be applied in projects in several of the business work packages. It is here that we achieve excellence in computer science research.



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

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

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

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

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

SIRIUS RESEARCH PROGRAMS

Analysis of Complex Systems: formal methods for executable design.

Silvia Lizeth Tapia Ta

Now a days industry is experiencing more and more the need for digitization and digital transformation. Supporting technologies to assist industry into such process needs to deal with complex systems. Characteristics of such systems are that they are usually very difficult to understand and analyse due to many interdependent processes happening at the same time. In the oil & gas domain, we can observe such systems in emergent supporting technologies such e.g., digital twins, big networks of heterogeneous autonomous systems, parallel processing of big data, etc. Formal methods are mathematical approaches to support the rigorous specification, design, and verification of the development of digital systems. They are very powerful to capture the behaviours and interactions of complex systems, helping with the understanding of what is currently happening and to further develop predictive and prescriptive analysis. Currently the use of Formal Methods is being more and more accepted in industry. This program uses and researches formal methods for system specifications and formal methods to describe, predict and prescribe the behaviours and interactions of system executions based on the analysis of models.

- **Objective 1 Theoretical development:**

  Analysis of formal models can span from lightweight automated simulations to heavyweight complex non-automated verification techniques. We aim to explore the middle ground, which we call systematic model and multi-model exploration.

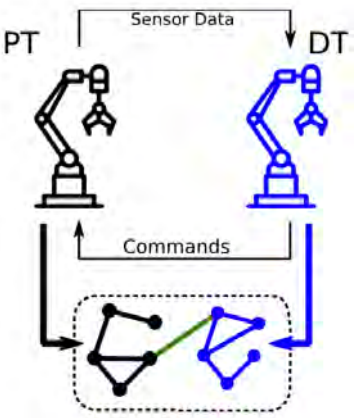
- **Objective 2 Tool development:**

  Tool development (both back- and front- end) for the developed methods. Knowledge extraction, visualization, and fill gaps in industries/market are important for tool development.

In the past year, our research program has focused on the development of new frameworks, targeting different application domains. We now detail some of these activities.

SMOL

- A Language-Based Integration of Knowledge Graphs and Object-Oriented Programs.



Digital Twins (DT) and similar applications typically connect simulators with data-rich components and domain knowledge, both commonly formalised as knowledge graphs. Engineering such applications poses challenges to developers, which we address using a language-based approach to enable their efficient development, as well as explore analysis and design: **SMOL** is an imperative, object-oriented research language which integrates semantic web technologies and numerical simulation blocks, and can serve as a test bed for creating digital twins.

SMOL[3] proposes a **language-based integration of knowledge graphs and object-oriented programs**. SMOL programs can contain queries to external knowledge graphs that contain, e.g., domain knowledge about an application domain, capturing information about a Physical Twin (PT). SMOL further proposes **semantic reflection** of programs

into knowledge graphs by lifting the runtime state of the program into an associated knowledge graph, which enables programs to directly query this semantic representation of itself at runtime. This way, programs can make use of domain knowledge in the knowledge graph. Semantic reflection in SMOL can be used in interesting ways by giving the program access to formalised domain knowledge about its own runtime state, for example for debugging but also for system reconfiguration.

SMOL can be used as a framework for developing digital twins. For digital twins, the knowledge graphs can be used to capture asset models. SMOL programs can then seamlessly interact with asset models and domain knowledge to configure and adapt, e.g., simulators. SMOL uses Functional Mock-Up Objects (FMOs) as a pro gramming layer to encapsulate simulators compliant with the FMI standard into OO structures and integrates FMOs into the class and type systems. By means of the semantic lifting, the FMOs can be integrated into knowledge graphs and used to ensure structural correctness properties for cyber-physical applications.

SMOL features:

- Imperative, object-oriented language
- Seamless integration of programs and knowledge bases
- Built-in reflection of program state in the knowledge base
- Knowledge bases can be queried from SMOL programs
- Encapsulation of simulators in objects based on the FMI standard

Open source. SMOL is released under the terms of the 3-clause BSD licence, the source is available at https://github.com/smolang/SemanticObjects.

REMARO
RELIABLE AI FOR MARINE ROBOTICS

REMARO: Reliable Marine Robotics

REMARO[1] is a European Training Network (ETN) on Reliable artificial intelligence for marine robotics. It started in December 2020 and it is funded by the EU Horizon 2020. REMARO ETN is a consortium of experts in submarine AI, software reliability and marine safety certification (DNV). REMARO ETN is created to educate 15 PhD students, two of them at University of Oslo (UiO). REMARO will develop technology for AI methods with quantified reliability, correctness in specifications, models, tests, and analysis & verification for autonomous systems. The two PhD topics at UiO will focus on safety in self-adaptive meta-controllers and on robustness of controllers with evolving knowledge-based systems. Technology developed in REMARO will be relevant for Digital Twins.

Smart Journey Mining (SJM)

SJM[2] is a research project on digitalization of services. It started in 2021 and it is funded by the The Research Council of Norway. SJM is a consortium of experts in service science, process mining and formal modelling and analysis. SJM is funding one PhD students at University of Oslo (UiO) and one postdoc at SINTEF. The PhD topic at UiO has been focussing on the development of automated analysis using event logs to improve the user experience of services. Technology developed in SJM can be relevant for the integrated digital planning and digitalization of information systems at SIRIUS.

Team Members

Rudolf Schlatte, Researcher	Paul Kobialka, PhD
Violet Pun, Assoc. Professor	Juliane Päßler, PhD
Einar Johnsen, Professor	Tobias John, PhD
Ingrid Yu, Assoc. Professor	Erik Voogd, PhD
S. Lizeth T. Tarifa, Researcher	Åsmund Kløvstad, PhD
Eduard Kamburjan, Postdoc	Chinmayi Baramashetru, PhD

Highlights of 2022

In 2022, our research program focused on developing new frameworks targeting different application domains. This work includes

- Progress in the further development of SMOL, an imperative, object-oriented research language which integrates semantic web technologies and numerical simulation blocks. SMOL has the potential to serve as a powerful test bed for creating digital twins.
- Develop automated analysis techniques using event logs to check properties, e.g., user experience in digital services.
- Develop formal frameworks to study safety in self-adaptive meta-controllers and the robustness of controllers with evolving knowledge-based systems. Such frameworks have significant potential to be used in the context of digital twins.

[1]Project: https://remaro.eu

[2]Project: https://smartjourneymining.no

[3]Project: https://smolang.org/

Ontology Engineering

Martin G. Skjæveland

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

Work in the research program is primarily performed in two projects: the pattern-based ontology engineering project and the information modelling framework project.

The pattern-based ontology engineering project has developed the **Reasonable Ontology Templates** (OTTR) framework. 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.

Members of the project are Martin G. Skjæveland, Leif Harald Karlsen, Christian Kindermann, Oliver Stahl. The framework is available as open source specifications and applications which are in active use by several industrial partners in and outside of SIRIUS, including DNV GL, Aibel, Grundfos and CapGemini.

For more information about OTTR, including interactive examples, specifications and research papers, see its homepage http://ottr.xyz.
In addition to the OTTR framework, the project works on research topics for identifying and characterising patterns in ontologies. OWL ontologies are built and maintained on the basis of all sorts of methods and methodologies using a wide range of tools. As ontologies are primarily published

Vision: 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.

as sets of axioms, their underlying design principles often remain opaque. However, a principled and systematic ontology design is likely to be reflected in regularities for axioms. Identifying such regularities may help to recover and unveil conscious design choices and otherwise recurring modelling practices.

This reverse-engineering of ontologies provides a starting point for high-level language services, e.g., automated rewritings from OWL into OTTR for the purpose of redundancy removal, model verification, data validation, only to name a few.

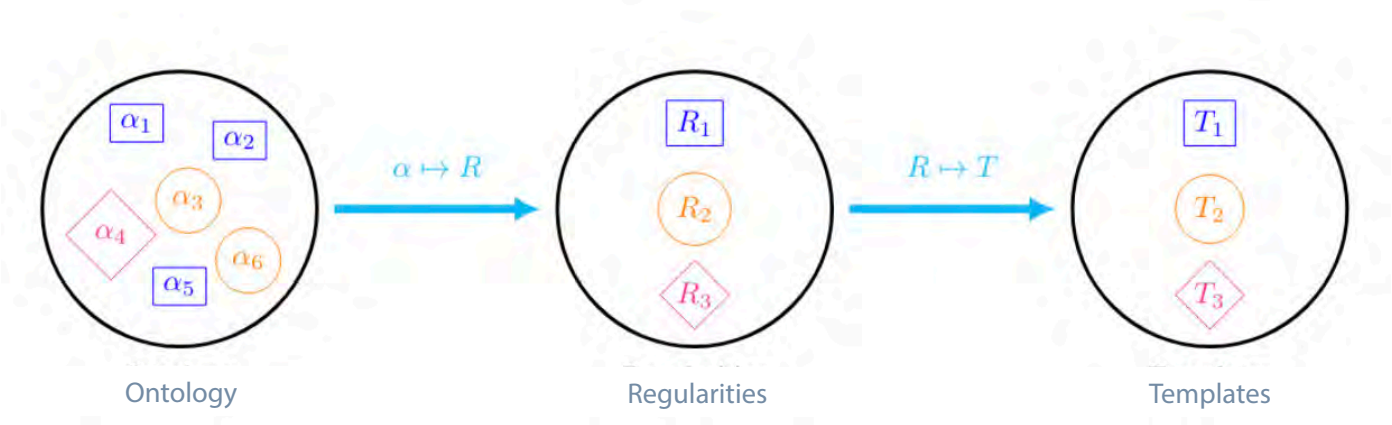
Several master students at UiO undertake a master thesis topic related to the OTTR project. The topics include:

- Characterising and implementing OTTR constructs and the expansion mechanism with a relational database
- Developing algorithms for synchronised updates and expansion of OTTR instance data
- Developing knowledge graph summarisation techniques using OTTR templates
- Developing algorithms for automatically generating bOTTR-mappings from relational databases

Highlights of 2022:

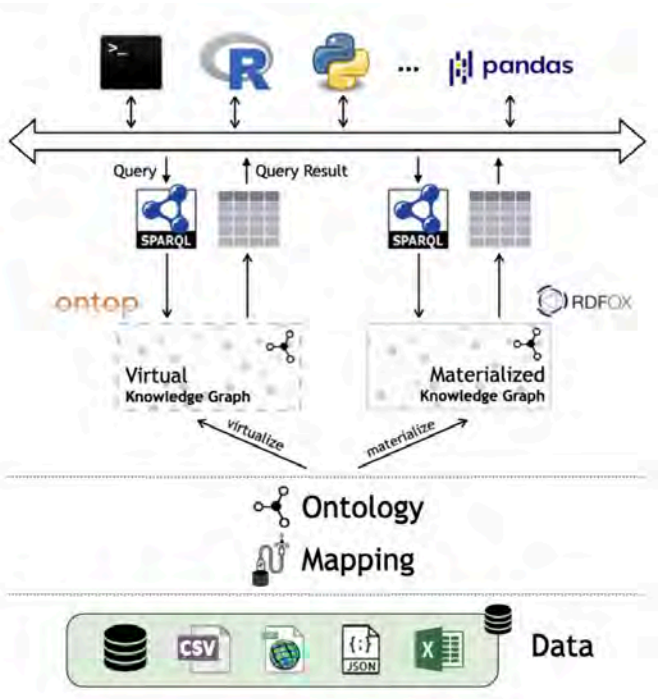
- The 3rd OTTR user forum took place (physically!) in April 2022 co-located with SIRIUS' general assembly. We had about 20 participants, and talks from Christian Hansen (Aibel), Veronika Heimsbakk (CapGemini), Moritz Blum (Bielefeld University), Mikkel Brynhildsen (Grundfos), Christian Kindermann (UiO) and Johan Klüwer (DNV).
- The OTTR framework continues to receive attention from our industrial partners and is a central part of projects that apply semantic technologies.
- Several master students defended their thesis on topics related to OTTR.

Figure: Reverse-engineering ontologies: Automated regularity extraction combined with automated template creation.



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) into Knowledge Graphs. It will demonstrate the efficacy of these tools through deployment in the beacon projects. Specifically, we work with ontology reasoners capable of supporting the development of large-scale ontologies and semantic data stores which answer realistic ontology-based queries over massive data sets.



Semantic Integration requires both solid theoretical foundation and proper tooling support. For the theoretical part, we need to understand the formal semantics of the semantic integration when considering different components: data sources (e.g., relational databases, JSON files), the mapping from these sources to the ontologies, the ontologies in different profiles (e.g., OWL 2 QL, OWL 2 RL, OWL 2), and the queries. We need to develop efficient algorithms and to make sure they are sound and complete. These algorithms need to be implemented in proper software tools. The tools should be validated in real-world use cases to show their effectiveness.

Guohui Xiao

Semantic Integration can be realized in two flavors:

Virtual Knowledge Graphs (VKGs). In the virtual approach, the data assertions are not materialized in a separate data store, but their presence in the KG is only virtual. Systems operating on VKGs are able to retrieve the data directly from the data sources only when it is required for a particular user query. In fact, query processing is delegated to the data sources. This is achieved by unfolding the mappings, thus translating user queries into queries over the data sources, whilst taking into account also the ontology background knowledge through a so-called query rewriting step. The advantage of VKGs is that information is always fresh and up-to-date with the data sources. For example, Ontop is a state-of-the-art Virtual Knowledge Graph system. Ontop implements 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.

Materialized Knowledge Graphs (MKGs). Despite the advantages of the virtual approach, it is sometimes convenient to actually materialize the data assertions. In such a case, we talk about Materialized Knowledge Graphs (MKGs). The main advantage of MKGs over VKGs is that usually a better performance in query answering can be achieved, especially in those situations where mappings are very complex and thus the unfolding of the virtual approach would give rise to complex queries over the data sources. This comes at the cost of maintaining a potentially very large MKG. For example, **RDFox** is a powerful system for Materialized Knowledge Graph. RDFox is a high-performance in-memory knowledge graph and semantic reasoner, optimised for speed and efficiency. Designed from the ground up with reasoning in mind, it outperforms other graph databases while also providing benefits and insights that cannot be achieved by alternatives.

Results

Ontop (https://ontop-vkg.org/) is the state-of-the-art open-source VKG engine. 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.

RDFox (https://www.oxfordsemantic.tech/product) is a high-performance in-memory knowledge graph and semantic reasoner, optimised for speed and efficiency, initially developed by University of Oxford, and currently by Oxford Semantic Technologies.

Ontopic Studio (https://ontopic.ai/en/ontopic-studio/) provides an integrated low-code environment for building RDF graphs from relational data sources. The comfortable KG authoring environment features a lightweight ontology editor and an advanced mapping editor for connecting relational datasets to ontologies, resulting in either complete Virtual Knowledge Graph setups or R2RML files for generating RDF data from relational data sources. Ontopic Studio is powered by Ontop.

Vision: Designing and developing scalable infrastructure for the integration of multiple large datasets and large-scale ontologies

Highlights of 2022

- Application of the Virtual Knowledge Graph approach for converting clinical data in relational DBs to FHIR RDF. This works is published at the top-tier Journal of Biomedical Informatics.
- The Ontopic Studio (https://ontopic.ai/en/ontopic-studio/) system becomes mature. Our slogan is "Design your Enterprise Knowledge Graph with ease".
- The work on optimising CQ answering over unrestricted OWL 2 ontologies resulted in the development of a new tool: ACQuA, an OWL reasoner for CQ answering combining RSAComb, PAGOdA, and a sprinkle of HermiT.

[1] Guohui Xiao, Emily Pfaff, Eric Prud'hommeaux, David Booth, Deepak K. Sharma, Nan Huo, Yue Yu, Nansu Zong, Kathryn J. Ruddy, Christopher G. Chute, and Guoqian Jiang. FHIR- Ontop-OMOP: Building clinical knowledge graphs in FHIR RDF with the OMOP common data model. Journal of Biomedical Informatics, 134:104201, 2022.

Team

UiO: Arild Waaler, Egor V. Kostylev, Martin Giese, Jens Otten, Vidar Klungre, Ratan Bahadur Thapa, Eduard Kamburjan, Adnan Latif

University of Oxford: David Tena Cucala, Federico Igne, Ian Horrocks, Jiaoyan Chen, Shuwen Liu, Stefano Germano, Temitope Ajileye, Xiaxia Wang, Xinyue Zhang, Yuan He

City University of London: Ernesto Jimenez-Ruiz, Roman Kontchakov

Bosch: Evgeny Kharlamov

Oxford Semantic Technology: Peter Crocker

Ontopic: Guohui Xiao, Benjamin Cogrel, Diego Calvanese, Peter

Prediktor: Magnus Bakken

Domain-Adapted Data Science

Within the artificial intelligence (AI) research community and beyond that community there is interest in developing strong AI, which means intelligent machines that are indistinguishable from the human mind or that go beyond human-level intelligence (superintelligence). However, despite the impressive progress in the field over the last decades, we still do not know how to achieve strong AI.

Basil Ell

As an example of human capability, imagine a child who 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.

This example shows 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.

In the AI community, there has been a long-standing frontier, labeled as the discussion about "symbolic vs. non-symbolic AI." In symbolic AI, information is structured in ontologies and deductions are made via reasoning. In sub-symbolic AI, information is obtained from data, and deductions are made via machine learning (ML). One hypothesis is that machines will become better at learning if we can combine these two types of information in the learning process. That is, combine information in terms of declarative statements and ontologies with information encoded in data made available through statistics and machine learning.

The focus of the SIRIUS domain adapted data science program is exactly to develop approaches that combine the use of structured knowledge with learning from data in the machine learning process. On one hand, this means that we try to bridge the traditional divide between symbolic and sub-symbolic learning, developing what we refer to as hybrid approaches. On the other hand, as it turns out, hybrid approaches yield improved machine learning results, and especially on "not so big-data". Combined with the fact that symbolically represented knowledge can often be very small and concise, this is a powerful tool that makes machine learning available on datasets that are otherwise too small, or otherwise unfit, for classical machine learning tasks. The overarching goal is a general methodology for how data science tasks can be enhanced through the combined use of symbolic and sub-symbolic knowledge.

Another intriguing feature of hybrid approaches is that the presence of symbolic knowledge in the machine learning process may lead to more explainable predictions. Within our research program, we also develop novel hybrid approaches that identify and exploit these capabilities.

One term that we use to refer to a particular class of hybrid approaches is 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.

In the context of our project domain adapted data science pipeline, we develop a catalogue of situations related to domain adaptation and describe how these situations are related to each other. For example, consider the situation that an organization performs machine learning given a tabular dataset, is interested in improving the performance of the approach, has selected the most appropriate approach, and has tried out the usual performance improvement strategies such as hyperparameter optimization, but has so far not made use of domain knowledge such as taxonomies or ontologies. (Often it is a case that "a little semantics goes a long way", which means that the ontology needs not to be very extensive. Instead, a couple of statements can already make a significant difference.) One way to go forward can be to check whether openly available domain knowledge exists (e. g., in the Wikidata knowledge graph) that can be used to contextualize the data so that the performance of the ML

Vision: We develop hybrid approaches that exploit both knowledge in data and knowledge in ontologies.

approach could be improved. Thus, we arrive at research questions such as: Given tabular data, how to find openly available domain knowledge in the Web of Data that could contextualize given tabular data? How to align tabular data to entities and properties in an RDF dataset (i. e., the domain knowledge represented in RDF format, which is the common format in the Semantic Web)? How to find out which parts of the external knowledge help most, so that when this information is improved it will have a significant impact on the performance? Which parts of the data have a negative impact on the performance and should thus be removed? How exactly can external knowledge be incorporated into an ML approach such as in a preprocessing step to improve the quality of existing training data? How can domain knowledge help in post-processing the output of the ML approach? How can the solution space exploration of the ML approach be guided by domain knowledge? How can the search space be pruned by making use of domain knowledge?

The situations and research questions that we collect can be ordered in two groups: those research questions that primarily address domain adaptation, such as, how to make use of (medical) taxonomies while training a (disease) classification model? Secondary research questions do not directly address domain adaptation but enable domain adaptation, e. g., embedding of knowledge graphs into vector spaces so that these can be processed by classical ML approaches, or they tackle the improvement of domain knowledge so that the performance of a domain adapted approach can be improved further, e. g., by anomaly detection in knowledge graphs and automatic knowledge graph completion.

The task of identifying and describing situations is both a top-down approach (guided by combinatorial exploration and brainstorming, zooming out and in) and a bottom-up approach (inspired by existing research). Not only does it allow us to structure existing approaches according to

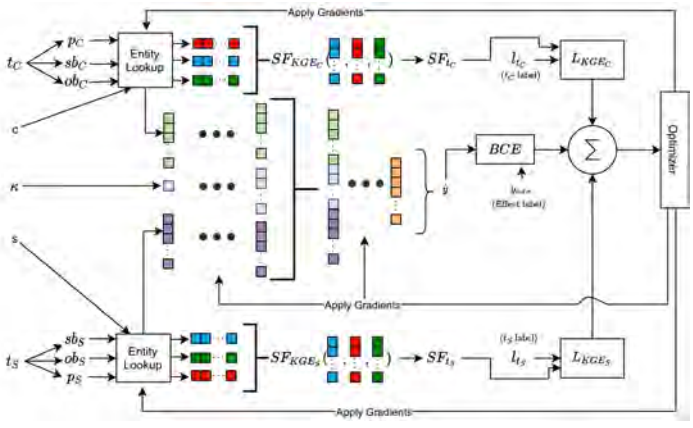


Figure 1. Image taken from [2]. Fine tuning model architecture where each component is shown with inputs and outputs. t_C and t_S are knowledge graph triples relating to chemicals and species each with a score SF and loss l. c, s, and k are the prediction input variables while ŷ is the predicted toxicity.

situations, but it also lets us identify research questions that have so far not been addressed sufficiently. Our vision is that we can develop a methodology for domain adaptation, where an individual or organization browses the graph of situations to learn about how to realize domain adaptation or at least finds pointers to relevant sources such as publications.

Knowing what research could be done, which is an outcome of our activities, needs to be complemented by what is relevant for our SIRIUS partners, so that we can focus on those tasks or challenges especially relevant to our partners for mutual benefit. Therefore, we plan a stronger involvement of our partners in 2023 to prioritize our activities and to create a roadmap for research.

Beyond domain adaptation or improving domain knowledge that can then be used for domain adaptation, we actively work on a couple of other topics, within a group that also includes non-SIRIUS researchers at the UiO Department of Informatics (i. e., Anne-Marie George, Thomas Kleine Büning, and Meirav Segal). For example, we investigate in which way domain knowledge can assist reinforcement learning tasks. UiO researchers in the NRC project "Safe and Beneficial Artificial Intelligence" investigate active learning problems involving human and societal preferences such as learning preferences from interactions, collaborating effectively with humans, and making repeated decisions that are fair in the long term. Here, background knowledge, e. g., behavioral conventions, structure of preferences, features, and their relations of states of the world, might be able to improve the learning. Reversely, elicitation schemes might be able to gather and structure knowledge.

Furthermore, we carry out research related to explainable AI (XAI). The current approaches and technologies in XAI mostly focus on shedding light on the behavior of black-box machine learning models (like deep neural networks) by explaining their decisions to the users. However, the least work has been done towards employing the information provided by the explanations for enhancing the models concerning accuracy, fairness, and robustness in a systematic way. In SIRIUS, we have studied this research area and devised explanation-based frameworks for investigating the accuracy and robustness of black-box ML classification models **[1]**.

Program members:

Basil Ell (program leader) , Daniel Bakkelund , Egor V. Kostylev , Erik Bryhn Myklebust , Ernesto Jimenez-Ruiz , Evgeny Kharlamov , Fábio Corrêa Cordeiro , Gong Chen , Ingrid Chieh Yu , Martin Giese , Jiaoyan Chen , Ole Magnus Holter ,Peyman Rasouli , Roxana Pop and Summaya Mumtaz

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 – see Figure 1 [2].

- Ole Magnus and Basil Ell develop approaches that make use of domain knowledge in the context of semantic parsing of textual requirements. Their goal is to formally represent (parts of) the meaning of textual requirements, so that the meaning of requirements becomes more accessible to machines and the management of requirements can be improved [3].

- Egor V. Kostylev and colleagues study theoretical and practical connections between graph neural networks (GNNs), a modern structure-aware machine learning architecture, and classic logic-based knowledge representation formalisms. In particular, they designed a family of monotonic GNNs that allow for an efficient translation to Datalog logic-based language, and developed an efficient INDIGO system for knowledge graph completion [4, 5].

- 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 [6]. The disputation of her PhD thesis took place in November 2021.

- 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 [7]. Daniel submitted and defended his PhD thesis in 2022.

- 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 [8, 9].

- Fábio Corrêa Cordeiro and colleagues have developed applications for information retrieval of structured and unstructured data for the energy industry. They have studied several aspects of information extraction and semantic search [24,25,26] and made available Petrolês (https://petroles.puc-rio.ai/index_en.html), a public repository of artifacts for Natural Language Processing applications in the petroleum domain in Portuguese [27]. For his PhD thesis, Fábio is researching a methodology for populating a knowledge graph with information extracted from technical documents.

- Actionable recourse (AR) techniques are a popular class of post-hoc interpretability approaches that help the users of ML models to obtain their desired decision from a machine learning model. Given an individual's preferences, an AR recommends feasible changes to their corresponding input that lead to the desired outcome by the model. To generate realistic ARs, it is important to capture and exploit the domain's information and the preferences of the users in the explanation process. Peyman Rasouli and Ingrid Chieh Yu are working on a model-agnostic framework that combines user/domain-level knowledge with model/data-level information to create plausible ARs that can guide individuals to obtain their desired decision from any ML classification and regression model in a simple and efficient manner.

- Current explainable artificial intelligence (XAI) techniques only rely on the observational data to analyze and explain the behavior of machine learning models. To increase the comprehensibility and faithfulness of explanations of ML

models, hence, it is essential to exploit domain knowledge that bridges between the models and human concepts. Peyman Rasouli and Ingrid Chieh Yu aim to integrate domain knowledge (in the form of knowledge graphs and taxonomies) with structured/tabular data to provide more feasible, comprehensible, and faithful explanations [22, 23]. Peyman will submit his thesis in 2023.

- Gong Cheng, Evgeny Kharlamov investigated keyword-based exploration of knowledge graphs [10,11] and proposed a novel method to generate smart snippets or summaries of large-scale knowledge graphs. Baifan Zhou, Evgeny Kharlamov and colleagues from SIRIUS showed how to facilitate development of ML models using semantic technologies [12]. Then, they investigated several practical aspects of knowledge graph management in connection to analytics and machine learning motivated by applications from Industry 4.0 [13,14]. That is, they showed how to scale usability of ML analytics and reshape industrial knowledge graphs. Moreover, Baifan and Evgeny consolidated a number of research directions into an advanced SIndAIS4 project (https://sirius-labs.no/sindais4-scaling-industrial-ai-with-semantics/) of SIRIUS that aims at Scaling Industrial AI with Semantics in four directions: human, data, methods, and applications. Within this project and together with Ahmet Soylu they selected several Bosch-funded interns – students of Ahmet – thus strengthening the Bosch-SIRIUS collaboration and disseminating it in two large Norwegian universities: NTNU and OsloMet.

- Basil Ell develops approaches to align symbolic data (i. e., ontologies) with sub-symbolic data (e. g., texts or tables). The alignment enables labeled training data to be generated via distant supervision for approaches such as information extraction (IE) for ontology population or natural language generation (NLG). Having symbolic and sub-symbolic data aligned means obtaining hybrid data that can be processed by hybrid approaches. **He received a best paper award at LDK 2021 – 3rd Conference on Language, Data and Knowledge**, for his work on mining association rules that help to bridge between text and data [15] – see Figure 2. Furthermore, he develops statistical approaches that are applied to symbolic data (KGs) for the purposes of identifying regularities and anomalies, for the prediction of missing facts, for the evaluation of the structural plausibility of facts, for bridging between structured and unstructured data (as in IE, question answering, NLG), and the structural classification of regions within graphs (which is similar to sequence labeling, but on graphs).

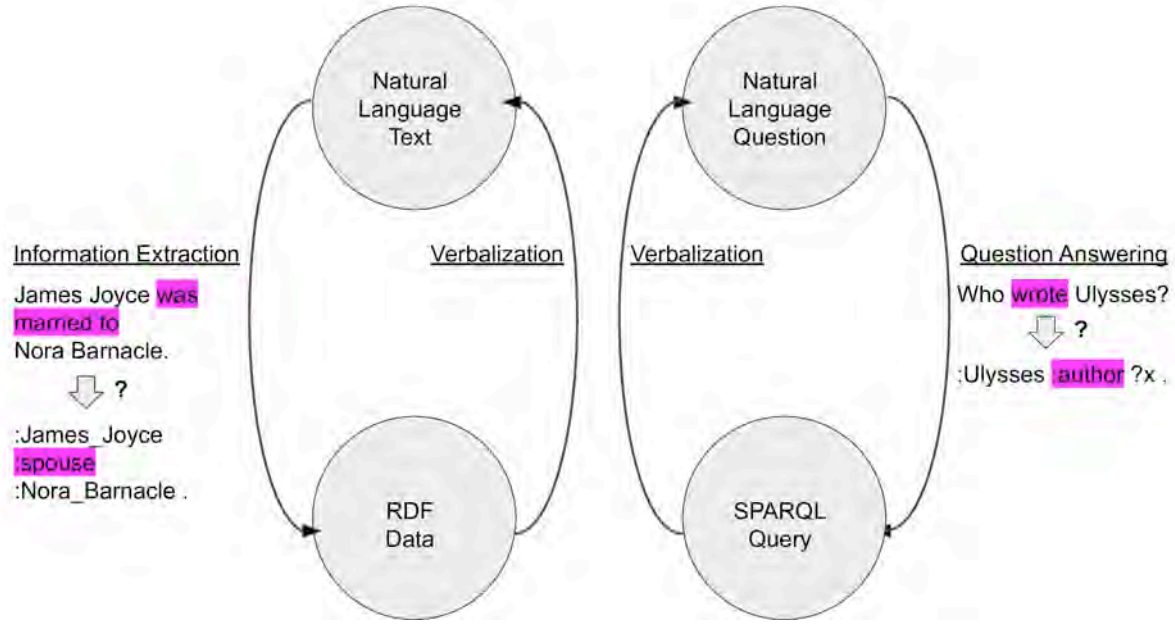


Figure 2. Image from [15]. There is a non-trivial lexical gap between expressions in natural language and terms in a knowledge graph, that needs to be bridged for a couple of tasks such as Information Extraction, Question Answering, and Verbalization (of RDF data or SPARQL queries).

Highlights of 2022:

- Daniel Bakkelund submitted and defended his thesis https://www.duo.uio.no/handle/10852/94452, and published a journal paper [21]
- Ernesto Jimenez-Ruiz was working as Knowledge Graph consultant for Samsung Research UK
- Best paper candidate in ISWC: "Yuan He, Jiaoyan Chen, Hang Dong, Ernesto Jiménez-Ruiz, Ali Hadian, Ian Horrocks: Machine Learning-Friendly Biomedical Datasets for Equivalence and Subsumption Ontology Matching. ISWC 2022: 575-591"
- Erik Bryhn Myklebust successfully defended his PhD
- The IG on KGs at the Turing has more than 300 members (chaired by Ian Horrocks, Ernesto Jimenez-Ruiz and others): https://www.turing.ac.uk/research/interest-groups/knowledge-graphs
- Ernesto Jimenez-Ruiz, Evgeny Kharlamov and Ian Horrocks are ranked amongst the top-10 scholars in the Knowledge Engineering field in their AI 2000 scholars ranking of 2023 https://www.aminer.cn/ai2000?domain_ids=5dc122672ebaa6faa962c073
- Roxana Pop got selected for a research stay at the Alan Turing Institute through NORA.
- We organized 2 workshops (Ontology Matching workshop, NeSy workshop)
- We organised 2 challenges (OAEI and SemTab)
- Ernesto Jimenez-Ruiz started as Research track chair last year for ESWC 2023
- We organised the 1st Symposium for the IG on KGs
- Egor V. Kostylev continued to work on the theory and practice of Graph Neural Networks; for example, he designed novel explainable GNN-based models over Knowledge Graphs (published at ICLR 2022) and neuro-symbolic approaches to query answering over incomplete Knowledge Graphs (published at ISWC 2022). Besides this, he established foundations of Reverse OTTR, a query language for knowledge graphs based on template ontology language OTTR.

Our ambitions for 2023 for the domain adapted data science pipeline project are to identify the most relevant data science activities relevant for our industry partners that greatly benefit from incorporating domain knowledge, to identify research gaps and to develop a roadmap for research, to develop methods to advance current methods so that domain knowledge can be incorporated, to develop methods to identify existing domain knowledge, and to develop methods that improve existing domain knowledge.

List of selected publications:

1. Peyman Rasouli and Ingrid Chieh Yu: "Explainable Debugger for Black-box Machine Learning Models." 2021 International Joint Conference on Neural Networks (IJCNN), IEEE, pp. 1-10, 2021. [https://ieeexplore.ieee.org/document/9533944]
2. Erik B. Myklebust, Ernesto Jiménez-Ruiz, Jiaoyan Chen, Raoul Wolf, Knut Erik Tollefsen: "Prediction of Adverse Biological Effects of Chemicals Using Knowledge Graph Embeddings." Semantic Web journal (2021). [http://www.semantic-web-journal.net/system/files/swj2804.pdf]
3. Ole Magnus Holter and Basil Ell: "Towards Scope Detection in Textual Requirements." 3rd Conference on Language, Data and Knowledge (LDK 2021). [https://drops.dagstuhl.de/opus/volltexte/2021/14567/pdf/OASIcs-LDK-2021-31.pdf]
4. David Jaime Tena Cucala, Bernardo Cuenca Grau, Egor V. Kostylev, Boris Motik: "Explainable GNN-Based Models over Knowledge Graphs." International Conference on Learning Representations (ICLR 2022). [https://openreview.net/pdf?id=CrCvGNHAIrz]
5. Shuwen Liu, Bernardo Cuenca Grau, Ian Horrocks, Egor V. Kostylev: "INDIGO: GNN-based inductive knowledge graph completion using pair-wise encoding." The 34th Annual Conference on Advances in Neural Information Processing (NeurIPS 2021). [https://proceedings.neurips.cc/paper/2021/file/0fd600c953cde8121262e322ef09f70e-Paper.pdf]
6. Summaya Mumtaz and Martin Giese: "Frequency-Based vs. Knowledge-Based Similarity Measures for Categorical Data." AAAI Spring Symposium: Combining Machine Learning with Knowledge Engineering (1). 2020. [http://ceur-ws.org/Vol-2600/paper16.pdf]

7. Daniel Bakkelund: "Order preserving hierarchical agglomerative clustering." Machine Learning (2021). [https://link.springer.com/content/pdf/10.1007/s10994-021-06125-0.pdf]

8. Jiaoyan Chen, Pan Hu, Ernesto Jiménez-Ruiz, Ole Magnus Holter, Denvar Antonyrajah, Ian Horrocks: "OWL2Vec*: embedding of OWL ontologies." Machine Learning 110.7 (2001): 1813-1845. [https://link.springer.com/content/pdf/10.1007/s10994-021-05997-6.pdf]

9. Jiaoyan Chen, Ernesto Jiménez-Ruiz, Ian Horrocks, Denvar Antonyrajah, Ali Hadian, Jaehun Lee: "Augmenting Ontology Alignment by Semantic Embedding and Distant Supervision." Extended Semantic Web Conference (2021): 392-408. [https://openaccess.city.ac.uk/id/eprint/25810/1/]

10. Yuxuan Shi, Gong Cheng, Trung-Kien Tran, Evgeny Kharlamov, Yulin Shen: "Efficient Computation of Semantically Cohesive Subgraphs for Keyword-Based Knowledge Graph Exploration." WWW 2021: 1410-1421. [https://dl.acm.org/doi/pdf/10.1145/3442381.3449900]

11. Yuxuan Shi, Gong Cheng, Trung-Kien Tran, Jie Tang, Evgeny Kharlamov: "Keyword-Based Knowledge Graph Exploration Based on Quadratic Group Steiner Trees." IJCAI 2021: 1555-1562. [https://www.ijcai.org/proceedings/2021/0215.pdf]

12. Baifan Zhou, Yulia Svetashova, Andre Gusmao, Ahmet Soylu, Gong Cheng, Ralf Mikut, Arild Waaler, Evgeny Kharlamov: "SemML: Facilitating development of ML models for condition monitoring with semantics." Journal of Web Semantics 71: 100664 (2021). [https://reader.elsevier.com/reader/sd/pii/S1570826821000391]

13. Baifan Zhou, Dongzhuoran Zhou, Jieying Chen, Yulia Svetashova, Gong Cheng, Evgeny Kharlamov: "Scaling Usability of ML Analytics with Knowledge Graphs: Exemplified with A Bosch Welding Case." IJCKG 2021: 54-63. [https://dl.acm.org/doi/pdf/10.1145/3502223.3502230]

14. Dongzhuoran Zhou, Baifan Zhou, Jieying Chen, Gong Cheng, Egor V. Kostylev, Evgeny Kharlamov: "Towards Ontology Reshaping for KG Generation with User-in-the-Loop: Applied to Bosch Welding." IJCKG 2021: 145-150. [https://dl.acm.org/doi/pdf/10.1145/3502223.3502243]

15. Basil Ell, Mohammad Fazleh Elahi, and Philipp Cimiano: "Bridging the Gap Between Ontology and Lexicon via Class-Specific Association Rules Mined from a Loosely-Parallel Text-Data Corpus." 3rd Conference on Language, Data and Knowledge (LDK 2021). [https://drops.dagstuhl.de/opus/volltexte/2021/14569/pdf/OASIcs-LDK-2021-33.pdf]

16. Xiaxia Wang, Gong Cheng, Tengteng Lin, Jing Xu, Jeff Z. Pan, Evgeny Kharlamov, Yuzhong Qu: "PCSG: Pattern-Coverage Snippet Generation for RDF Datasets." ISWC 2021: 3-2.

17. Roberto Avogadro, Marco Cremaschi, Ernesto Jiménez-Ruiz, Anisa Rula: "A Framework for Quality Assessment of Semantic Annotations of Tabular Data." International Semantic Web Conference (2021): 528-545. [https://openaccess.city.ac.uk/id/eprint/26426/1/]

18. Yuan He, Jiaoyan Chen, Denvar Antonyrajah, Ian Horrocks: "BERTMap: A BERT-based Ontology Alignment System." AAAI 2022. [https://arxiv.org/abs/2112.02682]

19. Jiaoyan Chen, Yuan He, Ernesto Jimenez-Ruiz, Hang Dong, Ian Horrocks: "Contextual Semantic Embeddings for Ontology Subsumption Prediction." Submitted to KR 2022. [https://arxiv.org/abs/2202.09791]

20. Peyman Rasouli and Ingrid Chieh Yu: "Analyzing and Improving the Robustness of Tabular Classifiers using Counterfactual Explanations." 20th IEEE International Conference on Machine Learning and Applications (ICMLA), pp. 1286-1293, 2021. [https://ieeexplore.ieee.org/document/9679972]

21. Daniel Bakkelund: "Machine part data with part-of relations and part dissimilarities for planted partition generation." Data in Brief 42 (2022): 108065. [https://www.sciencedirect.com/sdfe/reader/pii/S2352340922000956/pdf]

22. Peyman Rasouli and Ingrid Chieh Yu: "CARE: coherent actionable recourse based on sound counterfactual explanations." Int J Data Sci Anal (2022). [https://doi.org/10.1007/s41060-022-00365-6]

23. Peyman Rasouli, Ingrid Chieh Yu, and Ernesto Jiménez-Ruiz: "Interpreting Categorical Data Classifiers using Explanation-based Locality." 2022 IEEE International Conference on Data Mining Workshops (ICDMW), Orlando, FL, USA, 2022, pp. 163-170, doi: 10.1109/ICDMW58026.2022.00030. [https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=10031039]

24. L. L. de Oliveira et al., "Evaluating and mitigating the impact of OCR errors on information retrieval", Int J Digit Libr, jan. 2023, doi: 10.1007/s00799-023-00345-6. [https://link.springer.com/content/pdf/10.1007/s00799-023-00345-6.pdf]

25. J. D. B. Castro, S. W. A. Canchumuni, C. E. M. Villalobos, F. C. Cordeiro, A. M. A. Alexandre, e M. A. C. Pacheco, "Improvement Optical Character Recognition for Structured Documents using Generative Adversarial Networks", em 2021 21st International Conference on Computational Science and Its Applications (ICCSA), Cagliari, Italy, set. 2021, p. 285–292. doi: 10.1109/ICCSA54496.2021.00046. [https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9732375]

26. D. da S. M. Gomes et al., "Portuguese word embeddings for the oil and gas industry: Development and evaluation", Computers in Industry, vol. 124, p. 103347, jan. 2021, doi: 10.1016/j.compind.2020.103347.

27. F. C. Cordeiro e C. E. M. Villalobos, "Petrolês - How to Build a Specialized Oil and Gas Corpus in Portuguese", ROG, vol. 20, no 2020, p. 387–388, dez. 2020, doi: 10.48072/2525-7579.rog.2020.387. [https://icongresso.ibp.itarget.com.br/arquivos/trabalhos_completos/ibp/3/final.IBP0047_20_14072020_100743.pdf]

Industrial Digital Transformation

Thomas Østerlie

Increased availability and use of digital data hold the potential to transform how organizations work and collaborate as well as transforming their products and services. Yet, digital transformation is not easy. Both experience and research show that digital transformation frequently fall short of meeting expectations.

There are many reasons for this, but an often-overlooked aspect is the need to cultivate the organizational preconditions necessary for realizing digital data and technologies' transformative potential. What these preconditions entail in practice varies greatly and remains object of much scientific scrutiny across disciplines concerned with current shifts towards data-centric and -driven forms of work and organizing.

Our vision is to produce knowledge in close collaboration with industrial stakeholders, that

(1) inform collaborating companies in planning and organizing for digital transformation, and (2) advance scientific knowledge.

We pursue this vision through active engagement with digital transformation initiatives, where we apply our methods and theoretical frameworks to address important and difficult challenges and issues. Our focus is the relationship between social and technical factors during development, implementation, and use of digital data and technologies.

Our **scientific objective** is to develop empirically grounded insights and theory on digital transformation in general, with particular emphasis on transitions towards data-centric and -driven forms of work and organizing. We pursue this across multiple levels of scale from micro-level studies of data-centric work practices, via adoption and implementation of tools and data at the company level, to large-scale technological change at the industry level.

We have pursued our vision and scientific objective through a series of studies and projects throughout Sirius' lifecycle. During 2021, we pursued our research through two ongoing projects. These projects demonstrate the breadth of our research, from innovation to activities oriented towards basic research.

Digitalization of LCI exchange

This project addresses the problem complex associated with digitalization of LCI exchange. It is conducted in close collaboration with key industry stakeholders through our participation in the NORSOK Z-TI expert group.

Background. Structured and machine-readable life-cycle information (LCI) is a pre-requisite for the transition to data-centric and -driven approaches to design, construction, operations, and maintenance of oil and gas installations. Yet, much of it is today stored and exchanged between companies' internal legacy systems in form of unstructured data such as digital documents and images. What exists of structure information exchange is between proprietary systems on formats governed by single providers rather than as standards at the industry level.

While there is shared understanding of the need for structured and machine-readable LCI standards throughout the oil and gas industry, there is little to no coherence across companies to jointly solve central challenges associated with this. The result is that digitalization of LCI exchange is driven through disparate activities within individual enterprises, through larger capital expense projects or across various joint industry initiatives.

Results. This joint work has resulted in two innovations so far: (1) a novel standardization strategy and (2) a mechanism for industry-wide coordination of standardization.

These two innovations are part of the NORSOK national strategy for digitalization of LCI exchange

Future work will focus on evaluating and improving upon the innovations, as well as further development and detailing of key principles underpinning the national digitalization strategy. There is also ongoing work to introduce these innovations as basis for international standardization initiatives.

Relationship with other activities. Our work in NORSOK Z-TI builds and extends upon engagement in the READI JIP, predominantly through Mina Hagshenas PhD project. We also have an ongoing PhD project on digital tool support for early-stage design of subsurface facilities. This PhD is funded through the BRU21 program.

Further reading

Haghshenas, M. and T. Østerlie (2020). "Coordinating innovation in digital infrastructure: The case of transforming offshore project delivery", in Agrifoglio, R., R. Lamboglia, D. Mancini, and F. Ricciardi (Eds.), Digital Business Transformation: Organizing, Managing, and Controlling in the Information Age, Springer Verlag.

Haghshenas, M. and T. Østerlie (2020). "Navigating towards a digital ecosystem: The case study of offshore infrastructure industry", in Proceedings of the 11th Scandinavian Conference on Information Systems (SCIS2020)

Haghshenas, M. and T. Østerlie (2019). "Digital infrastructure innovation vs. digital innovation in infrastructure: Digital transformation in the offshore construction industry", in Proceedings of the 6th edition of The Innovation in Information Infrastructure (III) workshop.

Geological data preparation

This project is a collaboration with the Sirius research programs on semantic integration and digital geoscience, with Equinor as industrial partner.

Background. Analytics applied on vast quantities of digital data holds the potential for faster and better decisions. However, this is of limited use ifthe time it takes to prepare the data exceeds the decision window. This is the situation facing oil and gas companies in transitioning from traditional subsurface evaluation techniques towards data-driven decision-making. While existing data preparation techniques work perfectly with traditional (small data) subsurface evaluation, they do not scale to big data settings.

Results. This project is a continuation of our sustained engagement with transitions towards data-centric approaches in oil and gas exploration. This is a topic we have been engaged with since the beginning of Sirius; initially through the Geological Assistant beacon that we initiated together with Equinor and Schlumberger, and later through sustained collaboration with Equinor's data management community. Our output from this has been predominantly theoretical results, but has formed and continues to form the basis for innovations developed by the other research programs.

Future work. We will contribute to evaluating and improving the prototypes to be delivered by the other research programs as part of this study.

Further reading

Parmiggiani, E., T. Østerlie, and P.G. Almklov (2022). "In the backrooms of data science", Journal of the Association for Information Systems, 23(1), pp.139-164.

Monteiro, E., T. Østerlie, E. Parmiggiani, and M. Mikalsen (2018). "Quantifying quality: Towards a post-humanist perspective on sensemaking", in Schultze, U., M. Aanestad, M. Mähring, and K. Riemer (Eds.), Living with monsters? Social implications of algorithmic phenomena, hybrid agency, and the performativity of technolgy, Springer Verlag.

Østerlie, T., Parmiggiani, E., and Monteiro, E. (2017). "Information infrastructure in the face of irreducible uncertainty", in Proceedings of the 5th edition of Innovation in Information Infrastructure (III) workshop.

Research program team

Contact person: Thomas Østerlie, Associate Professor, Department of Computer Science, NTNU

Eric Monteiro, Professor, Department of Computer Science, NTNU

Elena Parmiggiani, Associate Professor, Department of Computer Science, NTNU

Mina Haghshenas, PhD student, Department of Computer Science, NTNU

Scalable Computing

The Subsurface work package organizes all the projects that meet business needs for subsurface activities in the oil and gas domain. This work includes geology, geophysics, reservoir simulation, petroleum technology and drilling.

The Scalable Computing (SC) research program is about making data access and processing faster to SIRIUS projects. This is achieved by building knowledge in High Performance Computing (HPC) and coupling this with scalable Cloud computing to support scalable big-data application processing. Specifically, we look at solutions for scalable and reconfigurable hardware, software design for parallel numerical simulations, and automatic cross-cloud application deployment and reconfiguration using hardware accelerators.

The target problems of HPC involve large scale computations that are beyond the capabilities of single 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 interconnected computers. The research topics of HPC encompass parallelization schemes, partitioning algorithms, communication overhead reduction strategies, software implementation and optimization techniques, use of heterogeneous clusters that consist of both conventional CPUs and cutting-edge hardware accelerators, in addition to adopting HPC for real-world applications.

The 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.

The Scalable Computing research have the following objectives:

Objective 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.

Objective 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.

Objective 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.

Activities and their contribution to the main objectives:

Objective 1:

- Work is ongoing to link UiO's Numascale 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.

Objective 2:

- 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 MORPHEMIC project developing proactive and polymorphic Cross-Cloud application management has successfully passed its mid-term evaluation with a real-world application demonstration.
- Ongoing PhD project in utility optimising Cross-Cloud autonomic application management.

Objective 3:

- 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 Cloud activities of Scalable Computing in SIRIUS was for the first three years focused on contribution to the MELODIC1 Horizon 2020 project. MELODIC supports Cross-Cloud application management through the Application Programming Interface (API) offered by the various Cloud providers and it is thereby able to deploy Cloud computing instances of the application's components in the form of virtual machines (VMs), containers, and serverless functions. 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 OW22, or as a supported and installed package on standard commercial terms. 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 (AWS3). Google Cloud4, and the Azure Cloud5 from Microsoft. Additionally, there is support for some smaller European Cloud providers, and the open-source Cloud infrastructure management platform OpenStack6, which is used by most academic Cloud installations worldwide, among them the he Norwegian Research and Education Cloud (NREC7).

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 TensorFlow8 processing. Hence, an application component may come as a standard Central Processing Unit (CPU) artefact, which is currently being deployed by MELODIC, or as artefacts compiled for one or more of the hardware accelerators. Furthermore, the accelerated 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 the SIRIUS Scalable Computing team leads the effort in the Horizon 2020 project MORPHEMIC9 to add this support.

The optimization of the application's Cloud deployment configuration in MELODIC is largely reactive as it is based on measured changes in the managed application's execution context. Acquiring 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 series prediction of the measurements of the running application's execution context and perform the optimization and reconfiguration early so that the Cloud resources will be available when they are needed by the application.

The MORPHEMIC project successfully passed the mid-term evaluation by European Commission in the autumn of 2021 demonstrating a real application running on the first release of the software platform. This is a major recognition given the Covid-19 restrictions in effect during the first half of the project. The SIRIUS researchers have successfully published four conference publications in 2021 related to MORPHEMIC, and work is ongoing on a further 6 publications expected for 2022. Work to integrate the HPC cluster from NumaScale into the NREC computing facilities allowing it to be used transparently by MELODIC and MORPHEMIC managed HPC applications started in 2020 will hopefully be completed early in 2022. The MORPHEMIC platform will be demonstrated at the SIRIUS General Assembly in the spring of 2022, and hopefully this will stimulate the industrial partners to test and evaluate advanced Cross-Cloud application management as it can provide significant performance improvements and cost savings for applications with variable resource need over longer execution times.

HPC support for reservoir simulation

This research topic aims to enable faster and more realistic reservoir simulations, which are the cornerstone in the workflow of oil reservoir management, thereby of vital importance to the entire oil and gas sector. The complexity and uncertainty of the subsurface geological properties require large-scale numerical computations that can only be treated with the technique of HPC. Besides simulating oil reservoirs, the same technique and software are also applicable to planning geological CO2 storage, which is becoming increasingly more important for reducing man-made climate changes. Specifically, the researchers in this topic investigate how to improve the various algorithmic and implemental aspects of such computations, so that reservoir simulations can more efficiently use modern computing hardware.

Smart Scalable PCI Express

I/O resources like Non-Volatile Memory express (NVMe), 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.

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. The CXL standard is also backed by all the large hardware vendors such as Intel, AMD, ARM, Nvidia, etc.

Team Members:

Geir Horn, UiO, Xing Cai, Simula Research Laboratory, Tor Skeie, UiO,

Andreas Thune, UiO , Håkon Kvale Stensland, Simula Research Laboratory, Marta Różańska, UiO, Thomas Hansen, UiO, Atle Vesterkjær, NUMAScale, Einar Rustad, NUMAScale, Hugo Kohmann, Dolphin


London weather
London weather

SIRIUS BEACON PROJECTS

SIRIUS Geo-Assistant

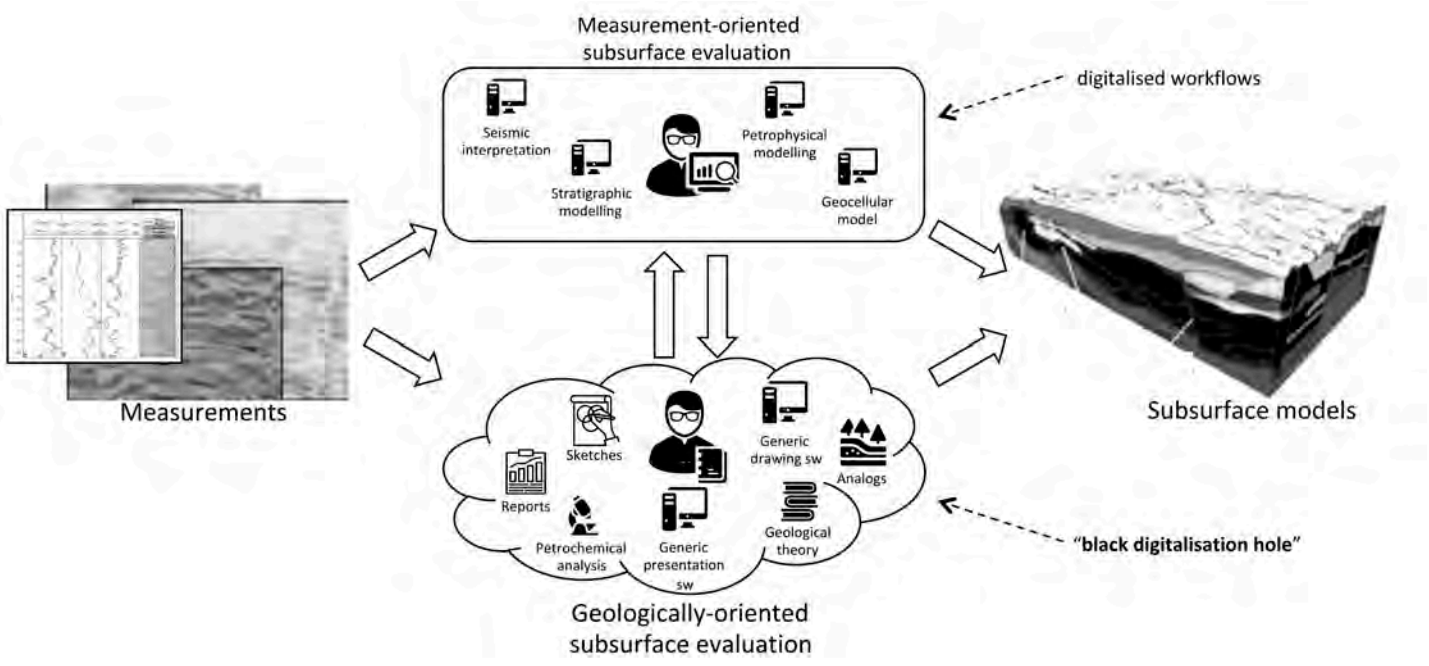
Geological Multi-scenario reasoner (GeMS).

The Geological Assistant is a SIRIUS innovation project between university researchers, Equinor, and SLB.

Interpretation of the subsurface in order to find out where hydrocarbons are located is a challenging task for explorationists. They need to be creative and come up with innovative ideas when defining and assessing new prospects, especially nowa- days when the easy to find, big fields have been already discovered. The challenges related to prospect assessment are (1) the geodata is uncertain, intermittent, sparse, multiresolution, and multi-scale, and (2) the explorationists often limit themselves to assess few possible scenarios

Recent advancements in computation, network and storage have led to numerous opportunities to improve these 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 thesse 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]



Geological Assistant team on field work in the Pyrenees (Aínsa, Spain)

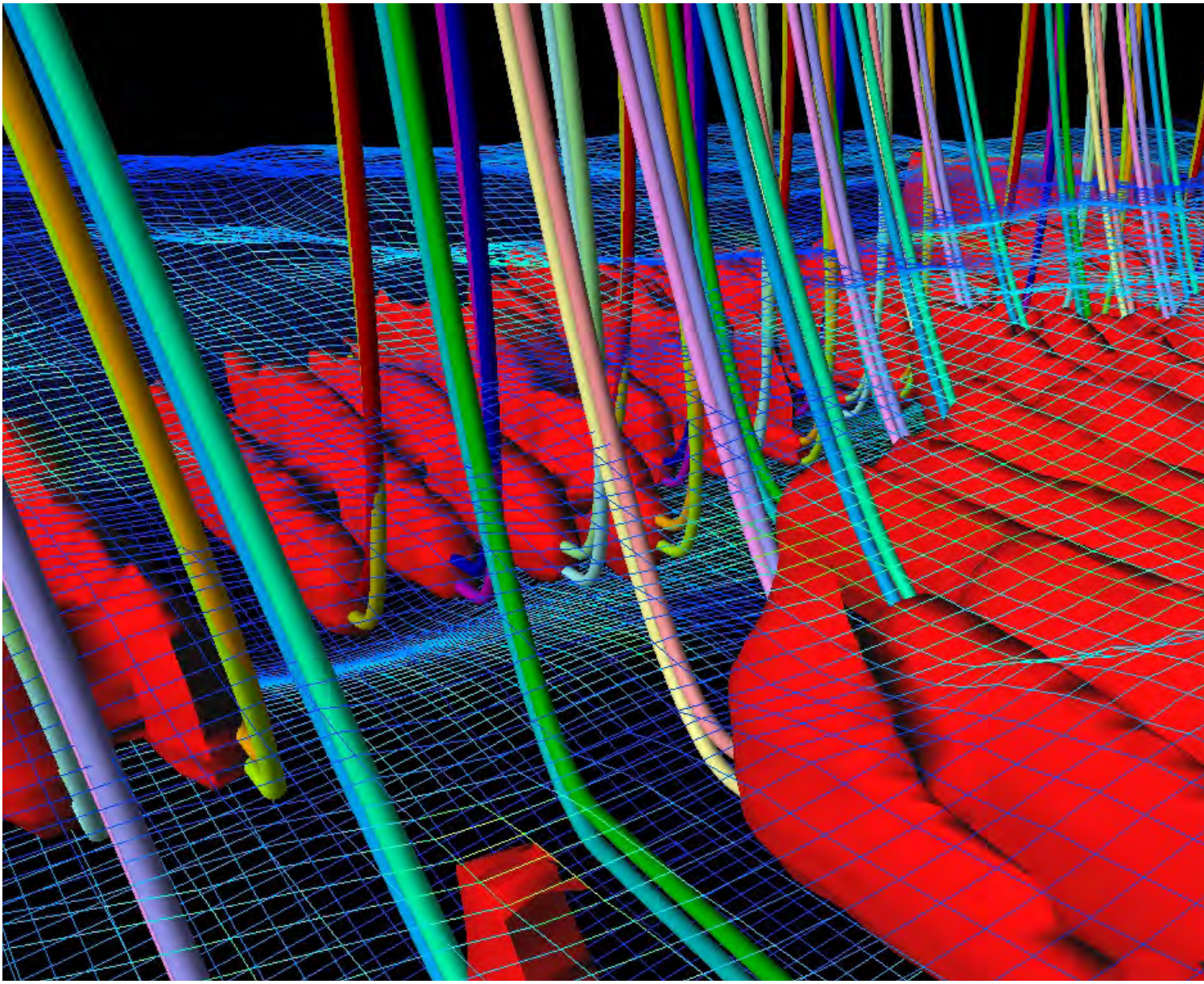
A recent study * of 97 wells drilled in the UK sector of the North Sea from 2003 to 2013 showed that some of the major reasons for unsuccessful exploration were the fact that several of the prospects relied too heavily only on data (seismic DHIs and amplitude). Regional play-based work for setting and context were repeatedly missed, and pre-drill analysis was often excluded the full range of possible outcomes.

The industry's current trend on moving from a physics-driven world to a purely data-driven world for a complex domain like Geoscience has proved inefficient. We believe that a significant success factor will be to combine geological reasoning- based evaluation with the insights derived from the Geological, Geophysical, and Petrophysical data.

The geological multi-scenario reasoning methodology developed in this project provides a hybrid approach by combining the data-driven seismic interpretation (faults and horizons) with geological reasoning (based on the encoded geological rules). It can significantly help subsurface experts to think out of the box to consider several geological models rather than relying on a single model. Further, it will enable Geoscientists to consider a full range of possible scenarios and corresponding outcomes beyond what is possible within human capacity.

We have 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.

* C.Mathieu.MorayFirth–centralnorthseapostwellanalysis.Oil&GasAuthority,2015.



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.

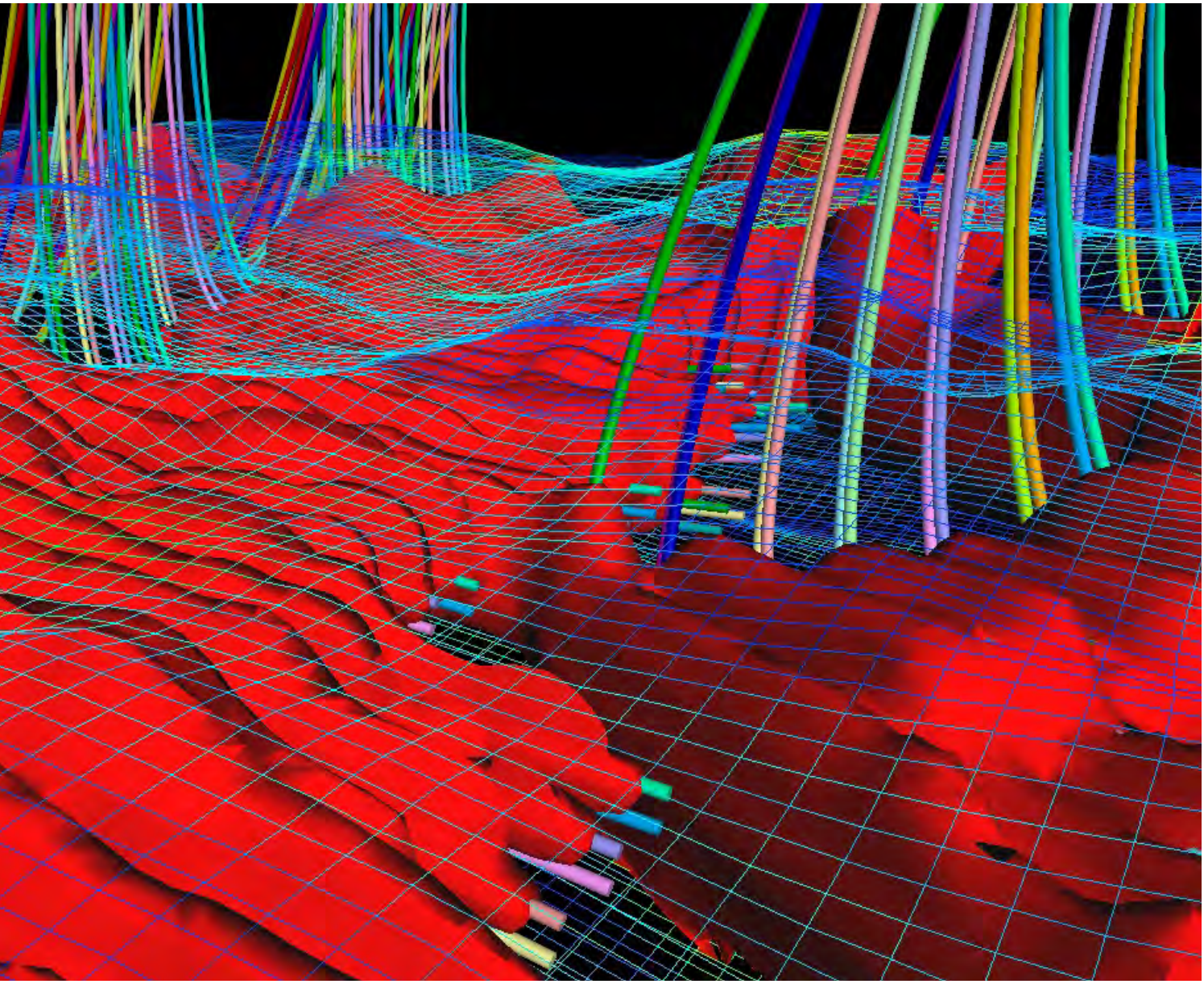
For a detailed information on the developed methodology, system architecture and its application on a real world (Geological) use case, please consult the publications listed in the Results section.

Results

Please see the project webpage for Source Code, Demo videos, posters, and other documentation

Relevant Publications

- Ingrid Chieh Yu, Irina Pene, Crystal Chang Din, Leif Harald Karlsen, Chi Mai Nguyen, Oliver Stahl, Adnan Latif. Subsurface Evaluation through Multi-Scenario Reasoning. In Interactive Data Processing and 3D Visualization of The Solid Earth, Springer (2022).
- Crystal Chang Din, Leif Harald Karlsen, Irina Pene, Oliver Stahl, Ingrid Chieh Yu, and Thomas Østerlie. Geological multi-scenario reasoning. NIK: Norsk Informatikkonferanse (2019).
- Vegar Skaret. Knowledge Representation and Concretization of Underdetermined Data. Master's Thesis (2020).



- Østerlie, Thomas, Elena Parmiggiani, and Eric Monteiro (2017). Information infrastructure in the face of irreducible uncertainty. The 5th Innovation in Information Infrastructure (III) workshop, November 7-9, Rome, Italy.

- Monteiro, Eric, Thomas Østerlie, Elena Parmiggiani, and Marius Mikalsen (2018). Quantifying quality: Towards a post-humanist perspective on sensemaking. In Aanestad, Margunn, Magnus Mffhring, Carsten Østerlund, Kai Riemer, and Ulrike Schultze (Eds.), in Living with Monsters? Social Implications of Algorithmic Phenomena, Hybrid Agency, and the Performativity of Technology (pp. 48-63), Cham, Switzerland: Springer International Publishing.

- Østerlie, Thomas, Elena Parmiggiani, and Petter Almklov (2019). History-based geological modelling: Some elements of a design theory. The 6th Innovation in Information Infrastructure (III) workshop, September 18-20, Guildford, UK

Team

Thomas Østerlie, Chi Mai Nguyen, Irina Pene , Oliver Stahl, Oliver Stahl, Crystal Chang Din, Ingrid Chieh Yu, Leif Harald Karlsen, Adnan Latif [Contact Person], Fabricio Rodrigues, Jens Otten, Michael Heeremans, Vegar Skaret, Eric Monterio, Elena Parmiggiani and Hallgrim Ludvigsen.

Subsurface Data access and analytics

SIRIUS is building on the Optique platform for ontology-based data access to demonstrate how repositories like DISKOS can be developed into digital platforms for exploration, research and innovation. Once this data is opened up, it needs to be analysed. For this reason we are also working with image analysis, data science and natural language applications in sub-surface data management.

Subsurface Data Access

SIRIUS working on a vision of providing a platform for innovation in the sub-surface. A recent book by Andrew McAfee and Erik Brynjolfsson, Machine Platform Crowd traced the role of platforms as enablers for innovation by crowds of workers. We believe that there is a need to open up subsurface data to researchers and innovators to try out their ideas on real data. We also believe that national data repositories, like DISKOS, have the potential to provide such a platform. However, for this to be done, we need to improve access to the data and allow it to be linked with data in other databases. We also need to improve access to unstructured text information in these databases. SIRIUS has several active projects to address the subsurface data access challenges

SIRIUS OBDA Pilot

Exploration digital transformation is about overcoming the bottleneck of data access and increasing the quality of interpretations by means of the better use of data. The data access bottleneck is substantial as up to 70% of exploration experts' time is spent finding, accessing, integrating, and cleaning data before analysis can even start.

One possible approach to address this challenge is to extend the OBDA (Ontology-based data access) theory and tools to support the data access challenges for the subsurface data. OBDA was extended in the Optique project to meet the needs of the oil & gas industry, but the solution has failed to be adopted due to its technological limitations.

The SIRIUS OBDA subsurface pilot project is addressing these shortcomings and aim to significantly broaden the applicability of the approach for use in subsurface projects. See the SIRIUS projects catalogue at Sirius-labs.no to read more about SIRIUS OBDA Pilot project.

In 2022, we developed the V1.0 of the OBDA subsurface pilot. This work includes setting up a large relational database from the publicly available G&G datasets, mappings, ontology, and integration. This Pilot was demonstrated at the SIRIUS GA, ECIM conference, and workshops with partners. Extensive feedback was collected in these sessions for the usefulness and usability of the developed Pilot. This feedback is now used to plan further development in 2023.

GeoDataPrep

Oil and gas companies are transitioning towards more data-driven decision-within the subsurface domain. By visualizing large volumes of complex data through dashboards and other forms of business analytics techniques, decision-makers are to make decisions faster and with greater confidence. However, such data-driven decision-making is moot if the time spent preparing subsurface data for analysis and visualization far exceeds the time saved by decision-makers.

The GeoDataPrep project targets data preparation workflows necessary for dashboarding and business analytics in the subsurface domain.

In 2022, we developed the V1.0 of the prototype in this project to address the naming variability problem (Predicting the mnemonics in the well log data). This project will continue in 2023.

SIRIUS Subsurface Lab

Equinor has made a complete set of data from a North Sea oil field (Volve field) available for research, study and development. This dataset consists of a variety of structured and unstructured subsurface data, comprising approximately 40,000 files from the Volve field which was in production

from 2008 to 2016. The data has been released to give students and scientists a realistic case to study and support learning, innovation and new solutions for the energy future.

The volume of the information available in this dataset is huge (approximately 5TB, 40,000 files), in proprietary and nonproprietary formats and with limited/missing metadata. This makes it challenging to use this dataset for experimentation. A substantial amount of time is required to make it usable (finding the required information from 40,000 files, reading and transforming formats, and compensating the missing meta data etc.). SIRIUS subsurface Lab project is focused on pre-processing the Vovle dataset and creating a sandbox environment for experimentation. See the SIRIUS projects catalogue at Sirius-labs.no to read more about SIRIUS Subsurface Lab project.

In 2023, we will work on improving the architecture of this Lab by building API to access the sandbox and publishing this dataset as semantic web data.

Subsurface Data Analytics

Faster access to relevant data is of interest only if the data can be used to create insight and drive decisions. The Exploration scoping workshop identified several challenges related to information extraction and structured and unstructured data usage. Some of the key areas we are working on.

Structured Data
Geoscientists use Reservoir Analogues to estimate the missing or uncertain reservoir parameters. Finding, selecting appropriate analogues, and extracting inferences depend on the team´s experience and limited human capacity.

To address this problem, a PhD project was started in 2017 and completed in 2021. In this project, techniques are developed to identify and quantify formal domain knowledge, thus predicting more accurate parameters for exploration modelling. The main objective is to extend a Machine learning model that can incorporate Oil & Gas domain information and recommend analogues to a reliable extent. See the SIRIUS projects catalogue at Sirius-labs.no to read more about Hierarchy-based Similarity Measures and Embeddings project.

Unstructured Data – Images
Finding a geological image based on its technical content from a large image database is difficult. Geoscientists use

the keyword search on the textual content of source documents to find relevant Images.

SIRIUS GeoAnnotator

To address this problem, an innovation project was initiated in 2019, and a prototype is developed. This tool supports executing complex queries to find geological images based on the geological content embedded in the images and significantly reduces the time and effort required to find the most relevant images and corresponding documents. See the SIRIUS projects catalogue at Sirius-labs.no to read more about SIRIUS Geo-Annotator project.

Unstructured Data- Documents
Domain experts spend a massive amount of time annotating corpora to train supervised statistical learning models for unstructured data.

To address this problem, a PhD project was started in 2017 and completed in 2021. In this project, techniques are developed to identify and quantify formal domain knowledge, thus predicting more accurate parameters for exploration modelling. The main objective is to extend a Machine learning model that can incorporate Oil & Gas domain information and recommend analogues to a reliable extent. See the SIRIUS projects catalogue at Sirius-labs.no to read more about *Low-Resource Adaptation of Neural NLP Models [PhD Project] project*.

Digital Field and Reservoir Management

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

Better HPC support for reservoir simulation

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

However, using supercomputers for efficient numerical simulations, such as reservoir simulations, is a highly non-trivial topic. Achieving high-performance computing (HPC) requires handling at least three challenges as follows:

- The overall computational work must be appropriately partitioned into pieces, which is the first step for using supercomputers because they are built upon the principle of parallel processing.
- The chosen numerical algorithm for a reservoir simulation to run on a supercomputer must be parallelizable and, moreover, scalable with an increasing number of processing units.
- The software implementation must be able to effectively utilize the capabilities of the processing units and the interconnect between them on a supercomputer.

For reservoir simulations, in particular, the above three challenges are intermingled together. First, the underground reservoirs are in general of irregular 3D shapes, thus unstructured computational meshes must be adopted to resolve the geometric irregularities. Second, partitioning unstructured meshes alone (without considering the actual computations to be executed on top of those) is already a demanding task. The added difficulty for reservoir simulations is that advanced numerical strategies are required to solve the involved mathematical equations. These complex algorithms, when parallelized, may require a different goal for the mesh partitioning task. For example, the mesh entities that are strongly coupled numerically will benefit from being assigned to the same partition, instead of being divided among multiple partitions. Therefore, the mesh partitioning task related to parallel reservoir simulations must balance between the standard partitioning criteria and the resulting effectiveness of the parallelized numerical algorithm. To make the situation even more challenging, the actual utilization of a parallel computer for reservoir simulations is also dependent on the hardware details, as well as the specific reservoir case at hand.

The goals of this SIRIUS project are therefore:

- To investigate the necessary balance between the standard criteria for partitioning unstructured meshes and the reservoir-specific criteria that are important for the parallelized numerical algorithm. Based on this investigation, we want to extend the standard mesh partitioning problem.
- To ensure a high level of performance of parallel reservoir simulations by avoiding non-contributing computations that can arise specifically from parallelization.
- To obtain a quantitative understanding of the communication overhead that is associated with any parallel implementation of a reservoir simulation, for the possibility of a better mapping of the different pieces to the actual hardware processing units.
- To avoid unnecessary waste of the computational effort by automatically identifying good choices of the inherent parameters of the numerical algorithm, so that accuracy is maintained while the numerical algorithm progresses faster than when less optimal parameters are used.

Most reservoir simulators are proprietary systems. This has hindered swift development of HPC techniques. The OPM initiative (https://opm-project.org), however, represents a

new community effort to promote openness in reservoir simulation research. OPM makes a number of real-world data sets openly available and coordinates collaborative development of open-source software for simulating porous media processes, like reservoirs. An extensive software framework has already been developed in OPM.

Two SIRIUS PhD projects (Andreas Thune & Erik Sæternes, supervised by Xing Cai (Simula), Alf Birger Rustad (Equinor) and Tor Skeie (Simula)) aim to enhance the HPC capability of OPM's reservoir simulator. In these projects, we rigorously tested and profiled the current reservoir simulator in the Open Porous Media (OPM) framework using an open data set. The numerical and parallel performance of the simulator, as well as the underlying code details, were carefully studied. See the SIRIUS projects catalogue at Sirius.labs.no to read more about *HPC support for reservoir simulation [2 PhD Projects].*

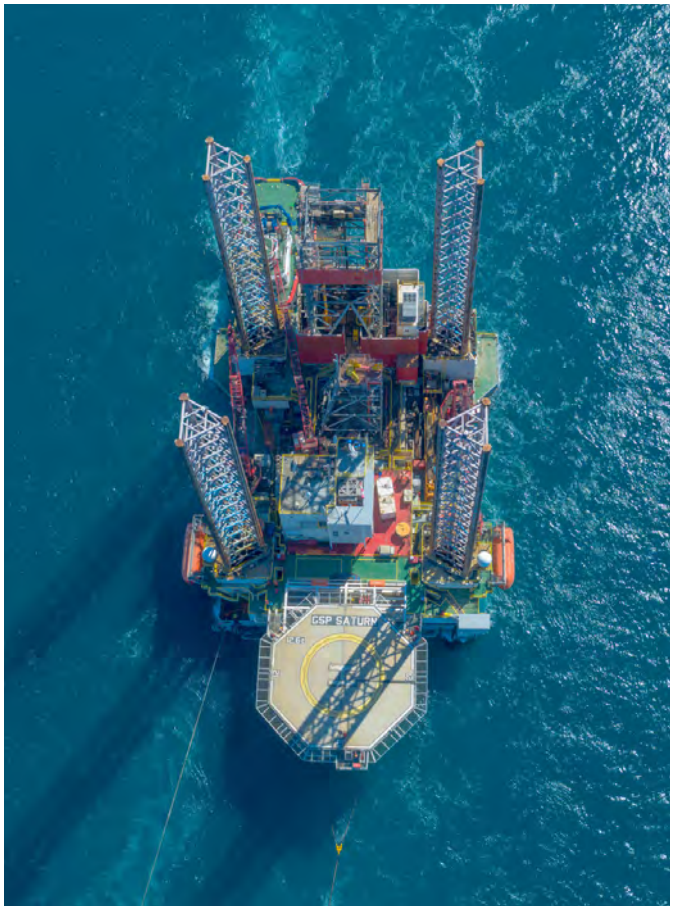
These two PhD projects are progressing as planned. The first subproject has published three papers in 2022, and two additional papers are in progress. This subproject is expected to be finished in early 2023. The second PhD project is also on track to publish results, and the PhD candidate is expected to defend his thesis in 2023.

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

Digital Integrated Field Management

SIRIUS has been working with the field management and petroleum technology groups in Equinor and Petrobras to identify IT challenges in digitalization of field development. Complex greenfield developments, like Libra and Johan Sverdrup, and brown-field projects, like Snorre, face difficult field management problems. Improvements in production and recovery require combination of petrotechnical analysis and data science. These analyses build upon a set of data and models that together provide a digital twin of the whole field: reservoir, wells, flowlines and production facilities. The components of this twin are commercially available but are locked into commercial and technical silos. Retrieval of data is time-consuming and requires database skills. Integration solutions that work for a single application cannot be rolled out to other facilities. Differences in data and application landscape make it difficult to collaborate between disciplines for a single field and within disciplines for different fields.

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



Analysis of Digital Twins

Einar Broch Johnsen

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.

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, and feeds technology to the PeTWIN and other Digital Twins projects.

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.
- 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 the PeTWIN demonstration cases.

Progress in 2022:

A prototype domain specific languages SMOL has been developed, which supports high-level ways to program the

Vision: Analysis support for digital twins, by building or combining tools which can leverage the information layer into insights

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. In 2022, we have integrated asset models into SMOL and used them to automatically configure and reconfigure digital twin simulators. Several papers have been published on SMOL and a two-day tutorial was given at the PhD-school of ICTAC 2022.

Researchers

Einar Broch Johnsen (project leader)
David Cameron
Martin Giese
Geir Horn
Eduard Kamburjan
Vidar Norstein Klungre
Rudolf Schlatte
Silvia Lizeth Tapia Tarifa

Selected Publications for Further Reading

- Eduard Kamburjan, Vidar Norstein Klungre, Rudolf Schlatte, Einar Broch Johnsen, Martin Giese: Programming and Debugging with Semantically Lifted States. In: ESWC 2021. Springer 2021.
- Eduard Kamburjan, Egor V. Kostylev: Type Checking Semantically Lifted Programs via Query Containment under Entailment Regimes. Description Logics 2021. CEUR Workshop Proceedings 2021.
- Eduard Kamburjan, Einar Broch Johnsen: Knowledge Structures Over Simulation Units. ANNSIM 2022. IEEE 2022.
- Eduard Kamburjan, Vidar Norstein Klungre, Rudolf Schlatte, Silvia Lizeth Tapia Tarifa, David Cameron, Einar Broch Johnsen: Digital Twin Reconfiguration Using Asset Models. ISoLA 2022. Springer 2022.
- Geir Horn, Rudolf Schlatte, Einar Broch Johnsen: Digital Twins for Autonomic Cloud Application Management. AINA 2022. Springer 2022.
- Eduard Kamburjan, Vidar Norstein Klungre, Martin Giese: Never Mind the Semantic Gap: Modular, Lazy and Safe Loading of RDF Data. ESWC 2022. Springer 2022.

Digital Design Basis

Prototyping a Shared Data Model for Early-phase Field Development | The Digital Design Basis project concluded this year and provided valuable experience for further work in the READI and IMF projects. The common data model, written using OWL and RDF was used to demonstrate how design basis information from an operator could be made available to engineering applications used by different suppliers. The project brought together Lundin Energy, AkerBP, Equinor, Aker Solutions, TechnipFMC and Aize. The work was organized as a SIRIUS Innovation project. The project results are freely available at the SIRIUS web site. Results were presented at the Advances in Process Digitalization conference and will be published in Digital Chemical Engineering.

David Cameron

The Digital Design Project started in 2019 and concluded in mid 2021. Since the end of the project, we have worked on preparing the results for publication and building on its results in the READI and IMF projects. We have developed and demonstrated a common digital model representation of the information in early-phase design bases for oil & gas field developments. The scope of the project was to develop a proof of concept for a Digital Design Basis that supports data-centric rather than document-based engineering.

The project established a standards-based data model that holds data about both the design basis and functional requirements decided by an operator. This model that can be implemented in any relevant software tools in a concept study, to ensure that information shared between operators and EPC vendors, with their different software tools, have the same meaning and understanding. The model is based on a common digitalized language for communication along the field development supply chain.

Semantic modelling made this representation possible and allowed data to be entered in a structured way and be consumed by engineering applications. We have validated the basic approach, which builds on reusing existing semantic models where possible. We have also demonstrated the feasibility of mixing the modelling approaches defined by ISO15926 and ISO/IEC81346.

We believe the industry needs to have more projects like this, where consortia along the supply chain work with academia and software vendors to agree on interoperability standards by working on real, non-trivial problems. Fortunately, it appears that the European Union, World Economic Forum, and International Organization of Oil & Gas Producers agree with this goal.

Our approach here is not restricted to the oil & gas industry. The system breakdown and modelling of fluid properties can be extended straightforwardly to chemical, fine chemicals, and energy applications. A good first step would be an extension of the RDS for Oil & Gas to ensure that it covers the unit operations in these other domains. We are working further with the READI partners to do this.

This work was experimental, where we were seeking to prove that recent advances in system modelling and ontologies could be applied to a real design basis problem. This meant the integrations with tools tended to be pragmatic rather than user-friendly and scalable. Further work is needed to provide the tools that are necessary to integrate models like this into engineering workflows.

Semantic technology tools are too low-level to be used by practising engineers. OTTR templates have addressed some of these usability challenges, but further work must focus on developing a set of tools to simplify configuration of the model and access to data.

A graphical tool is needed for building system-oriented models by selecting nodes and connecting these nodes with topological and semantic relationships. This tool should also allow the configuration of design basis data in a guided, but

flexible sequence. This interface can exploit the semantic content of the model to provide flow and check consistency.

This interactive tool must be supplemented with tools that allow data to be entered into the DDB in bulk, using tabular data. OTTR provides some of this functionality, but this needs to be lifted up into interactive tools. Mappings need to be developed towards common process simulators and engineering design databases. We need to both read design variables from the DDB and write calculated results back to the DDB. Here we need to work together with the industry so that our models and vendor's models converge over time to an actual or de-facto standard. We are cautiously optimistic about the possibilities of this being successful. We see that many influential vendors are interested in exposing their data using semantic schemas and open formats.

We should aim for work practices where a DDB harvests data from engineering tools without intervention from the engineer. The proof-of-concept has also helped us to develop a more systematic approach to defining the digital design basis. Time and organizational constraints meant that this first effort was more inductive than deductive. The modelling was driven by the data we had to represent, and we then drew systematic conclusions from the solutions developed.

The lessons from this project have been taken up in further, ambitious initiatives by each of the partners. We are contributing to the revisions of READI IMF and RDS for Oil & Gas. This work aims to address the tooling challenges above in the context of several on-going field development projects. It is also developing a formal systematization of modelling and use of data in engineering projects. The results of this work have also been taken into the development of the forthcoming Part 14 of the ISO15926 standard. We hope that these initiatives together will provide elements for establishing a practical, scalable framework for sharing information in the process engineering sector.

PeTWIN

Collaboration with Brazil around Digital Twins in Oil & Gas | The PeTWIN project was finally able to kick off at the end of 2021. Work has been made difficult by COVID-19, as we have not been able to collaborate in person with our partners in USA and Brazil. However, we have made good initial progress, with three post-doctoral researchers and a PhD fellow in place. We held three workshops and a course, held by Shell, on agile development. The work in Oslo is looking at the combination of semantics with formal simulation of cyber-physical systems, linkages to knowledge representation in the READI and IMF projects, and semantic modelling of subsurface in production.

PeTWIN is a Petromaks Brazil project, financed by the Research Council of Norway. It is part of a program where the Research Council collaborates with FINEP, the Brazilian national innovation funding organization, to finance projects with Brazilian and Norwegian partners. PeTWIN's Brazilian Partners are the Federal University of Rio Grande do Sul (UGRGS) and the Libra Field Development Consortium, hosted by Petrobras. The Norwegian Partners are the University of Oslo, Shell and Equinor. The project started in late 2020 and will run until 2024. The project team in Oslo consists of David Cameron, project manager, Martin Giese, scientific leader, three post-doctoral researchers (Eduard Kamburjan, Christian Kindermann and Vidar Klungre) and a PhD fellow (Irina Pene). The project team is supplemented by an internally financed PhD fellow (Yuanwei Qu) and a post-doctoral researcher financed by the Digiwell project (Baifan Zhou). Digiwell is a Petromaks project, led by the University of South Eastern Norway, with University of Oslo, Equinor, Kongsberg Digital, SINTEF, MIT and Imperial College as partners. Coordination between PeTWIN and Digiwell is beneficial to both projects, as they have overlapping interests.

The aim of PeTWIN is literally to write the book about digital twins in the oil and gas industry. We plan to publish a textbook on this topic in 2024. We have prepared the table of contents and are negotiating with publishers. We are doing this by working on the fundamentals of digital twins in engagement with applications in Petrobras / Libra, Shell and Equinor. The use cases cover the field management chain, from reservoir monitoring, through production optimization to facility operations and maintenance. Our researchers bring complementary skills to specific aspects of the digital twin puzzle. For example, Eduard Kamburjan is researching how we can link semantic data access and integration with dynamic simulation of cyber-physical systems. He has worked this year with a Modelica library for facility simulation that was provided by Equinor. This work promises new ways to link our semantic technologies to complex digital twin models of facilities in a way that allows us to ensure the quality of the models and their results. Eduard is working also with Vidar Klungre, who provides the semantic technologies skills needed to realize the work.

Irina Pene and Yuanwei Qu are working on semantic modelling at the other end of the production chain, namely the sub-surface. Here we have a linkage to the Exploration beacons, in that the modelling work done by them is coordinated with the semantic models used in that work package. Our hope is that we can demonstrate the use of these models in modelling and using the knowledge from monitoring of fields in production.

Christian Kindermann joined PeTwin late in 2021. He is looking at the application of semantic domain modeling and reasoning to digital twins, drawing on the results of the READI and IMF projects. We have seen that it would be profitable to use the results or the READI and IMF work as a framework for structuring and delivering digital twins. Our vision is still to provide common tools and standard knowledge models that can benefit Libra, Shell and Equinor.


SHELF DRILLING
SCEPTER
H
22
22
22
22
9.3t

Modelling and Standardizing Process Engineering:

DEXPI Plus

In 2022 SIRIUS workers have developed and published a new and effective way of modelling the information in process design for chemical and oil & gas facilities. The approach builds on the Information Modelling Framework (IMF) and integrates existing standards for modelling engineering information. The results have been published, with supporting datasets. The new DEXPI plus standard for process flow diagrams builds on this work.

The Information Modelling Framework (IMF) brings together systems thinking with the organization of data and information in a design basis. Earlier, SIRIUS enabled and sponsored a joint innovation project called Digital Design Basis, where Lundin Energy, AkerBP, Aker Solutions, Aibel and TechnipFMC worked on creating a prototype digital representation of early-phase design data.

In 2022 this work was taken further by considering how the IMF could be applied to organizing the information and drawings used in the process design of a facility, such as an oil platform. We started with the ISO/IEC81346 Reference Designation System for oil and gas defined by the READI joint industry project. This system defines types of systems that can be used to design and specify a process facility, classified into high-level "Technical Systems" and low-level "Component Systems".

We built on this system by aligning the list with standard types of process equipment and unit operations. For equipment, this can be found in existing standards such as DEXPI, CFIHOS and ISO15926. These have definitions for equipment items such as butterfly valves, centrifugal pumps and pressure vessels. However, we found that there was no corresponding standard for chemical processes and unit operations, such as separation, heat exchange, chemical reaction, compression or size reduction.

Our main finding is that process plant design can be broken down into two steps. First, we are interested in the process: what we want to happen in the plant. We represent the process by systems that stand for unit operations. Once the process is defined, we can then begin to design the plant that will perform the process. We represent the plant by systems that stand for specific pieces of equipment in the plant. Here we see the idea of aspects described in ISO/IEC81346. Process design corresponds to the product aspect, while plant design corresponds to the product aspect.

The benefit of this approach is that we can separate process requirements from equipment requirements. This allows us also to structure operational data so that we can optimize within process constraints and ensure safe equipment behaviour.

In this work, we build linkages between process engineering and systems engineering disciplines. Our library of systems is modelled using the SysML language. This means that standard systems engineering tools can be used to work with these models. However, these tools are too complex and expensive for use by process engineers. For this reason, SIRIUS is working with Equinor to develop simpler tools to build these models.

An Open Example for Experiments

The work we are doing with oil and engineering companies is confidential. However, the methods and library of systems blocks is something we want to share as much as possible. For this reason, we published and example, with full supporting documentation, that applied our approach to a well-known example process, namely the Tennessee Eastman process.

The paper reference is Cameron, D.B., Waaler, A., Fjøsna, E., Hole, M., Psarommatis, F., 2022. A semantic systems engineering framework for zero-defect engineering and operations in the continuous process industries. Front. Manuf. Technol. 2, 945717.

The DEXPI Plus Standard

We applied this work to modelling the information in Process Flow Diagrams (PFD) and Block Flow Diagrams (BFD) for the NOAKA (now Yggdrasil) field development in the North Sea. At the same time, the DEXPI standardization collaboration (www.dexpi.org) started working one extending their standard for representing Piping & Instrumentation Diagrams towards modelling PFD and BFD. David Cameron from SIRIUS took part in the working group and was the main author of the draft standard, which will be issued for discussion early in 2023. The proposed standard is completely consistent with the IMF approach to modelling systems developed in SIRIUS. This international standard gives a basis for using our approach to simplify how engineering design is done. We plan to extend these models to support integration with process simulators and support for graph-based artificial intelligence.

Cross-Domain Applications

The cross-domain application work package aims to show the adaptability of methods and tools developed by SIRIUS researchers in the oil and gas sector for use in other areas of industry, biomedicine, and environmental applications. In 2022, cross-domain efforts focused on three key areas: Healthcare, Earth Science, and Biodiversity.

- **Transfer to Technology to Healthcare**
  In the past, our emphasis was on semantic integration and semantic mapping tasks, as seen in the earlier BIGMED project, along with formal methods for process planning in health organization logistics. In 2022, we undertook multiple scoping projects to utilize technologies created by SIRIUS researchers in achieving healthcare goals. The following are some examples of these efforts.

- Rudi Schlatte and Einar Broch Johnsen contributed to the development a digital twin concept for preventing the development of global pandemics from local epidemics. The work included modeling the global spread of COVID in the three first months of 2020 to validate the digital twin concept against hospitalisation numbers worldwide, and then the study of hypothetical what-if scenarios to study how different interventions could contribute to slow down the spread of a similar epidemic. The work resulted in a paper which has been published in PNAS: **https://www.pnas.org/doi/10.1073/pnas.2220080120**

- Christian Kindermann has been developing techniques for ontology summarisation and maintenance. The main goal is to help practitioners navigate and interact with large and complex ontologies in the biomedical domain. The idea is to present information on different levels of abstraction with design patterns that can be made more concrete on demand. This hierarchical organisation of information in terms of design patterns also opens possibilities for ontology maintenance and quality control by verifying and validating an ontology's underlying design.

- Violet Ka I Pun has been working on a resource sensitive formal language and a static cost analyser cross-organisational workflow model that can be applied in domains like healthcare. The pathology workflow in the Haukeland hospital has been used as a case study in this work. A related publication is "Cost analysis for a resource sensitive workflow modelling language" **https://doi.org/10.1016/j.scico.2022.102896**

- **Scalable Computing for Earth Observation**
  The Execution Modelling and Analysis group has conducted research on cloud computing, specifically focusing on developing distributed deep learning techniques for analyzing satellite images, emphasizing sea ice detection. The CIRFA SFI managed this project in Tromsø, and Einar Broch Johnsen was a member of the supervision team for a PhD student. The PhD project was successfully completed in 2022.

- **Semantics for Biodiversity**
  The Norwegian Biodiversity Information Centre (Arts-databanken) is working on implementing a Species Traitbank (Egenskapsbanken) by integrating data on Norwegian plants and animals. OTTR templates may potentially serve as an efficient method for supporting the transparent integration and utilization of digital biodiversity data, fostering successful collaboration and data reuse.

- The SIRIUS Ontology Engineering research program is engaged in the challenge of identifying, constructing, and maintaining ontologies for Egenskapsbanken. Additionally, the program is investigating the potential application of OTTR templates in developing the Traitbank.

Other cross-domain scoping activites have been carried out by Silvia Lizeth Tapia Tarifa, her work covers event logs analysis and self-adaptive systems:

- automated analysis techniques using event logs to check properties, e.g., user experience in digital services, with resulting publications **https://link.springer.com/chapter/10.1007/978-3-031-17108-6_16** and **https://ebjohnsen.org/publication/22-edba/22-edba.pdf** (To appear EdbA 2022)

- formal frameworks targeting autonomous underwater vehicles to study safety and uncertainty in self-adaptive systems and robustness of controllers with evolving knowledge-based systems, with resulting publications **https://link.springer.com/chapter/10.1007/978-3-031-19849-6_32** and **https://arxiv.org/abs/2303.09220** (To appear - SEAMS 2023)

SIRIUS' Partners

SIRIUS draws together a consortium of leading industrial organisations across the oil & gas value chain, including operators (Equinor), service companies (SLB, Aibel, Aker Solutions, Bosch, TechnipFMC and DNV) and IT companies (Computas, TietoEvry, Dolphin Interconnect Solutions, IBM, Kadme, Numascale, OSISoft, Billington Process Technology, ONTOPIC, Oxford Semantic Technologies,Prediktor, Envester, metaPhacts GMbH, Bouvet and SAP).

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

Partnership for Small and Medium Enterprises (SMEs)

SIRIUS partners, with a few exceptions, 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 two 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. The Spring 2020 General Assembly for SIRIUS approved a program in which Small and Medium Enterprises could become SIRIUS partners. As a result, SIRIUS welcomed Bouvet ASA as partner from February 2022.

Bouvet ASA

Bouvet ASA is a Scandinavian company that provides consultancy and development services in information technology and integrated communication to public and private sector enterprises. The company offers services in the areas of creative design, communication services, web portals, custom application development, system integration, SAP, business intelligence, application management, and education and training. It serves the commerce, energy, public, retail, banking and finance, and service industries.

In SIRIUS, this partnership will be focused on field development beacons, especially around the Information modelling framework related projects. Bouvet has extensive experience, tools and in-house competency in this domain and will have a key role in driving these projects.

Governance

SIRIUS' General Assembly is the body that makes final decisions in the centre. It consists of one high-level representative from each partner. It held two meetings in 2022: one in May, hosted at Radisson Blu Plaza Hotel in the centre of Oslo and the second in November, hosted at the University of Oslo. The spring meeting focused on centre strategy and results, whereas the autumn meeting was a business meeting to review and approve the work plan for 2023.

The spring meeting was one of the largest general assembly we ever had, turning a general assembly into a major scientific event. More than 100 participants from both industry and academia joined this event. The technical program for this Spring GA was organized to cover the research programs in SIRIUS, with the most focus on how these research programs and industrial applications can cross-fertilize each other.

This program consists of:

- An overview presentation of activities in the research programs
- Research talks from each research program
- Technical workshops for in-depth technical sessions on particular topics.
- Panel discussions on the Value of research and the Voice of youth
- Social session, Meet the researchers.

Executive Committee of the General Assembly

The Executive Committee of the General Assembly held regular meetings in 2022. The executive committee acts on behalf of the General Assembly to exercise regular over-sight over plans and progress for the centre.

This Committee in 2022 consisted of:

- Knut Sebastian Tungland, Equinor, Chairman
- Per Eivind Solum, SLB
- Elisabeth Nøst , Technipfmc
- Terje Pedersen, IBM
- Thomas Østerlie, NTNU
- Einar Rustad, Numascale
- Thomas, Computas

Knut Sebastian Tungland

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

- På Rylandsholm, DNV
- Hugo Kohmann, Dolphin Interconnect Solutions
- Arild Waaler, University of Oslo, with Ingrid, University of Oslo, as deputy

Operations Board/Centre Management

The Operations Board (or Centre Management team) is responsible for the centre's day-to-day operations, definition of the work plan that implements the decisions of the Executive Board. It is chaired by the Operations Manager (Centre Coordinator) and consists of the Centre Leader, Administration Manager, Mentor and Education Coordinator, Work Package Leaders and Research Program Leaders. Operations bord had regular meetings in 2022.

Roles

Role	Name	Affiliation
Chairman of General Assembly	Knut Sebastian Tungland	Equinor ASA
Centre Director	Arild Waaler	University of Oslo
Deputy Centre Director	Ingrid Chieh Yu	University of Oslo
Centre Coordinator	Adnan Latif	University of Oslo
Scientific Coordinator	Evgeny Kharlamov	University of Oslo/BOSCH
Administration Manager	Lise Reang	University of Oslo
Finance Manager	Geir Ulvestad	University of Oslo
Subsurface Beacons	Adnan Latif	University of Oslo
Operations Beacons	David Cameron	University of Oslo
Cross-Domain Beacons	Laura Slaughter	University of Oslo
Strategy and Outreach	Einar Broch Johnsen	University of Oslo
Analysis of Complex Systems	Silvia Lizeth Tapia Tarifa	University of Oslo
Ontology Engineering	Martin G. Skjæveland	University of Oslo
Data Science	Basil Ell	University of Oslo
Semantic Integration	Guohui Xiao	University of Oslo/Ontopic
Scalable Computing	Geir Horn	University of Oslo
Industrial Digital Transformation	Thomas Østerlie	NTNU

List of Staff

Name	Institution/ Funding	Nationality	Period/ Duration	Sex	Main research area
Key Researchers					
Arild Waaler	University of Oslo	Norwegian	From 2016	M	Knowledge Representation
Martin Giese	University of Oslo	German	From 2016	M	Knowledge Representation
Einar Broch Johansen	University of Oslo	Norwegian	From 2016	M	Knowledge Representation
Ingrid Chieh Yu	University of Oslo	Norwegian	From 2016	F	Execution Modelling & Analysis
Ian Horrocks	University of Oxford	British	From 2016	M	Knowledge Representation
Boris Motik	University of Oxford	British	From 2018	M	Databases
Eric Monteiro	NTNU	Norwegian	From 2016	M	Working Practices
Jan Tore Lønning	University of Oslo	Norwegian	From 2016	M	Natural Language
Geir Horn	University of Oslo	Norwegian	From 2016	M	Scalable Computing
Laura Slaughter	University of Oslo	American	From 2016	F	University of Oslo
Evgeny Kharlamov	University of Oslo Bosch	German	From 2018	M	Knowledge Representation
Basil Ell	University of Oslo	German	From 2018	M	Natural Language
Xing Cai	Simula Research	Norwegian	From 2016	M	Scalable Computing
Dirk Hesse	University of Oslo	Norwegian		M	Knowledge Representation
Egor Kostylev	University of Oslo	British	From 2020	M	Knowledge Representation
Soylu Ahmet	University of Oslo	Norwegian		M	Knowledge Representation
Martin Skjæveland	University of Oslo	Norwegian	From 2016	M	Knowledge Representation
David Cameron	University of Oslo	Australian	From 2016	M	Knowledge Representation
Adnan Latif	University of Oslo	Norwegian	From 2017	M	Knowledge Representation
Lewis Keith	University of Oslo	American	From 2018	M	University of Oslo
Stahl Oliver	University of Oslo	German	From 2017	M	Knowledge Representation
Postdoctoral researchers with financial support from the Centre budget					
Dag Hovland	University of Oslo	Norwegian	2016-2022	M	Knowledge Representation
Jens Otten	University of Oslo	German	2016-2022	M	Knowledge Representation
Jiaoyan Chen	University of Oslo	Chinese	2019-2022	F	Knowledge Representation
Chi Mai Nguyen	University of Oslo	Vietnamese	2019-2022	F	Analysis of Complex Systems
Laura Slaughter	University of Oslo	American	From 2016	F	Knowledge Representation
Ernesto Jiménez-Ruiz	University of Oslo	Spanish	From 2016	M	Knowledge Representation
Rudi Schlatte	University of Oslo	Austrian	From 2016	M	SIRIUS Laboratory
Violet Pun	University of Oslo	Macanese	From 2017	F	Analysis of Complex Systems

Name	Institution/ Funding	Nationality	Period/ Duration	Sex	Main research area
Postdoctoral researchers working on projects in the centre with financial support from other sources					
Thomas Østerlie	NTNU	Norwegian	From 2016	M	Industrial Digital Transformation
Lizeth Tapia	RCN FRINATEK	Peruvian	From 2017	F	Analysis of Complex Systems
Vidar Klungre	RCN PETROMAKS	Norwegian	2020-2023	M	Knowledge Representation
Eduard Kamburjan	RCN PETROMAKS	Norwegian	2020-2023	M	Analysis of Complex Systems
Christian Kindemann	RCN PETROMAKS	German	2021-2023	M	Knowledge Representation
PhD students with financial support from the Centre budget					
Sigurd Kittilsen	University of Oslo	Norwegian	2017-2022	M	Analysis of Complex Systems
Marta Rozanska	University of Oslo	Polish	2019-2022	F	Scalable Computing
Ratan Bahadur Thapa	University of Oslo	Nepali	2019-2022	M	Analysis of Complex Systems
Peyman Rasouli	University of Oslo	Iranian	2019-2022	M	Analysis of Complex Systems
Ole Magnus Holter	University of Oslo	Norwegian	2019-2023	M	Knowledge Representation
Yuanwei Qu	University of Oslo	Chinese	2019-2023	M	Digital Geosciences
Erik Hide Sæternes	Simula	Norwegian	2020-2023	M	Scalable Computing
PhD students working on projects in the centre with financial support from other sources					
Chinmayi Prabhu Baramashetru	University of Oslo	Indian	2020-2023	F	Analysis of Complex Systems
Irina Pene	RCN PETROMAKS	Rumanian	2020-2022	F	Digital Geosciences
Master degrees					
Justyna Ozog	University of Oslo		2021-2022	F	Domain adoped Data science
Eivind Grønlie Guren	University of Oslo		2020-2021	M	Domain adoped Data science
Rohullah Akbari	University of Oslo		2022-2023	M	Domain adoped Data science
Henrik Syversen Johansen	University of Oslo		2021-2022	M	Domain adoped Data science
Johanna Peersdatter Haarseth	University of Oslo		2021-2022	F	Ontology engineering
Erik Snilsberg	University of Oslo		2021-2022	M	Ontology engineering
Shanshan Qu	University of Oslo		2021-2022	M	Ontology engineering
Anders Søberg	University of Oslo		2022-2023	M	Ontology engineering
Henrik Syversen Johansen	University of Oslo		2022-2023	M	Ontology engineering
Marianne Santos Alvær	University of Oslo		2021-2022	M	Ontology engineering
Ahmet Abbas	University of Oslo		2021-2022	M	Ontology engineering

Publications in 2022

Abusdal, Ole Jørgen; Kamburjan, Eduard; Pun, Violet Ka I; Stolz, Volker. A Notion of Equivalence for Refactorings with Abstract Execution. Lecture Notes in Computer Science (LNCS) 2022 ;Volum 13702. s.259-280 UiO HVL

Ahrendt, Wolfgang; Beckert, Bernhard; Bubel, Richard; Johnsen, Einar Broch. The Logic of Software. A Tasting Menu of Formal Methods - Essays Dedicated to Reiner Hähnle on the Occasion of His 60th Birthday. Springer 2022 (ISBN 978-3-031-08165-1) 519 s. Lecture Notes in Computer Science (LNCS)(13360) UiO

Aminifar, Amin; Matin, Shokri; Rabbi, Fazle; Pun, Violet Ka I; Lamo, Yngve. Extremely Randomized Trees With Privacy Preservation for Distributed Structured Health Data. IEEE Access 2022 ;Volum 10. s.6010-6027 UiO HVL UiB

Baramashetru, Chinmayi; Tapia Tarifa, Silvia Lizeth; Owe, Olaf; Gruschka, Nils. A Policy Language to Capture Compliance of Data Protection Requirements. Lecture Notes in Computer Science (LNCS) 2022 ;Volum 13274. s.289-309 UiO

Cameron, David B.; Waaler, Arild Torolv Søetorp; Fjøsna, Erlend; Hole, Monica; Psarommatis, Foivos. A semantic systems engineering framework for zero-defect engineering and operations in the continuous process industries. Frontiers in Manufacturing Technology 2022 UiO

Cameron, David; Skogvang, Arnljot; Fekete, Mihaly; Martinsson, Henrik; Strand, Morten; Castro, Marcel; Helle, Kirsten; Gjerver, Anders; Hansen, Christian Mahesh; Johnsen, Kristin Høines. The Digital Design Basis. Demonstrating a framework to reduce costs and improve quality in earlyphase design. Digital Chemical Engineering 2022 ;Volum 2.(100015) UiO

Damiani, Ferruccio; Hähnle, Reiner; Kamburjan, Eduard; Lienhardt, Michael; Paolini, Luca. Variability Modules. Journal of Systems and Software 2022 ;Volum 195. UiO

d'Avila Garcez, Artur; Jimenez-Ruiz, Ernesto. Preface. CEUR Workshop Proceedings 2022 ;Volum 3212. UiO

Dongzhuoran, Zhou; Zhou, Baifan; Zhuoxun, Zheng; Ahmet, Soylu; Ognjen, Savkovic; Egor V., Kostylev; Evgeny, Kharlamov. ScheRe: Schema Reshaping for Enhancing Knowledge Graph Construction. 31st ACM International Conference on Information and Knowledge Management (CIKM); 2022-10-17 - 2022-10-21 UiO OSLOMET

Ghadrdan, Maryam; Cameron, David B.. An Overview of the Evolution of Oil and Gas 4.0. I: Industry 4.0 Vision for Energy and Materials: Enabling Technologies and Case Studies. John Wiley & Sons 2022 ISBN 9781119695936. s.241-268 UiO

Gu, Zhenzhen; Corcoglioniti, Francesco; Lanti, Davide; Mosca, Alessandro; Xiao, Guohui; Xiong, Jing; Calvanese, Diego. A systematic overview of data federation systems. Semantic Web Journal 2022 UiO UiB

He, Yuan; Chen, Jiaoyan; Dong, Hang; Jimenez-Ruiz, Ernesto; Hadian, Ali; Horrocks, Ian. Machine Learning-Friendly Biomedical Datasets for Equivalence and Subsumption Ontology Matching. Lecture Notes in Computer Science (LNCS) 2022 ;Volum 13489. s.575-591 UiO

Horn, Geir; Schlatte, Rudolf; Johnsen, Einar Broch. Digital Twins for Autonomic Cloud Application Management. I: Advanced Information Networking and Applications: Proceedings of the 36th International Conference on Advanced Information Networking and Applications (AINA-2022), Volume 3. Springer 2022 ISBN 978-3-030-99618- 5. s.141-152UiO

Horrocks, Ian; He, Yuan; Chen, Jiaoyan; Antonyrajah, Denvar. BERTMap: A BERT-Based Ontology Alignment System. Thirty-Sixth AAAI Conference on Artificial Intelligence; 2022-06-28 UiO

Horrocks, Ian; He, Yuan; Chen, Jiaoyan; Jimenez-Ruiz, Ernesto; Hadian, Ali. Machine Learning-Friendly Biomedical Datasets for Equivalence and Subsumption Ontology Matching. The Semantic Web –International Semantic Web Conference (ISWC) 2022; 2022-10-16 UiO

Horrocks, Ian; Olivares, Jordi; Cocchi, Valerio; Motik, Boris; Roy, Dylan. The Dow Jones Knowledge Graph. European Semantic Web Conference (ESWC) 2022: The Semantic Web; 2022-05-31 UiO

Johnsen, Einar Broch. Digital Twins: An Emerging Paradigm for Model-Centric Engineering. Keynote at FormaliSE 2022; 2022-05-19 UiO

Johnsen, Einar Broch. Digital Twins: An Emerging Paradigm for Model-Centric Engineering. Keynote at PNSE 2022; 2022-06-20 UiO

Johnsen, Einar Broch. Digital Twins as Evolving Model-Centric Systems. Invited talk in the Engineering Digital Twins Seminar Series; 2022-05-02 UiO

Johnsen, Einar Broch. Digital Twins: From Engineering to Science. Distinguished Lecture Series at UiB; 2022-02-17 UiO

Johnsen, Einar Broch; Wimmer, Manuel. Fundamental Approaches to Software Engineering - 25th International Conference, FASE 2022, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2022, Munich, Germany, April 2-7, 2022, Proceedings. Springer 2022 (ISBN 978-3-030-99428-0) 346 s. Lecture Notes in Computer Science (LNCS)(13241)UiO

Kamburjan, Eduard; Din, Crystal Chang; Schlatte, Rudolf; Tapia Tarifa, Silvia Lizeth; Johnsen, Einar Broch. Twinning-by-Construction: Ensuring Correctness for Self-adaptive Digital Twins. Lecture Notes in Computer Science (LNCS) 2022 ;Volum 13701. s.188-204 UiO UiB

Kamburjan, Eduard; Johnsen, Einar Broch. Knowledge Structures Over Simulation Units. I: Annual Modeling and Simulation Conference, ANNSIM 2022. IEEE Press 2022 ISBN 978-1-71- 385288-9. s.78-89 UiO

Kamburjan, Eduard; Klungre, Vidar Norstein; Giese, Martin. Never Mind the Semantic Gap: Modular, Lazy and Safe Loading of RDF Data. Lecture Notes in Computer Science (LNCS) 2022 ;Volum 13261. UiO

Kamburjan, Eduard; Klungre, Vidar Norstein; Schlatte, Rudolf; Tapia Tarifa, Silvia Lizeth; Cameron, David; Johnsen, Einar Broch. Digital Twin Reconfiguration Using Asset Models. Lecture Notes in Computer Science (LNCS) 2022 ;Volum 13704. s.71-88 UiO

Kamburjan, Eduard; Rama Fiorini, Sandro. On the Notion of Naturalness in Formal Modeling. Lecture Notes in Computer Science (LNCS) 2022 UiO

Kamburjan, Eduard; Wasser, Nathan. The Right Kind of Non-Determinism: Using Concurrency to Verify C Programs with Underspecified Semantics. Electronic Proceedings in Theoretical Computer Science (EPTCS) 2022 UiO

Kaminski, Mark; Kostylev, Egor; Grau, Bernardo Cuenca; Motik, Boris; Horrocks, Ian. The Complexity and Expressive Power of Limit Datalog. Journal of the ACM 2022 ;Volum 69.(1) UiO

Karami, Farzane; Basin, David; Johnsen, Einar Broch. DPL: A Language for GDPR Enforcement. I: 2022 IEEE 35th Computer Security Foundations Symposium (CSF). IEEE Press 2022 ISBN 978-1- 6654-8417-6. s.112-129 UiO

Khaleghian, Salman; Ullah, Habib; Johnsen, Einar Broch; Andersen, Anders; Marinoni, Andrea. AFSD: Adaptive Feature Space Distillation for Distributed Deep Learning. IEEE Access 2022 ;Volum 10. s.84569-84578 UiO UiT

Kindermann, Christian; Skjæveland, Martin G. A Survey of Syntactic Modelling Structures in Biomedical Ontologies.. Springer 2022 (ISBN 978-3-031-19432-0) ;Volum 13489.865 s. Lecture Notes in Computer Science (LNCS)(1) UiO

Kobialka, Paul; Tapia Tarifa, Silvia Lizeth; Bergersen, Gunnar R.; Johnsen, Einar Broch. Weighted Games for User Journeys. Lecture Notes in Computer Science (LNCS) 2022 ;Volum 13550. s.253-270 UiO

Lin, Tengteng; Chen, Qiaosheng; Cheng, Gong; Soylu, Ahmet; Ell, Basil; Zhao, Ruoqi; Shi, Qing; Wang, Xiaxia; Gu, Yu; Kharlamov, Evgeny. ACORDAR: A Test Collection for Ad Hoc Content-Based (RDF) Dataset Retrieval. I: SIGIR '22: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, July 11 - 15, 2022. Association for Computing Machinery (ACM) 2022 ISBN 978-1-4503-8732-3. s.2981-2991 OSLOMET UiO

Matentzoglu, Nicolas; Balhoff, James P.; Bello, Susan M.; Bizon, Chris; Brush, Matthew; Callahan, Tiffany J.; Chute, Christopher G.; Duncan, William D.; Evelo, Chris T.; Gabriel, Davera; Graybeal, John; Gray, Alasdair; Gyori, Benjamin M.; Haendel, Melissa; Harmse, Henriette; Harris, Nomi L.; Harrow, Ian; Hegde, Harshad B.; Hoyt, Amelia L.; Hoyt, Charles T.; Jiao, Dazhi; Jimenez-Ruiz, Ernesto; Jupp, Simon; Kim, Hyeongsik; Koehler, Sebastian; Liener, Thomas; Long, Qinqin; Malone, James; McLaughlin, James A.; McMurry, Julie A.; Moxon, Sierra; Munoz-Torres, Monica C.; Osumi-Sutherland, David; Overton, James A.; Peters, Bjoern; Putman, Tim; Queralt-Rosinach, Núria; Shefchek, Kent; Solbrig, Harold; Thessen, Anne; Tudorache, Tania; Vasilevsky, Nicole; Wagner, Alex H.; Mungall, Christopher J.. A Simple Standard for Sharing Ontological Mappings (SSSOM). Database: The Journal of Biological Databases and Curation 2022 ;Volum 2022. UiO

Pardo, Raul; Johnsen, Einar Broch; Schaefer, Ina; Wasowski, Andrzej. A Specification Logic for Programs in the Probabilistic Guarded Command Language. Lecture Notes in Computer Science (LNCS) 2022 ;Volum 13572. s.369-387 UiO

Päßler, Juliane; Aguado, Esther; Rezende Silva, Gustavo; Tapia Tarifa, Silvia Lizeth; Hernández Corbato, Carlos; Johnsen, Einar Broch. A Formal Model of Metacontrol in Maude. Lecture Notes in Computer Science (LNCS) 2022 ;Volum 13701. s.575-596 UiO

Petricek, Tomas; van den Burg, Gerrit J. J.; Nazábal, Alfredo; Ceritli, Taha; Jimenez-Ruiz, Ernesto; Williams:, Christopher K. I.. AI Assistants: A Framework for Semi-Automated Data Wrangling. IEEE Transactions on Knowledge and Data Engineering 2022 UiO

Qu, Yuanwei; Zhou, Baifan; Kharlamov, Evgeny; Giese, Martin. Industrial Geological Information Capture with GeoStructure Ontology. First International Workshop on Semantic Industrial Information Modelling (SemIIM); 2022-05-29 - 2022-06-02

UiO

Róaska, Marta; Skrzypek, Pawe; Materka, Katarzyna; Horn, Geir. An Architecture for Autonomous Proactive and Polymorphic Optimization of Cloud Applications. I: Advanced Information Networking and Applications: Proceedings of the 36th International Conference on Advanced Information Networking and Applications (AINA-2022), Volume 3. Springer 2022 ISBN 978-3-030-99618-5. s.567-577UiO

Schlatte, Rudolf. A Note on Idleness Detection of Actor Systems. Lecture Notes in Computer Science (LNCS) 2022 ;Volum 13360. s.437-445 UiO

Schlatte, Rudolf; Johnsen, Einar Broch; Kamburjan, Eduard; Tapia Tarifa, Silvia Lizeth. The ABS simulator toolchain. Science of Computer Programming 2022 ;Volum 223. UiO

Tapia Tarifa, Silvia Lizeth. Locally Abstract Globally Concrete Semantics of Time and Resource Aware Active Objects. Lecture Notes in Computer Science (LNCS) 2022 ;Volum 13360. s.481-499 UiO

Tapia Tarifa, Silvia Lizeth; Proença, José. Formal Aspects of Component Software. 18th International Conference, FACS 2022, Virtual Event, November 10–11, 2022, Proceedings. Springer 2022 (ISBN 978-3-031-20871-3) ;Volum 13712.245 s. Lecture Notes in Computer Science (LNCS)(13712) UiO

Thapa, Ratan Bahadur; Giese, Martin. Mapping Relational Database Constraints to SHACL. Lecture Notes in Computer Science (LNCS) 2022 UiO

Xiao, Guohui; Pfaff, Emily; Prud'hommeaux, Eric; Booth, David; Sharma, Deepak K.; Huo, Nan; Yu, Yue; Zong, Nansu; Ruddy, Kathryn J.; Chute, Christopher G.; Jiang, Guoqian. FHIR-Ontop-OMOP: Building clinical knowledge graphs in FHIR RDF with the OMOP Common data Model. Journal of Biomedical Informatics 2022 ;Volum 134:104201. s.1-10 UiB UiO

Yahya, Muhammad; Zhou, Baifan; Zheng, Zhuoxun; Zhou, Dongzhuoran; Breslin, John G.; Ali, Muhammad Intizar; Kharlamov, Evgeny. Towards Generalized Welding Ontology in Line with ISO and Knowledge Graph Construction. 19th Extended Semantic Web Conference (ESWC); 2022-05-29 - 2022-06-02 UiO OSLOMET

Zheng, Zhuoxun; Zhou, Baifan; Dongzhuoran, Zhou; Ahmet, Soylu; Evgeny, Kharlamov. ExeKG: Executable Knowledge Graph System for User-friendly Data Analytics. 31st ACM International Conference on Information and Knowledge Management (CIKM); 2022-10-17 - 2022-10-21 UiO

Zheng, Zhuoxun; Zhou, Baifan; Zhou, Dongzhuoran; Cheng, Gong; Jimenez-Ruiz, Ernesto; Soylu, Ahmet; Kharlamov, Evgeny. Executable Knowledge Graphs for Machine Learning: A Bosch Case of Welding Monitoring. The 21st International Semantic Web Conference (ISWC); 2022-10-23 - 2022-10-27 UiO OSLOMET

Zheng, Zhuoxun; Zhou, Baifan; Zhou, Dongzhuoran; Cheng, Gong; Jimenez-Ruiz, Ernesto; Soylu, Ahmet; Kharlamov, Evgeny. Query-Based Industrial Analytics over Knowledge Graphs with Ontology Reshaping. Lecture Notes in Computer Science (LNCS) 2022 ;Volum 13384. s.123-128 OSLOMET UiO

Zheng, Zhuoxun; Zhou, Baifan; Zhou, Dongzhuoran; Khan, Akif Quddus; Soylu, Ahmet; Kharlamov, Evgeny. Towards A Statistic Ontology for Data Analysis in Smart Manufacturing. The 21st International Semantic Web Conference (ISWC); 2022-10-23 - 2022-10-27 UiO OSLOMET NTNU

Zheng, Zhuoxun; Zhou, Baifan; Zhou, Dongzhuoran; Soylu, Ahmet; Kharlamov, Evgeny. Executable Knowledge Graph for Transparent Machine Learning in Welding Monitoring at Bosch. I: Proceedings of the 31st ACM International Conference on Information & Knowledge Management, Atlanta, GA, USA, October 17-21, 2022. Association for Computing Machinery (ACM) 2022 ISBN 978-1-4503-9236-5. s.5102-5103UiO OSLOMET

Zhou, Baifan; Pychynski, Tim; Reischl, Markus; Kharlamov, Evgeny; Mikut, Ralf.

Machine Learning with Domain Knowledge for Predictive Quality Monitoring in Resistance Spot Welding. Journal of Intelligent Manufacturing 2022 UiO

Zhou, Baifan; Tan, Zhipeng; Zheng, Zhuoxun; Zhou, Dongzhuoran; He, Yunjie; Zhu, Yuqicheng; Yahya, Muhammad; Tran, Trung-Kien; Stepanova, Daria; Gad-Elrab, Mohamed H.; Kharlamov, Evgeny. Neuro-Symbolic AI at Bosch: Data Foundation, Insights, and Deployment. The 21st International Semantic Web Conference (ISWC); 2022-10-23 - 2022-10-27 UiO OSLOMET

Zhou, Baifan; Zheng, Zhuoxun; Zhou, Dongzhuoran; Cheng, Gong; Jimenez-Ruiz, Ernesto; Trung-Kien, Tran; Stepanova, Daria; Gad- Elrab, Mohamed H.; Nikolov, Nikolay; Soylu, Ahmet; Kharlamov, Evgeny. The Data Value Quest: A Holistic Semantic Approach at Bosch. Lecture Notes in Computer Science (LNCS) 2022 ;Volum 13384. s.287-290 OSLOMET SINTEF NTNU UiO

Zhou, Baifan; Zheng, Zhuoxun; Zhou, Dongzhuoran; Tan, Zhipeng; Savkovic, Ognjen; Yang, Hui; Zhang, Yujia; Kharlamov, Evgeny. Knowledge Graph-Based Semantic System for Visual Analytics in Automatic Manufacturing. The 21st International Semantic Web Conference (ISWC); 2022-10-23 - 2022-10-27 UiO OSLOMET

Zhou, Dongzhuoran; Zhou, Baifan; Zheng, Zhuoxun; Kostylev, Egor; Cheng, Gong; Jimenez-Ruiz, Ernesto; Soylu, Ahmet; Kharlamov, Evgeny. Enhancing Knowledge Graph Generation with Ontology Reshaping – Bosch Case. Lecture Notes in Computer Science (LNCS) 2022 ;Volum 13384. s.299-302 OSLOMET UiO

Zhou, Dongzhuoran; Zhou, Baifan; Zheng, Zhuoxun; Soylu, Ahmet; Cheng, Gong; Jimenez-Ruiz, Ernesto; Kostylev, Egor; Kharlamov, Evgeny. Ontology Reshaping for Knowledge Graph Construction: Applied on Bosch Welding Case. The 21st International Semantic Web Conference (ISWC); 2022-10-23 - 2022-10-27 UiO OSLOMET

Zhou, Dongzhuoran; Zhou, Baifan; Zheng, Zhuoxun; Soylu, Ahmet; Savkovic, Ognjen; Kostylev, Egor V.; Kharlamov, Evgeny. ScheRe: Schema Reshaping for Enhancing Knowledge Graph Construction. I: Proceedings of the 31st ACM International Conference on Information & Knowledge Management, Atlanta, GA, USA, October 17-21, 2022. Association for Computing Machinery (ACM) 2022 ISBN 978-1- 4503-9236-5. s.5074-5078 UiO OSLOMET

Zhou, Dongzhuoran; Zhou, Baifan; Zheng, Zhuoxun; Tan, Zhipeng; Kostylev, Egor; Kharlamov, Evgeny. Towards Executable Knowledge Graph Translation. The 21st International Semantic Web Conference (ISWC); 2022-10-23 - 2022-10-27 UiO OSLOMET

Zhuoxun, Zheng; Zhou, Baifan; Zhou, Dongzhuoran; Soylu, Ahmet; Kharlamov, Evgeny. ExeKG: Executable Knowledge Graph System for User-friendly Data Analytics. I: Proceedings of the 31st ACM International Conference on Information & Knowledge Management, Atlanta, GA, USA, October 17-21, 2022. Association for Computing Machinery (ACM) 2022 ISBN 978-1- 4503-9236-5. s.5064-5068 OSLOMET UiO

Annual Accounts

Costs

All figures in 1000 NOK	2015	2016	2017	2018	2019	2020	2021	2022
Personnel and indirect costs	539	5188	10087	12450	20922	25379	22102	16970
Purchase of research services	-	600	2113	6231	6167	7787	5796	1187
Equipment	-	31	122	44	-	-	-	-
Other operational costs	62	9505	10424	16355	17202	17694	11189	16957
Total Sum	**601**	**15324**	**22746**	**35080**	**44291**	**50860**	**39087**	**35114**

Funding

All figures in 1000 NOK	2015	2016	2017	2018	2019	2020	2021	2022
Research Council	-	7173	8180	12430	16961	17805	15792	9766
University of Oslo	601	-864	5596	5604	7721	14358	8048	9788
Public partners	-	144	813	1143	1137	1059	-	-
Private partners	-	8510	7910	15131	17700	17638	15247	15560
International partners	-	361	250	772	772	-	-	-
Total Sum	**601**	**15324**	**22749**	**35080**	**44291**	**50860**	**39087**	**35114**

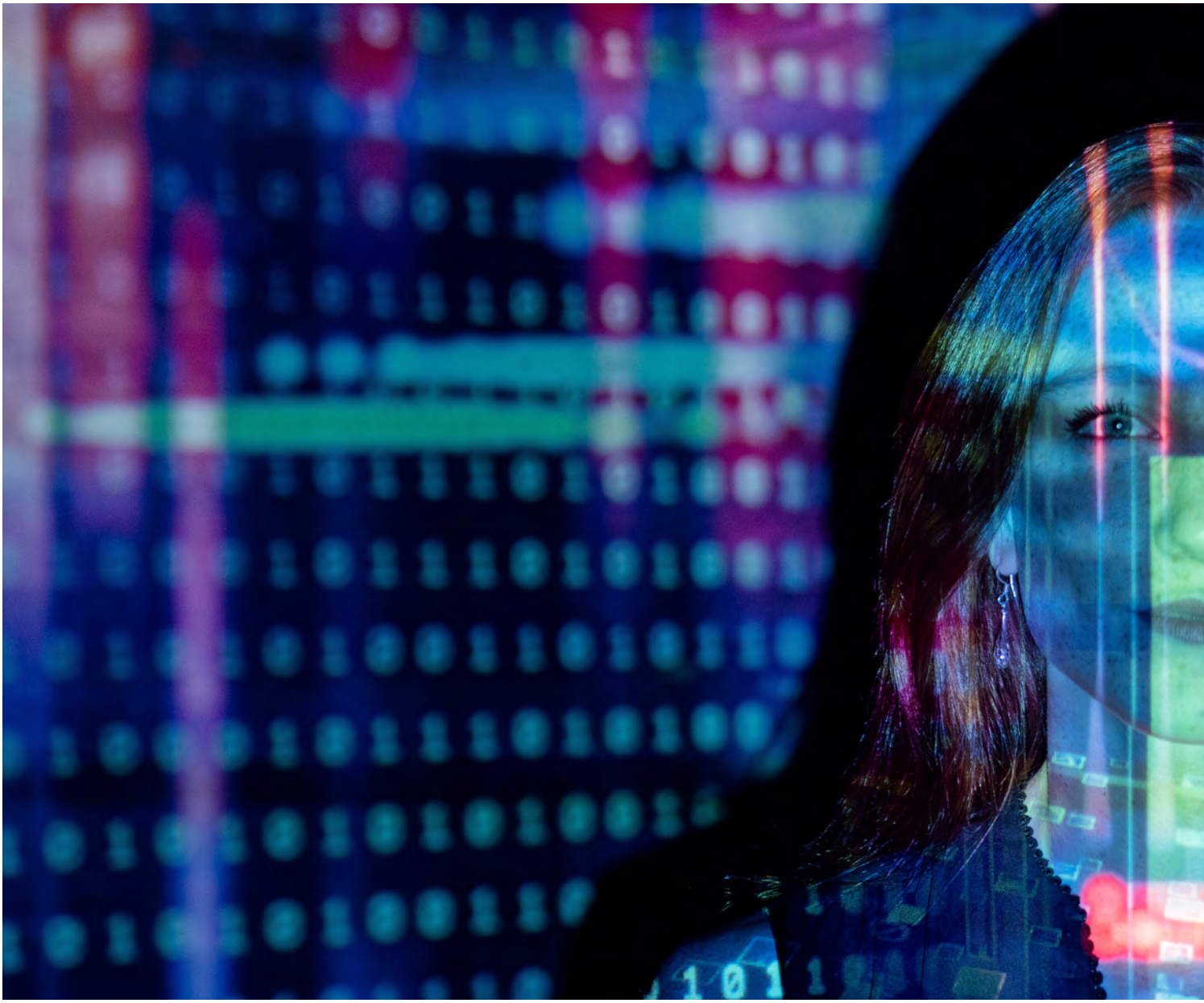
Photo credits
Istock
Unsplash
The Noun Project

Editor
Adnan Latif

Layout
FÆRD AS

SIRIUS

ANNUAL REPORT 2022



Sirius-labs.no