Project Periodic Report

Grant Agreement number: FP7-318338

Project acronym: Optique

Project title: Scalable End-user Access to Big Data

Funding Scheme: Integrated Project

Date of latest version of Annex I against which the assessment will be made: 20 June 2012

Periodic report: 1st ☐ 2nd ☒ 3rd ☐ 4th ☐

Period covered: **from** 1 November 2013 **to** 30 October 2014

Name, title and organisation of the scientific representative of the project's coordinator:

Arild Waaler, Prof. Dr.
Department of Informatics
University of Oslo

Tel: +47 952 12 954

E-mail: arild@ifi.uio.no

Project website address: http://www.optique-project.eu

Declaration by the Scientific Representative of the Project Coordinator

I, as scientific representative of the coordinator of this project and in line with the obligations as stated in Article II.2.3 of the Grant Agreement declare that:

- The attached periodic report represents an accurate description of the work carried out in this project for this reporting period;
- The project (tick as appropriate):
  - ☒ has fully achieved its objectives and technical goals for the period;
  - ☐ has achieved most of its objectives and technical goals for the period with relatively minor deviations;
  - ☐ has failed to achieve critical objectives and/or is not at all on schedule.
- The public website, if applicable
  - ☒ up to date
  - ☐ is not up to date
- To my best knowledge, the financial statements which are being submitted as part of this report are in line with the actual work carried out and are consistent with the report on the resources used for the project (Section **??**) and if applicable with the certificate on financial statement.
- All beneficiaries, in particular non-profit public bodies, secondary and higher education establishments, research organisations and SME's, have declared to have verified their legal status. Any changes have been reported under Section **??** (Project Management) in accordance with Article II.3.f of the Grant Agreement.

Name of scientific representative of the Coordinator: Arild Waaler

Date: 2014/ 07 / 22

Signature of scientific representative of the Coordinator: ______________________

Contents

1 Publishable Summary

Optique™

Scalable End-User Access to Big Data

http://www.optique-project.eu/

UiO : University of Oslo

UNIVERSITY OF OXFORD

TUHH Technische Universität Hamburg-Harburg

SAPIENZA UNIVERSITÀ DI ROMA

Freie Universität Bozen
Libera Università di Bolzano
Free University of Bozen · Bolzano

HELLENIC REPUBLIC
National and Kapodistrian University of Athens

SIEMENS

Statoil

fluidOps

DNV·GL

Optique: Improving the competitiveness of European industry

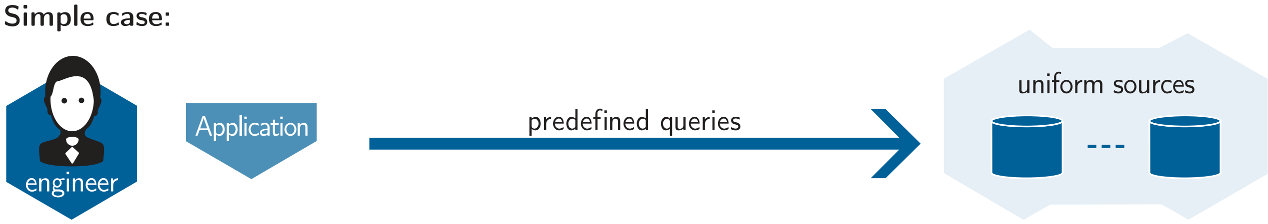
For many advanced end users, accessing the *relevant* data is becoming increasingly difficult due to the explosion in the *size* and *complexity* of data sets.

How much time do engineers in European industry spend searching for data?

Optique targets the key bottleneck limiting exploitation of "Big Data":

- Massive amounts of data are accumulated, in real time and over decades.
- Accessing relevant parts of the data requires in depth knowledge of the domain *and* of the organisation of data repositories.
- End users' domain-specific applications limit data access to a restricted set of predefined queries (see Simple case figure below).

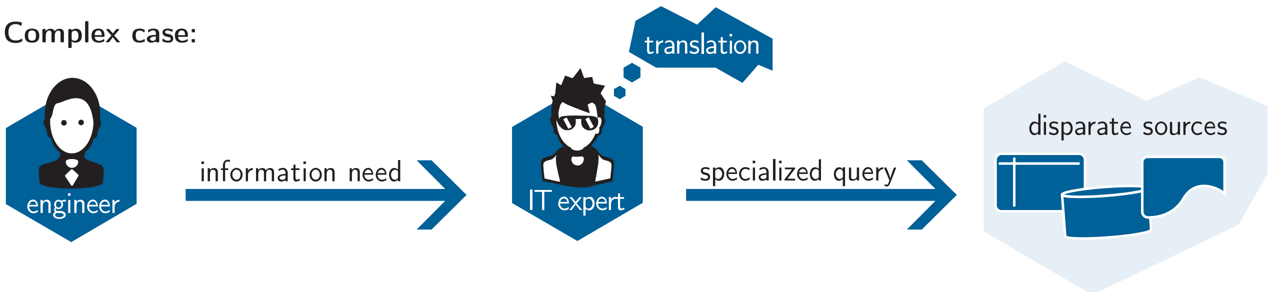
Simple case:



Maximally exploiting data requires flexible access – engineers need to *explore* the data in innovative ways not supported by current applications. This typically requires an IT-expert in order to write special purpose queries and optimise the queries for efficient execution (see Complex case figure below). With this process, accessing the data can take several days. In data-intensive industries, engineers spend up to 80% of their time on data access.

How much value could they create in that time?

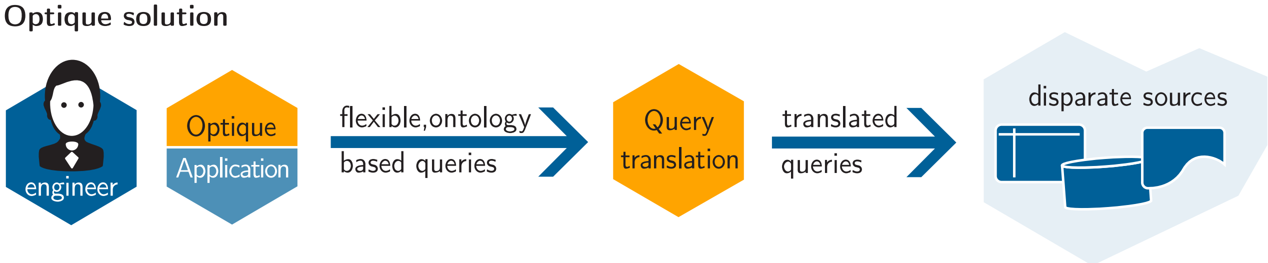
Complex case:



Apart from the enormous *direct cost*, freeing up expert time would lead to even greater *value creation* through deeper analysis and improved decision making.

The goal of Optique is to enable end users to formulate optimised special purpose queries on their own, without direct assistance from IT experts. In order to achieve this the **Optique platform** uses an *ontology* to capture (possibly multiple) user conceptualisations, so as to allow users to formulate queries using their own conceptualisations of the data. These user queries are then transformed into *complete, correct and highly optimised queries* over the data sources, which may include streams:

Optique solution



Development of the Optique platform will be informed by and evaluated against the requirements of *complex real-world challenges*, with Siemens Energy Services and Statoil Exploration providing the project with comprehensive use cases.

The Optique Platform

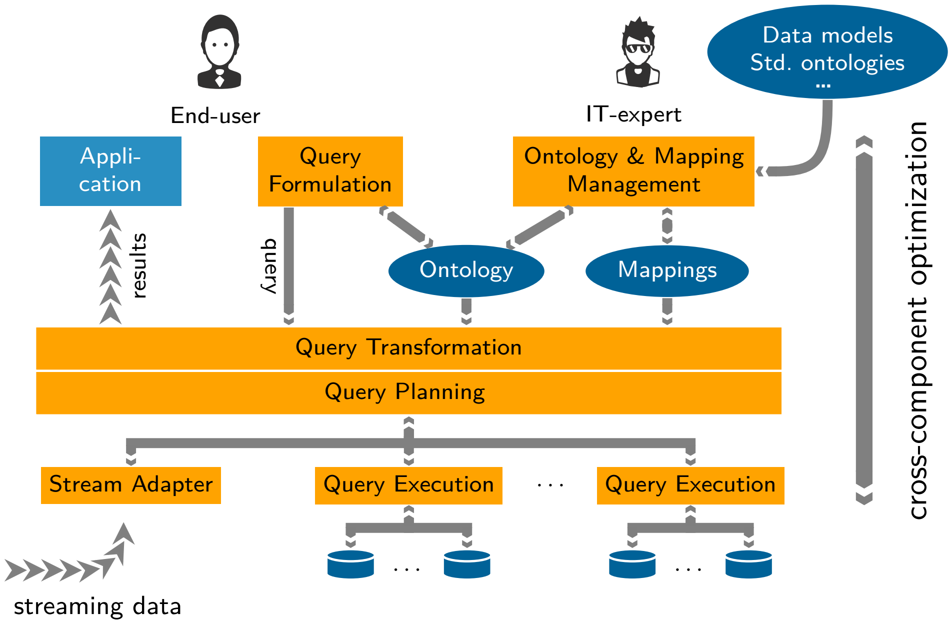


Figure 1.1: Optique platform architecture

The Optique platform exploits recent ground-breaking European research on semantic technologies, in particular related to *query rewriting*, and combines this with techniques for scaling up query evaluation, in particular *massive parallelism*. These are integrated in a comprehensive and extensible platform that builds on open standards and protocols.

The core architecture is illustrated in Figure 1.1. *Front-end scalability* is achieved by abstracting away from data sources, allowing users to construct queries in terms of an ontology that reflects their vocabulary. Intelligent support for query construction is provided by a query formulation component that combines ontology based "query by navigation" with a context-sensitive query editor. Query results are presented to the user via existing tools, with interfaces to these provided by the platform.

While the ontology captures crucial knowledge of the domain, the knowledge of the IT experts is captured by a set of declarative mappings that relate the ontology to the data sources. A novel ontology & mapping management component supports the construction and maintenance of ontology and mappings. This includes support for semi-automatic "lazy" construction driven by the requirements of user queries, minimising the need for *a priori* ontology and mapping development.

Back-end scalability is achieved by optimising the query transformation component that rewrites user queries into queries over the data sources. The query planning component then devises an optimised plan for evaluating these queries, which are handled by a query execution or stream adaptor component for each autonomous source. Further scalability gains are achieved by taking a holistic approach to optimisation that goes beyond considering components individually, which is a distinguishing feature of Optique. Query transformation is optimised so as to produce queries that can be more effectively planned and executed, and query planning and execution are optimised for the kinds of queries that result from query transformation. These optimisations exploit features of typical (user) queries, ontologies and mappings, and the system is tuneable so as to maximise performance in specific applications and even for specific queries.

The platform is non-invasive, in the sense that it runs on existing IT infrastructure and does not require data migration. The source data can be of a wide variety of types (including relational and semantic data formats as well as streams), provided that they can be accessed via structured query interfaces. However, in order to maximise back-end scalability, we also explore an approach in which data *is* migrated to an infrastructure that implements massive parallelisation.

The platform runs against open data sources in a *Public showcase* installation, in addition to the two installations within the premises of Siemens and Statoil.

The Optique Partner Programme: A community of early adopters

Optique targets enterprises with big data challenges, and in particular challenges related not only to large *volumes* of data, but also to schemas and infrastructures of great *complexity*, new data coming in with high *velocity*, or data sources being structured according a large *variety* of schemas and formats.

Big data challeges are rooted deeply in organisations and their workflows; enterprises facing such challenges thus need to address not only new technologies, but also how well they know their data and their maturity for adopting new solutions. The Optique dissemination and exploitation strategy for targeted industry, summarised in Figure 1.2, assumes that a decision to adopt Optique will depend on a successful evaluation of a proof-of-concept implementation. This implementation should target a use case that is sufficiently complex, yet simple enough to be implemented and tested within a short timeframe and without requiring a major organizational effort.

Optique has identified five key phases of an enterprise project seeking to adopt the Optique platform, starting with exploration of introductory material, going through learning and assessment of business opportunities, and then to proof-of-concept implementation and evaluation. Optique delivers a comprehensive suite of resources in order to support the first three of these phases:

- Key personnel have to *observe* the features of Optique to be in position to decide on setting up a proof-of-concept project. Optique provides a proverbial *executive summary*, introductory presentations and *white papers* so as to enable staff to understand how Optique can meet their needs.
- In order to set up teams with the sufficient level of expertise for a proof-of-concept project, the enterprise has to go through a phase of *learning*. Optique delivers one-day *courses* to allow industry staff to explore how Optique can be applied to their specific needs and a *curriculum* with text-book character that can serve as reference material for implementation projects.
- A phase of *appraisal* is needed in order to identify suitable use cases. Optique provides analytic tools to support a focused work plan, leading up to an installed Optique system that delivers company data to expert users. Implementation guidelines and business model templates assist in choosing a use case, finding the right mix of people, and establishing an appropriate work process.

These resources will be freely available. Targeted events are offered to members of the community of early adopters, organised in the *Optique Partner Programme*. The Partner Programme is a framework for arranging binding, predictable agreements and a path to adoption that settles cost, responsibilities, and expectations. Ad hoc collaboration between a research programme and a commercial company is in general far from straightforward to set up. With the Optique Partner Programme, companies can sign up on a level appropriate to their interests, ranging from basic training on to use case evaluation and pilot project work.

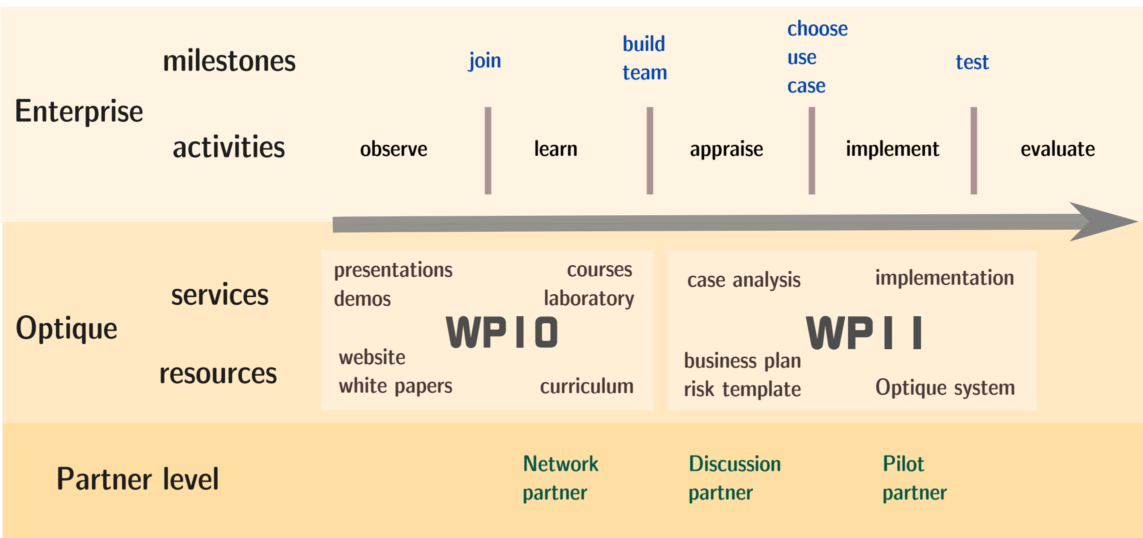


Figure 1.2: The Optique dissemination and exploitation strategy for targeted industry.

Project objectives

Use cases from Siemens Energy Services and Statoil Exploration drive the development of the Optique platform. Optique follows a four phase work plan; each phase starts November 1 and ends October 31 the following calendar year. The main priorities for each project year are outlined in Table 1.1. The overall project objectives are stated in Table 1.2.

Year	Optique Platform and Use Cases	Community Building
1	Early prototype incl. visual query interface – basic architecture – combine existing Background components Build initial Query Library Prototype processes simple queries	Make Optique widely known through: – publications, talks and demonstrations – targeted events in Siemens and Statoil – meetings with industry – development of fact sheets and web site
2	Functional prototype w/all components Build comprehensive Query Library Prototype processes realistic queries	Establish the Optique Partner Programme – targeted dissemination to network partners Release of first version of Public Showcase
3	Fully functional platform – new components fully integrated – Siemens: combine streams & static data – Statoil: federation of several sources	Extensive user training in Siemens/Statoil Partner Programme fully operational – deliver first courses – recruiting Discussion and Pilot Partners
4	Cross-component optimisation – increased usability & performance – additional functionality Most End User information needs covered	Broad showcasing in Siemens/Statoil Release of tutorials and Public Showcase Pilot Partners implementing demos Cultivation of user groups

Table 1.1: Phases of the Optique Work Plan

Main Objective. To design and implement an end-to-end solution to the problem of providing comprehensive and timely access to large scale data sets.	
Objective 1.	Achieve the combination of intelligence, flexibility and scalability needed to meet the requirements of the use case challenges.
Objective 2.	Provide a front end that supports end-users and helps them to formulate queries.
Objective 3.	Extend query rewriting techniques with support for temporal queries and queries over streams.
Objective 4.	Integrate scalable methods for database management into an optimised Ontology-Based Data Access framework.
Objective 5.	Develop the Optique platform in response to requirements from use cases provided by Siemens and Statoil and deploy and evaluate the platform in these use cases.
Objective 6.	Widely disseminate Optique scientific results and technologies to the academic sector, to industry, and to the general public.
Objective 7.	Establish the Optique eco-system in order to disseminate and exploit the project results.

Table 1.2: The Optique project objectives

Project results up to Month 24 (cf. Table 1.2)

Main Objective. At Month 24, a fully functional prototype has been implemented and evaluated—with very promising results—in the two industrial use case scenarios.

Objective 1.	• A generic architecture has been designed that can be adapted to any domain that requires scalable data access and efficient query execution. • Based on this architecture, an initial prototype of the Optique Platform has been developed. • The prototype has been deployed at the use case partners and is able to answer realistic user queries over large datasets (up to 1TB).
Objective 2.	• A stable prototype of the ontology-based visual query formulation subsystem has been implemented, based on multiple coordinated interaction paradigms, in particular graph navigation and facet refinement. Novel techniques to support query formulation have been developed. • Use case evaluations were carried out both with students and endusers from Statoil and Siemens with encouraging results showing that endusers can formulate realistic queries with mild training.
Objective 3.	• STARQL—a temporal and streaming extension of the SPARQL query language—has been developed, and an analysis of the architectural requirements for the time and streams subsystem has been carried out.
Objective 4.	• A prototypical ontology and mapping management subsystem has been implemented; it provides basic and advanced bootstrapping wizards, and supports ontology bootstrapping from schema constraints, ontology integration and novel ontology approximation techniques. • A query rewriting subsystem has been implemented, based on the Ontop system; it supports OWL 2 QL, R2RML mappings, and most of the SPARQL query language, and improves performance by exploiting mappings and data dependencies. • A JDBC interface has been implemented for the ADP-based distributed query execution subsystem, and queries imported from use-cases have been used to benchmark the subsystem and its component algorithms.
Objective 5.	• Relevant user interface components, such as those for visual query formulation and browsing, have been presented to groups of end users at use-case workshops. • Feedback suggests that the overall approach is promising and that end-users are able to use the system.
Objective 6.	• Approx. 100 refereed publications, more than 100 presentations for the research community, more than 100 presentation for industry • The Optique Public Showcase, with a queryable data endpoint and sample queries and mappings. • An Executive White paper. • Several releases of the Query Transformation component as part of the open source ontop project.
Objective 7.	• The Optique Partner Program has been established. • An initial partner event with more than 20 companies have been arranged. • Partner events for 2015 have been scheduled.

2 Project Objectives for the Period

Structure of the work packages

The project development in Optique is driven by a push-pull process, the pull coming from end users in Siemens and Statoil, the push coming from the S&T partners. In terms of the work plan:

- Requirements from end users in Siemens and Statoil are collected, compiled, and turned into specifications to the research work packages, WP3–WP7.
- Novel and innovative ideas, implemented in the research work packages WP3–WP7, are demonstrated for the end users and evaluated by them.

The interplay between these two forces is managed in WP1, covering both requirements analyses, evaluation, and the overall scientific management of the project.

Schematically the work packages in Optique are categorised in five groups: Research (WP3–WP7), Integration (WP2), Use Case (WP1, WP8, WP9), Impact (WP10, WP11), and Management (WP12).

- The Research, Integration and Use Case groups are interconnected to form a cycle where the outcome of the Research WP's are integrated by WP2 and put to use in the Use Case WP's, which again gives input to the Research group, as shown in more detail in Figure 2.2.
- The combined outcome of these three groups acts as input to the Impact group which will refine the results for public dissemination and exploitation.
- Project Management ensures overall progress, but is not connected to the other groups in other ways.

The relationship and dependencies between the groups are illustrated in Figure 2.1.

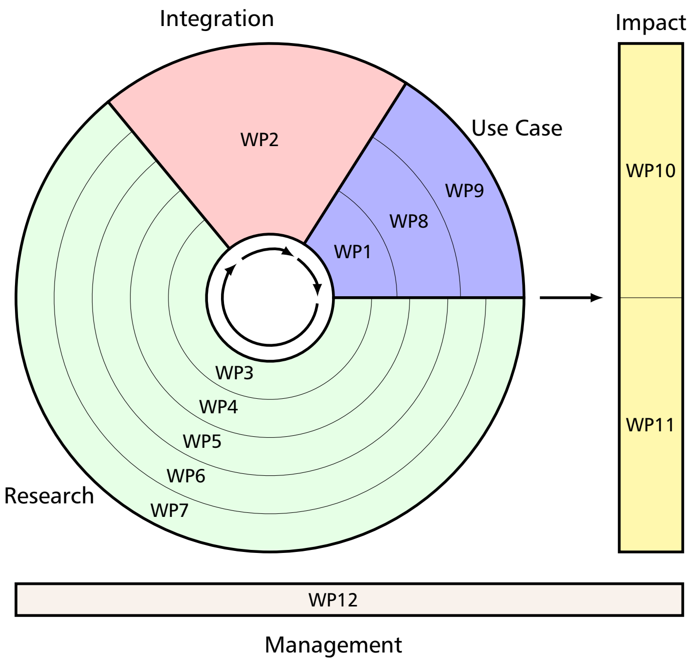


Figure 2.1: Relationships and dependencies between the five WP groups

Context of the Reporting Period within the Work Plan

Development of the Optique Platform

The Optique Platform is developed following four 1-year iterations of a complete development cycle; the yearly development cycle is illustrated in Figure 2.2. Work Packages 8 and 9 supply requirements to Work Package 1, which informs the research work packages (Work Package 3–7). Work Package 2 integrates the results of Work Packages 3–7, and the new version of the platform is installed, tested and evaluated by the industry partners in Work Packages 8 and 9, resulting in new requirements for the next cycle.

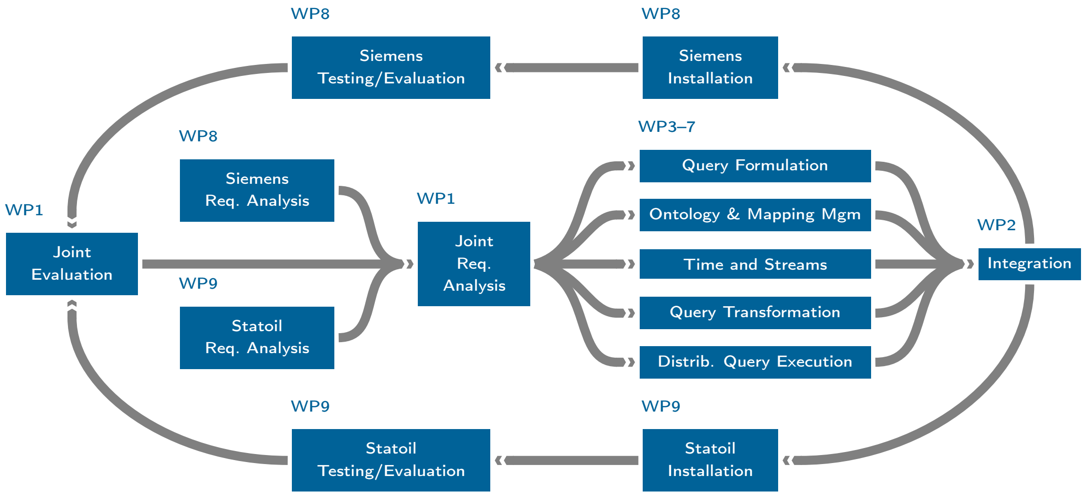


Figure 2.2: The Optique yearly development cycle.

Development of the platform in Year 2

In Year 1 the project succeeded in implementing a working prototype in which the main components were integrated to the extent that the platform could successfully evaluate simple queries from the use cases.

In Year 2 the integration has been much deeper and more robust. Also, all the novel components that are under development in the technical work packages have prototypical versions that have been integrated.

Impact-generating activities

The Optique dissemination strategy comprises a range of activities reflecting both the specific needs of the different target groups and the different phases of the Optique project. The Optique project continually revises its dissemination strategy, so as to be able to concentrate its dissemination efforts as efficiently as possible. The strategy as of the end of Year 2 is stated in Chapter 1 of deliverable D10.5.

Implementation of Recommendations from the Reviewers

After the first project review, the Commission recommendations to be implemented were as follows. A short comment on how the project has addressed these points is inserted in *italics*, referring to the Year 2 objectives in Table 2.1.

1. Stream processing and querying against federated data sources shall be high priority items for the second reporting period's project development.

   The Y2 focus on stream processing is reflected in Objective Y2.2. In order to accommodate stream processing, WP5 and WP2 have reworked the system architecture design to allow handling of *streaming data. A working group with members from WP2, WP3, WP5, and WP8 has been set up to address the displaying results of streaming queries. Additionally, the Siemens use case has been adapted to focus on predictive analysis (in the DoW scheduled for Y3) rather than reactive diagnosis, so as to supply test environments for streaming data.*

The Y2 focus on federation has been reflected in Objective Y2.3. In order to support querying against federated sources, the project has adapted the work plan accordingly, moving research on federation forwards from Y3 to Y2. A working group with members from WP1, WP2, WP7, WP8, and WP9 addresses federation via the ADP back end.

2. Careful cross-optimization between the various components of the platform is necessary to ensure optimal performance.

   Cross-component optimization is part of Objective Y2.1. In order to implement this in the work plan, a working group with members from WP4, WP6, WP7, WP8, and WP9 has been set up with particular responsibility for the optimisation of tricky queries across the various components.

3. Compatibility of STARQL with all the Optique components (e.g. in task T5.2) is recommended.

   Actions for the integration of key features with STARQL in Y2 include:

   - *Translation of STARQL queries in SQL with the help of ontop*
   - *Execution of the translated SQL queries with ADP*

4. Addressing requirements that are not explicitly mentioned the DoW but became evident during the requirements analysis phase (i.e. geospatial queries and provenance of query results) and data quality issues) should be duly considered.

   The project has placed due emphasis on these points. In particular,

   - *Geospatial queries are addressed collaboratively by WP5 and WP9, exploiting results from the TELEIOS project*
   - *Provenance of query results is addressed in WP4*
   - *Data quality issues are addressed in WP8 and WP9*

5. Regarding the overall application development process in the OPTIQUE project phases, some possible duplication of effort or lacking clear responsibility could follow from the slightly imprecise distinction between the content of tasks T1.1, T1.3, and related tasks in the industrial WPs WP8 and WP9. Clarifying the overall workflow, including requirements gathering across WPs, implementation monitoring, and user workshops, and other evaluation steps is recommended, if still possible in D1.2 (due in M15) or, at the latest, in D1.3 due in M24.

   Due to their confidential nature, end user requirements, including the Query Libraries, and evaluation results shall be reported in WP8 and WP9 (D8.2 / D9.2). WP1 shall publicly report scientific requirements and implementation requirements (but not their rationale) in D1.3. The overall workflow will be clarified in D1.3.

6. The set of mappings that are managed in order to successfully align the ontology with the underlying sources can become large and difficult to maintain. Care should be taken to tackle potential issues arising from that, including the possibility that the large number of mappings could cause a bottleneck in practical adoption and adaptability of the OPTIQUE technology given big data analytics' dynamic requirements.

   In Y2 a specialised working group has been set up between WP4 and WP2 to address the editing and debugging of mappings.

Project Objectives for Year 2

Main Objectives for Y2

During the Optique plenary meeting on June 12, 2014, the coordinator team and the work package leaders decided on objectives for the second year review. These objectives are given in Table 2.1.

Main Objectives	
Y2.1	Significantly improve the functionalities of the Y1 demonstrator wrt. • cross-component optimization • handling of mappings • support for geospatial queries and data
Y2.2	Back-end processing of STARQL queries over streaming data • rely on ADP that supports time windows and streaming answers to execute continuous queries • rely on Ontop to rewrite STARQL queries and on in-house strategy to unfold rewritten queries into SQLite • demonstration on streaming data from the Siemens use case
Y2.3	Implement federation of multiple data sources • using ADP as back-end • demonstrated on federated queries over the Statoil central data store EPDS and the NPD data sets from the Norwegian Petroleum Directorate
Y2.4	Demonstrate an integrated platform with the core end-to-end functionality, i.e., from bootstrapping to query execution
Y2.5	Demonstrate the platform and perform a usability study at Statoil and Siemens with the real endusers, queries, and data.
Y2.6	Document and widely disseminate the Optique platform targeting various research communities, enterprises, and the general public, with special emphasis on the consortium partners from industry
Y2.7	Establish the Optique Partner Programme

Table 2.1: Main Objectives in Y2

The objectives were compiled on the basis of:

- The Y2 deliverables and milestones in the DoW, summarised in Section 4
- The recommendations from the Y1 review, summarized below
- The objectives of the individual work packages
- A general consideration of the progress of the project.

The objectives reflect the following changes in the work plan of the DoW:

- Specific focus on federation and streaming is moved forward, from Y3 to Y2
- Predictive analysis in the Siemens use case is moved forward, from Y3 to Y2, in order to provide real-world basis for the work on streaming
- The focus on geospatial data was not anticipated in the DoW; it is now addressed in WP3, WP7, and WP9.

The objectives in the Publishable Summary have been adjusted so as to properly reflect the objectives for Y2.

Sub-Objectives in Y2

The main objectives Y2.1–Y2.6 are broken down into specific objectives for each individual work package. These sub-objectives are given below, structured according to the work package grouping in Figure 2.1.

Research and Demonstration Main objectives Y2.1–Y2.3 are related to research activities leading to the development of new and innovative components in the Optique platform. The objectives for the corresponding work packages WP3–7 are given in Table 2.2.

Integration Main objective Y2.4 is related to integration of all components of the Optique platform, which is the responsibility of WP2. The objectives for WP2 are given in Table 2.3.

Use Cases Main objective Y2.5 is related to the use cases. The objectives for the corresponding work packages WP1,8,9 are given in Table 2.4.

Impact Main objective Y2.6 is related to impact. The objectives for the corresponding work packages WP10 and WP11 are given in Table 2.5.

Research and Demonstration Objectives	
WP3	Develop techniques enhancing the utility of Query by Navigation over OBDA: • by increasing the number of potentially useful queries constructible by a user • by enhancing the ease of this construction Demo a stable prototype of the visual query interface for Query by Navigation including (basic) support for temporal/streaming and geospatial queries • Enhance functionality of VQS including support for building a query log. • Implement functionality for geospatial queries • Implement basic functionality for streaming queries: query formulation, registration of continuous queries, specifying result modalities • Conduct user studies with students and endusers from Statoil and Siemens
WP4	Complete the ontology and mapping bootstrapping component; extend ontology approximation with LHS existentials; demonstrate a prototypical mapping engineering component; address provenance of query results • Finalise and demonstrate the bootstrapping component • Extend ontology approximation • Mapping editing and visualization • Mapping analysis • Provenance of query results
WP5	Demonstrate continuous query over streaming Siemens data • Integrate streaming into query reformulation • Show query execution via ontop and ADP as well as visualization of results • Prepare demonstrator with local data in case Siemens systems are inaccessible • Improve performance, test scalability
WP6	Evaluate realistic Statoil queries • Analyze, optimize for problematic Statoil queries • Improve coverage of the query catalogues • Implement semi-automatic configuration and optimization
WP7	Demonstrate federated query evaluation over multiple Statoil sources, and streaming query evaluation over Siemens data • Improve the integration of ADP into the Optique platform • Support streaming sources, continuous queries • Install ADP at Statoil, Siemens; simplify the installation process • Execute test queries over federated sources at Statoil • Execute test queries over streaming sources at Siemens • Optimize and adapt ADP (e.g., for Oracle)

Table 2.2: Research Objectives in Y2

Integration Objectives	
WP2	Demonstrate an integrated platform that supports the core end-to-end functionality including data streaming and federation. • Implement a revised backend architecture for streaming • Improve integration, including the federation backend • Enhance display of (streaming) query results • Demonstrate new features: mapping editor, administration interface, etc. • Improve deployment and installation procedure

Table 2.3: Integration Objectives in Y2

Use Cases Objectives	
WP1	Establish requirements for the whole project and monitor and guide their implementation • Prepare user evaluation and requirements gathering workshops • Follow up on requirements identified during Y1 • Coordinate query catalogue creation, data availability and on-site installation activities
WP8	Demonstrate an integrated platform running at Siemens and answering realistic streaming queries • Install the Optique platform on Siemens premises • Prepare a test installation of streaming data • Initial evaluation of the platform at Siemens
WP9	Demonstrate an integrated platform running at Statoil and answering realistic queries over EPDS and NPD FP • Test and evaluate bootstrapping component on EPDS • Study usability of VQS in Statoil • Evaluate and optimise performance of query execution over EPDS • Test federated queries over EPDS and NPD FP

Table 2.4: Use Cases Objectives in Y2

Impact Objectives	
WP10	Enhance visibility of Optique in academia, industry, and to the general public • Continue academic publishing and presentation activities • Continue presentation activities for the industrial consortium partners and targeted industry • Establish the Optique Partner Programme and prepare training workshops for end users • Create white papers and other user-oriented documentation • Enhance Optique's web presence, including providing public access to the Public showcase
WP11	Establish the Optique Partner Programme • Create the Optique Alpha package for pilot customers • Implement a test installation at an Optique Pilot Partner • Test the Optique messaging with customers

Table 2.5: Impact Objectives in Y2

3 Work Progress and Achievements during the Period

Progress towards the objectives of the reporting period

Progress on Y2 Main Objectives (cf. Table 2.1)	
Y2.1	All components are now integrated in the Optique platform, which provides the core end-to-end functionality including bootstrapping, ontology and mapping management, query formulation, query transformation, and query execution over both federated and streaming sources. Cross component optimisation is ongoing, and has already significantly improved query answering performance. Mapping management now includes a fully integrated editor and prototypical analytical functions. Extended support for geospatial queries includes, e.g., map-based query formulation.
Y2.2	The Optique platform architecture has been revised to facilitate the integration of streaming data at all levels, from query formulation to query execution. A STARQL prototype has been implemented and tested using stream-extended ADP for back-end evaluation, and a cross-WP working group is addressing the visualisation of streaming query results. A demo of streaming functionality was given at the recent WP8 end-user workshop.
Y2.3	The Optique platform now supports federated query evaluation via the ADP back-end, and this was demonstrated at the recent WP9 end-user workshop integrating data from EPDS (the Statoil central data store) and NPD (the Norwegian Petroleum Directorate data set). Moreover, we are currently commissioning an 8-node cluster at Statoil that will be used for large scale federation in Y3 and Y4.
Y2.4	The fully integrated end-to-end functionality of the Optique platform was successfully demonstrated at recent WP8 and WP9 end-user workshops.
Y2.5	At the above mentioned workshops, end-users from both Siemens and Statoil were able to use the Optique platform to formulate and execute realistic queries over real use-case data, and gave very positive feedback on the usability of the platform.
Y2.6	The project has in Year 2 given more than 100 presentations of project results, including keynotes at conferences, presentations at major scientic events, presentations to high-level management staff and presentations to external companies. The project has produced 9 YouTube videos with in total 566 views in Year 2. The project was invited to present itself at a press conference arranged by the Norwegian Minster of Education along with the presentation of the Government's new research strategy.
Y2.7	The Optique Partner Programme has been established, with the Norwegian partners in lead, for a partner network in Norway and Sweden; networks lead by Optique parters in other countries planned for Y3. The first meeting was arranged for specially invited only, with 50% of the invited persons attending. A contact list with 300 names has been established. Detailed plans for two annual partner meetings in Y3 and Y4 have been developed. New pilot installations of Optique, fully funded by members of the Pertner Programme, are planned for Y3.

Table 3.1: Progress on main Year 2 Objectives

WP1: Requirement Analysis and Evaluation

UiO : University of Oslo

How do we evaluate the system performance and usability with respect to the general requirements in more specific scenarios?

- What are the global requirements?
- Do we genuinely advance the SOTA?
- Do we provide generic solutions?

TUHH Technische Universität Hamburg-Harburg

HELLENIC REPUBLIC National and Kapodistrian University of Athens

- What performance measures do we use for Optique Software?
- How do we assess its usability?

- Methods for requirement elicitation?
- Which Use Cases and Scenarios?
- Which performance indicators?

HELLENIC REPUBLIC National and Kapodistrian University of Athens

- Who are the main stake holders?
- What do we expect to accomplish?

UNIVERSITY OF OXFORD

Is the overall scientific and technical framework coherent?

Work Package 1: Requirement Analysis and Evaluation

WP1 Main Objectives (from Table 2.4) and progress made:

Establish requirements for the whole project and monitor and guide their implementation

- Prepare user evaluation and requirements gathering workshops

  The end user evaluation and requirements workshops were planned together with WP8 and WP9, and successfully conducted on October 30 and September 29–30, respectively. The results are reported in D1.3.

- Follow up on requirements identified during Y1

  D1.1 guided Optique development activity during Y2. The end user requirements for Y2 were evaluated against the Y2 prototype during the end user workshops. The results are reported in D1.3.

- Coordinate query catalogue creation, data availability and on-site installation activities

  Progress was made on all these objectives. The WP9 query catalogue guided cross-WP optimization work on query rewriting and execution throughout Y2. At the end of Y2, WP8 and WP9 are running an Optique installation on-site, using real data with ontologies and mappings evaluated at the end user workshops.

Task 1.1: Framework for Requirement Analysis

► Task definition according to DoW

This task defines a general framework for effectively capturing user requirements in the Siemens and Statoil use cases. The framework identifies the proper combination of methods for requirements elicitation to be used, and describes how to recast the end user requirements, initially expressed in scenarios and user stories, into technical requirements directed at Optique's technical work packages.

Progress

- The project delivered Deliverable D1.2 (Requirement Analysis and Evaluation framework), at the end of M15. This framework will be used during the next end user requirements analysis workshops.

No Significant Deviations from Annex I

All Objectives achieved according to Schedule

No Corrective Actions Required

Task 1.2: Framework for Usability and Performance Evaluation

► Task definition according to DoW

This task defines a general framework for carrying out usability and performance evaluation of the Optique software. The framework will define a detailed process according to which requirements will be elicited and made explicit, adjusted to Optique's overall vision, reformulated and communicated to the technical work packages, and finally used to evaluate technical project results.

Progress

- The project delivered Deliverable D1.2 (Requirement Analysis and Evaluation framework), at the end of M15. This framework will be used during the next end user requirements analysis workshops.

Meetings

No Significant Deviations from Annex I

All Objectives achieved according to Schedule

No Corrective Actions Required

Task 1.3: Joint Requirement Analysis

► Task definition according to DoW

This task generates concrete requirements for WP3–7 from state-of-the-art analysis and the annual requirements analysis and usability testing activities. This is an ongoing task which will be reported at the end of each project year, where each year's results form the basis of implementation activities in the following year.

Progress

- The project delivered D1.3 (Joint Phase 2 Evaluation and Phase 3 Requirement Analysis) in M24. End user requirements were gathered from different stakeholder groups within Siemens and Statoil in the course of the on-site end user workshops. Scientific requirements were gathered from the DoW and project-internal needs.

Deviations from Annex I

- Due to organisational reasons, the Siemens annual end-user workshops will be held in the first or second week of October. We anticipate no significant delays for D1.3.

All Objectives achieved according to Schedule

No Corrective Actions

Task 1.4: Joint Evaluation

► Task definition according to DoW

In this task the team works closely with the users to compile the results of usability tests and system performance analyses into a coherent and generic whole such that the research partners have access to the complete Optique development picture. This is an ongoing task which will be reported at the end of each project year (in D1.1 and D1.3–5, respectively). Each year's results form the basis of implementation activities in the following year.

Progress

- The project delivered D1.3 (Joint Phase 2 Evaluation and Phase 3 Requirement Analysis) in M24. Usability studies were conducted both with end users from Statoil and Siemens using domain-specific ontologies, and with a larger group using a movie ontology. Performance evaluation of Ontop is ongoing within WP6 and has led to significant speedups of generated queries against Statoil data.

No Deviations from Annex I

All Objectives achieved according to Schedule

No Corrective Actions Required

Task 1.5: Scientific and technical coordination

► Task definition according to DoW

This task coordinates evaluation activity and constantly analyses the results to ensure that the development of the platform is on a trajectory that will enable it to meet relevant performance targets and overall project goals, ensuring that the technical infrastructure is developed according to a coherent design, that it provides a generic solution to the problem of information extraction from large and complex data sets, and that it meets the (possibly changing) requirements of our use case applications with respect to both features and performance. This will involve, in particular, coordinating activities that cut across multiple work packages, including optimisation and evaluation.

Progress

During the last six month there was a significant coordination effort on demonstration activities and use cases:

- Leading the demonstration integration, i.e., development of a consolidated scenario across several WP's, preparation and filming of the demo video, etc.. This demonstration significantly extends the one we were leading during Months 6-12. As the outcome of this activity we submitted a demonstration paper to an international conference.
- Leading extensive experiments with the Optique platform over the Statoil database Slegge. As the outcome of this activity we submitted to an industrial track of an international conference.
- Leading use case description of monitoring and diagnostic tasks in Siemens service centres and consolidation of the experience with the use case. As the outcome of this activity we submitted to a in-use track of an international conference.

Meetings

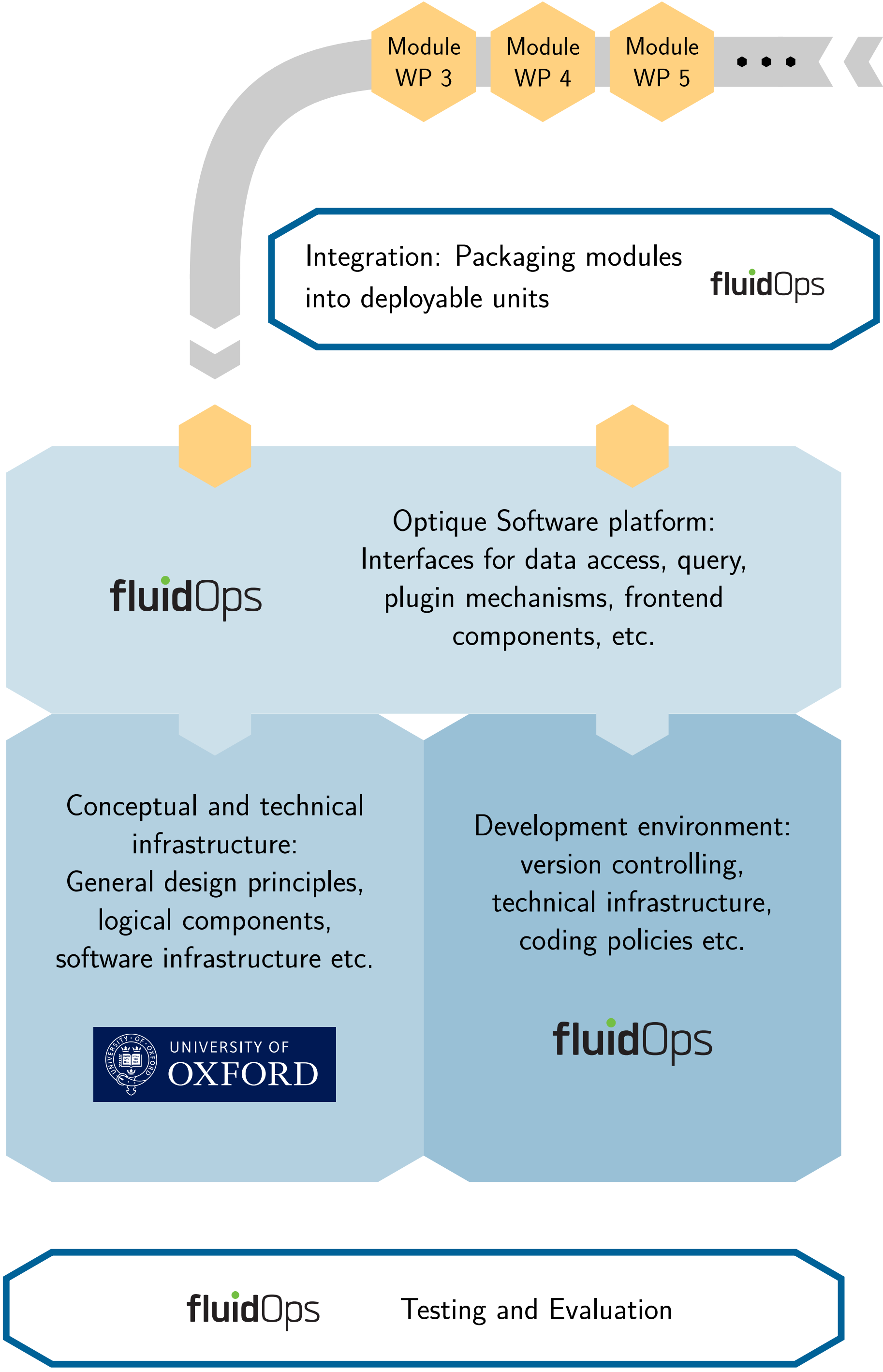
- Teleconferences: Teleconferences were used extensively for coordination within and across WP's; this maximises cohesion while minimising travel. Overall project coordination was facilitated by monthly WP leader teleconferences; these were used to discuss technical issues that cut across WP's, as well as challenges and directions more broadly.
- Visit to Oslo: during a research visit to UiO within 17.01-01.02.2014 we participated in a WP9 workshop and started extensive experiments with the Optique platform over the Statoil database Slegge.

No Significant Deviations from Annex I

All Objectives achieved according to Schedule

No Corrective Actions Required

WP2: Implementation Infrastructure



Work Package 2: Implementation Infrastructure

WP2 Main Objectives (from Table 2.3) and progress made:

Demonstrate an integrated platform that supports the core end-to-end functionality including data streaming and federation.

- Implement a revised backend architecture for streaming

  The platform architecture was extended to enable the execution of continuous queries. A new datasource for streams has been implemented to seamlessly connect to the ADP stream processing back-end and to allow visualization widgets or external applications do process continuous result streams directly.

- Improve integration, including the federation backend

  An API for mappings following the R2RML standard has been developed, and other components adapted to use it. A new datasource for the ADP federation back-end has been implemented to ease the configuration and to handle the management of federation members transparently.

- Enhance display of (streaming) query results

  A new widget has been implemented to visualize a continuous stream of results. A new KML writer has been integrated to ensure compliance with existing GIS visualization and processing software.

- Demonstrate new features: mapping editor, administration interface, etc.

  A fully functional R2RML mapping editor has been implemented and integrated. The mapping editor interacts seamlessly with the module for mapping analysis and bootstrapping. The administration front-end has been revised to allow seamless end-to-end configuration of all modules.

- Improve deployment and installation procedure

  The bundling and publishing process of platform builds have been completely re-factored to ease the deployment, installation and update procedure.

Task 2.1: Development Environment

► Task definition according to DoW

This task involves setting up and maintaining a shared development infrastructure for the Optique project, which provides the tools and methodologies for the Optique development process including the installation and maintenance of code sharing and versioning facilities as well as an environment for continuous build integration. Activities in this task include providing support for developers w.r.t. technical infrastructure questions, resolving check-in and build problems, defining coding policies and providing facilities for running automated test cases.

Progress

The task has been completely finished, in the remainder of the project the only major efforts that need to be spent in the task are maintenance works. During the second year, these maintenance activities included in particular system maintenance, security checks, re-organization of the storage infrastructure, improvements to the build infrastructure and re-structuring of the code repositories.

No Significant Deviations from Annex I

All Objectives achieved according to Schedule

No Corrective Actions Required

Task 2.2: Conceptual and Technical Architecture

► Task definition according to DoW

This task deals with integrating the various components of the Optique platform in a flexible modular architecture while at the same time achieving the performance necessary to handle very large data sets. The conceptual architecture – defined in terms of general design principles, logical components, their connections, and their interplay – will be constantly reviewed and potentially revised in response to the evaluation of individual components, of the platform as a whole, and experiences made when deploying the resulting platform in the use cases.

The conceptual architecture is complemented by the general technical system architecture, covering the application layer, its components, and its specific interfaces. As a central point, the technical architecture also includes the specification of project-wide conventions and standards. The agreement on such standards is a central, early topic of this WP, in order to assert that the modules and components developed by the project partners can be seamlessly integrated into the platform.

Progress

Evaluation of the architecture as presented in D2.1 is ongoing. In particular, small adjustments have been made w.r.t. to the integration of streams and the communication channels for result streams respectively. Evolved version of the architecture will be presented in D2.2.

Meetings Several individual synch-up meetings between FOP and TUHH regarding possible refinements of the architecture w.r.t. to interfaces and standards for querying temporal streams.

No Significant Deviations from Annex I

All Objectives achieved according to Schedule

No Corrective Actions Required

Task 2.3: Optique Software Platform

► Task definition according to DoW

This task produces the integration platform for the Optique software modules, i.e., a core platform that is extensible through open interfaces and plugin mechanisms and enables the project partners to integrate the components from the technical WP's. Based on the standards defined in Task 2.2, this software platform provides interfaces for data access, querying, and plugin-mechanisms for query language extensions as well as front-end components. An initial version of the platform, based on the Information Workbench product by fluid Operations was made available during phase 1 of the project (cf. Task 2.1).

The implementation process follows an agile software development paradigm, characterised by short, incremental development cycles (typically lasting only few weeks), followed by short testing and QA cycles.

Progress

During the second year the Optique platform has been extended in several directions. Existing interfaces have been re-factored and improved and new plugin-mechanism have been implemented:

- New data source concept & re-factored metadata management API to ease the connection to disparate data sources. Enhanced metadata handling (fully transparent, caching, multi-schema support) and support for different database.
- New facilities for platform configuration & monitoring
- New platform life-cycle listener to ease the proper registration and configuration of components and services during different all states of the platforms' life-cycle

User interface elements have been improved to ease end-to-end configuration workflows. We have created a new structure and design of the administration dashboard as a single point of entry for administrative tasks. Furthermore, we intensively worked on additional features in order to support new requirements (D1.1). These developments include:

- New mapping catalogue and fully integrated, build-in mapping editor. The mapping editor is fully standard compliant and eases the process of maintaining large collections of mappings.
- Improved facilities to ease the export, import, delete and visualization of ontologies
- New query catalogue, fully standard compliant (SPIN), support for parametrisation
- New and improved visualization and export capabilities such as for visualization of stream results, visualization of large data point collections as well as extension points to integrate with geographic information systems.
- Built-In SPARQL federation, to enable transparent federation semantic repositories

No Significant Deviations from Annex I

All Objectives achieved according to Schedule

No Corrective Actions Required

Task 2.4: Integration

► Task definition according to DoW

The goal of this task is the integration of the software components that will be developed in the technical work packages during the project. This integration will be carried out on top of the Optique platform developed in Task 2.3, which will be extensible by design, in order to make it easy to plug in the individual modules for query formulation, ontology and mapping management, stream processing, and query rewriting. This task also encompasses the work to further package these components into deployable units. One of the main goals of this task is to enable seamless interaction between the individual modules.

The integration of the components will be carried out in close cooperation between FOP and the respective technical WP leaders. Combined with the software platform developed in Task 2.3 and the components from the other technical WP's, this task will result in the four prototypes at the end of the four phases (cf. D2.3–D2.6).

Progress

We refined existing interfaces and added new extension points to facilitate closer integration of modules:

- Extended integration of component for ontology and mapping management
  - back-end integration of the component for mapping analysis
  - front-end integration of the mapping analysis into the mapping catalogue as well as into the mapping editor

  - close integration of the module for ontology and mapping bootstrapping with the mapping editor and ontology catalogue

- Devlopment of an API for the R2RML standard for relational-to-RDF mappings. This API is now used for the representation of mappings in other components like ontop, bootstrapping, etc. The intention is to establish the R2RML API as a de-facto standard, also outside of the Optique project. Care was therefore taken to make it independent of any particular underlying RDF framework.

- Extended integration of ADP component
  - ADP implementation of the new data source concept for standard, distributed JDBC mode
  - "ADP Federated" implementation of the new data source concept for JDBC federation mode
  - initial templates for automated provisioning using the platforms native cloud management functionalities

- Closer integration of the module for visual query formulation
  - implementation of the platform widget concept to ease configuration and improve back-end-communication
  - integration of extended back-end functionality
  - integration with query catalogue (queries can be stored and be re-initialxized from the catalogue)
  - initial version of the semi-automatic ontology annotation extractor has been integrated into the ontology catalogue

- Enhanced integration of the query transformation component
  - fully configurable, ability to add user defined database constraints
  - improved exception handling, simplified exception messages are propagated to the end-user

- Initial back-end integration of component for handling time and streams
  - implementation as an extension to the standard query transformation component
  - to operate on a newly implemented ADP "StreamingDataSource", instead of standard JDBC datasource
  - transparent registration and transformation of STAQRL queries to a StreamingDataSource
  - new visualization widget to operate on the streaming data source

- Integration of KML result writer for GIS

Finally, the packing and publishing process of the platform has been completely re-factored to ease the installation and deployment of the platform.

Meetings Among monthly WP2 phone conferences (refinements of requirements, interfaces, responsibilities, and status updates), we had various bilateral phone conferences between partners to discuss possible refinements of the platform interfaces. In particular, we had dedicated synch-up meetings to agree on

- new extension points for the analysis and management of mappings (FOP, UNIROMA1)

- additional back-end functionality for query formulation (UiO, UOXF, FOP).

- workflows and preparation of templates for provisioning of worker nodes for distributed query execution (FOP and UoA)

- interfaces to handle temporal stream queries asynchronously(FOP and TUHH)

No Significant Deviations from Annex I

All Objectives achieved according to Schedule

No Corrective Actions Required

Task 2.5: Design of Evaluations and Analysis of Results

► Task definition according to DoW

While the individual components will be evaluated as stand-alone modules in the work packages in which they are developed, in this task the platform as a whole will be continuously tested and evaluated, in order to monitor progress and to ensure fitness for purpose. The task thus comprises both the design and implementation of a suitable evaluation framework and the analysis of the testing and evaluation results. The evaluation results will be used as as continuous feedback for the design and development of the technical infrastructure.

Following common software development principles, the testing and evaluation infrastructure will include both unit tests for the individual modules (e.g., functional test) as well as integration test cases that cover the interaction of internal and external modules and system components. These integration tests will be guided by and evaluated against the real-world use cases provided by the use case partners. The correctness tests are complemented by a continuous performance evaluation throughout all stages of the project, to assess the general quality of component interplay, global optimisations, and the adequacy of the system and implementation strategies on top of practical use cases. The task will result in a comprehensive evaluation report at the end of the project (D2.7).

Progress

We implemented new facilities for integrated unit testing of components. Furthermore, an improved continuous integration infrastructure makes new changes immediately accessible to non-developers as well as guarantees seamless interaction between the modules. Finally, several ADP worker nodes have been deployed to work in a clustered mode with in the testing environment.

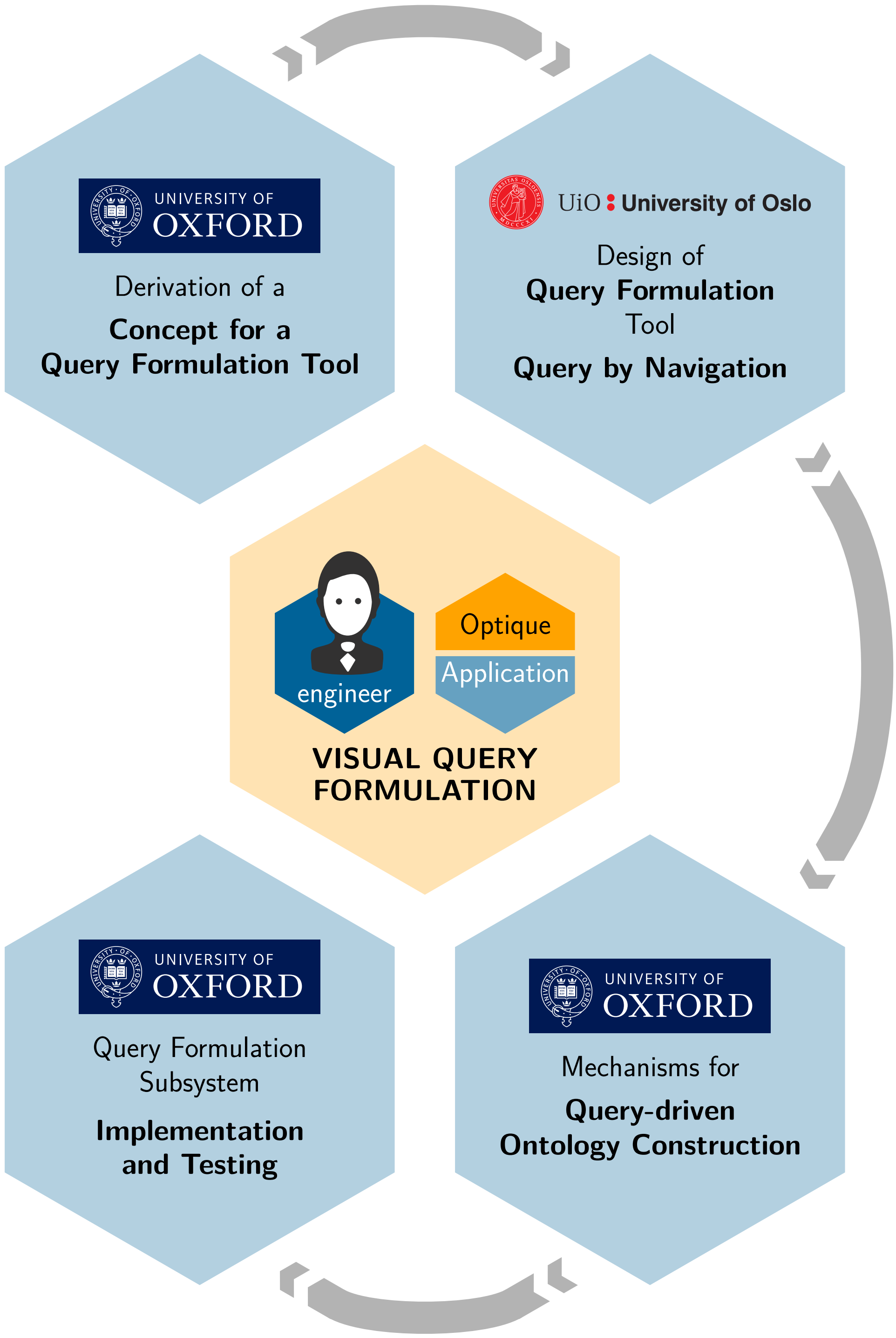
Meetings As part of the monthly WP2 phone conference we discussed requirements and defined further steps towards an integrated benchmarking framework.

No Significant Deviations from Annex I

All Objectives achieved according to Schedule

No Corrective Actions Required

WP3: Query Formulation



Work Package 3: Query Formulation

WP3 Research and Demonstration Objectives (from Table 2.2) and progress made:

Develop techniques enhancing the utility of Query by Navigation over OBDA:

- by increasing the number of potentially useful queries constructible by a user

  Generic techniques of graph based ontology projections for navigation over complex ontologies were developed. Such projections encode classes semantically related via properties which can be suggested to users during the query formulation process.

- by enhancing the ease of this construction

  Ranking techniques (based on a catalog of useful queries) for concepts and relations that form the suggestions were developed.

Demo a stable prototype of the visual query interface for query by navigation including (basic) support for temporal/streaming and geospatial queries

- Enhance functionality of VQS including support for building a query log.

  A general architecture for the VQS backend has been designed in which the behavior of the VQS is controlled by annotations of the ontology. Several new functionalities has been introduced in the VQS and existing ones were refined, including loading and saving queries to a catalog, storing executed queries in a query log, experimental mode of the VQS.

- Implement functionality for geospatial queries

  Ontology annotations to identify concepts with geospatial extent were added, and a map widget was added to the Query Formulation interface, to allow selecting entities on a map. These improvements were well received by end users in Statoil.

- Implement basic functionality for streaming queries: query formulation, registration of continuous queries, specifying result modalities

  Specialised widgets for time ranges and time windows for streaming queries were implemented. The current implementation supports most of the temporal and streaming features for the Siemens query catalog. Work on translating interface queries into STARQL is under way.

- Conduct user studies with students and endusers from Statoil and Siemens

  We conducted evaluation of OptiqueVQS with 15 students from the University of Oslo and then with endusers from Statoil and Siemens during two workshops.

Task 3.1: Derivation of a concept for a query formulation tool

► Task definition according to DoW

This task exploits the initial joint requirement analyses in order to derive a concept for a tool for query formulation support to be used by end-users and IT-experts to access information. The work is based on an analysis of the methodologies currently employed for query formulation and the current state of the art with respect to tool support. The results of this analysis will be reported in D3.1.

Progress

We worked on the requirement analyses for the use cases. In particular, we received a query catalogue made by Statoil engineers which we analysed. As a result, we updated the requirements for the interface.

- *Analyses of query catalogs:*
  We analysed the catalogue in order to understand which queries from it could be formulated in the current version of the interface. Based on this we figured out the missing constructs that are required by Statoil specialists for formulating necessary queries.

- *New requirements:*
  We found that there is a need to support the selection of a geographical region, to which the composed query should be applied.

- *Publications:*
  A detailed overview of our analyses of related work has been submitted to an international journal and is currently under review.

Meetings

- Regular teleconferences were held during the reported period, including general WP3 meetings.
- Project meeting in Oslo: January 17–February 1, 2014, between UiO (Martin Giese and Ahmet Soylu) and UOXF (Evgeny Kharlamov, Ernesto Jiménez-Ruiz, and Dmitriy Zheleznyakov).
- Statoil enduser workshop in Stavanger: Sept 28–30.
- Siemens enduser workshop in Oslo: October 30.

Deviations from Annex I

In Annex I this task was foreseen to end in Y1. New user requirements in the Y1 evaluation made it necessary to extend the work in this task beyond Y1.

Lapses in Schedule

Due to user requirements in the Y2 evaluation, this task will continue in Y3.

Task 3.2: Query Formulation

► Task definition according to DoW

In this task we investigate and develop techniques to support query formulation, and design a suitable visual query formulation system (VQS) with direct editing of support. The VQS aims to exploit the domain model captured in the ontology in order to allow users to formulate queries in a way that is intuitive to them, that matches their own conceptualisation of the domain, that uses vocabulary that they are familiar with, and that does not require any knowledge of the structure of the underlying data sources. Direct editing of queries should also be possible using a suitable component, e.g., a context sensitive Query Editor (using the ontology to provide pre-completion and menu selection of terms, and the structure of the query language to provide pre-completion and menu selection of language constructs). The editor should be tightly connected to the VQS so that the user is able to switch freely between QbN and direct editing of the query at any stage of query formulation; this will also facilitate collaboration between more and less technically sophisticated users. The integrated query formulation tool should generate coherent queries that can be passed to the query translation sub-system (WP6).

Progress

Based on the analyses carried out in T3.1, we made the following progress:

- *Development of ontology projection techniques:*
  Our Query by Navigation approach essentially relies on a paradigm where a user construct queries by exploring the system's ontology in a graph traversal fashion, i.e., by navigation through ontological classes via properties. To increase the number of potentially useful classes that a user ca reach in this way, we proposed ontology projection techniques and studied properties of such projections.

- *Definition of a ranking model:*
While a user constructs queries via an ontology traversal, he may encounter an overwhelming number of classes and properties. To address this issue we developed a ranking model for classes and properties that is based on the analyses of query catalogs. The model suggests the user classes and properties that follow the most frequent patterns captured in the catalogs and asked by users.

- *Separation of logical and visualisation ontological parts:*
We separated the ontology into two parts: for logical reasoning and for visual query formulation. The latter one is based on ontology annotations and we extended the backend of OptiqueVQS to support treatment of annotations.

- *Temporal and streaming queries:*
A preliminary query formulation support for temporal and streaming queries was added. Users can set these queries in specialised widgets imbedded in OptiqueVQS.

- *Geospacial queries:*
An initial implementation of a geospacial component of OptiqueVQS was conducted. Users can select areas of interest via an interactive map which is supported by a specialised widget imbedded in OptiqueVQS.

- *Publications:*
We published our work in several international conferences and workshops, see D3.2 for details.

Meetings

- Regular teleconferences were held during the reported period, including general WP3 meetings.
- Project meeting in Oslo: January 17–February 1, 2014, between UiO (Martin Giese and Ahmet Soylu) and UOXF (Evgeny Kharlamov, Ernesto Jiménez-Ruiz, and Dmitriy Zheleznyakov).
- We also participated in several meetings on other relevant work packages in order to get a better understanding of query formulation needs.
- Statoil enduser workshop in Stavanger: Sept 28–30.
- Siemens enduser workshop in Oslo: October 30.

No Significant Deviations from Annex I

All Objectives achieved according to Schedule

No Corrective Actions Required

Task 3.3: Query-driven ontology construction

► Task definition according to DoW

This task investigates and develops mechanisms for query-driven ontology construction (QOC), i.e., extension of the domain ontology "on the fly", needed for query construction. Extensions to the ontology might be required for anything from single new vocabulary terms to the entire structure of the query. We investigate how techniques for ontology learning and alignment can be adapted to deal with such cases, in particular how effectively new terms and conceptual structures can be derived from the query and aligned with the existing domain ontology, and how to extend the query formulation tool designed in Task 3.2 so as to support QOC. This task also works closely with WP4 to ensure a seamless integration of QOC with ontology and mapping management tools.

Progress

We are currently working on query driven ontology construction module. In particular, we started on the subproblem of designing on techniques for learning ontologies from query catalogs, the result of which will be tested on Siemens and Statoil query catalogs.

Meetings

- Several teleconferences were held during the reported period.

Deviations from Annex I

Implementation of the the query driven ontology construction module is planned for Year 3.

Task 3.4: Query Formulation Sub-System Implementation and Evaluation

► Task definition according to DoW

The result of this task is a query formulation sub-system for integration in the Optique platform. The sub-system will provide all the functions described in the previous tasks, including a visual query system, a query editor and query driven ontology construction, and will be continuously tested and evaluated as part of the implementation process.

Progress

- *Architecture revision:*
  A refined architecture was defined for the backend of the query formulation component, that allows to support our new developments such as ontology separation, ranking, etc., as well as to control the behaviour of the frontend based on the ontology, data, previous queries, and manual instructions.
- *New functionality of OptiqueVQS:*
  - `Query management`: users can save/load queries from the platform.
  - `Reversible actions`: users can undo/redo their actions over their working query.
  - `Subclass refinement`: users can refine a selected node to one or more of its subclasses, which are treated as multi-select field in the form-based widget.
  - `Widget binding to form elements`: system allows binding input widgets to form fields, for instance a map widget is attached to representative attributes of concepts that have geographical presence, so that a user can select an instance from the map widget rather than typing its name.
  - `Parameterised initialisation`: system now can be initiated with a set of predefined parameters such as, ontology to use, a stored query id to load, repository etc.
  - `Experiment mode`: system now can be operated on an experiment mode, in which the use of tool is controlled and session data is collected for experimental analysis. For instance, user attempts are stored along with time to complete each attempt and a simple interface is made available for pre-loading tasks for an experiment.
- *User studies:*
  - We conducted a user study at the University of Oslo with 15 students with a non-technical background. The results are quite positive and has been submitted to an international journal.
  - For end user evaluation in Statoil and Siemens we prepared questionaries and query tasks and conducted the evaluation during two enduser's workshops.

- *Experiments with ontologies and queries of use-cases:*
  We tested the interface on the Siemens and Statoil ontologies and, as the result, we extended the ontologies with annotations that support facilitation of query formulation. In particular, they support such features as image representation and sliders for representing numerical values, etc.

- *Publications:*
  We published and submitted to publication demonstration papers with different stages of the current version of OptiqueVQS and proposed approaches to enhance visualisation, see D3.2 for details.

Meetings

- Regular teleconferences were held during the reported period, including general WP3 meetings.

- Project meeting in Oslo: January 17–February 1, 2014, between UiO (Martin Giese and Ahmet Soylu) and UOXF (Evgeny Kharlamov, Ernesto Jiménez-Ruiz, and Dmitriy Zheleznyakov).

- Statoil enduser workshop in Stavanger: Sept 28–30.

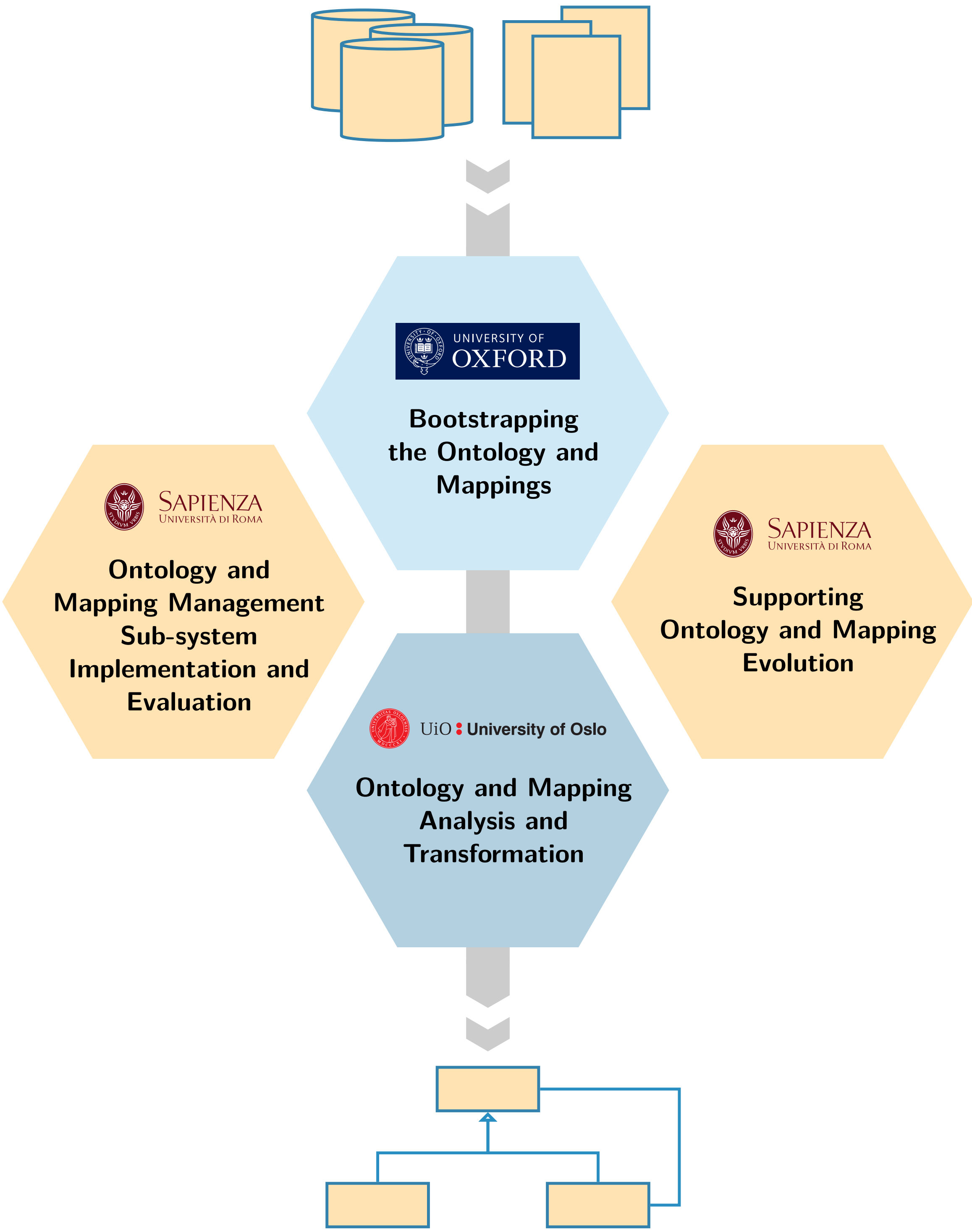
- Siemens enduser workshop in Oslo: October 30.

No Significant Deviations from Annex I

All Objectives achieved according to Schedule

No Corrective Actions Required

WP4: Ontology & Mapping Management



Work Package 4: Ontology and Mapping Management

WP4 Main Objectives (from Table 2.2) and progress made:

Complete the ontology and mapping bootstrapping component; extend ontology approximation with LHS existentials; demonstrate a prototypical mapping engineering component; address provenance of query results.

- Finalise and demonstrate the bootstrapping component

  The bootstrapping components of the Year 1 prototype have been improved and extended, and new bootstrapping techniques have been developed, implemented, integrated in the Optique platform, and evluated through experiments.

- Extend ontology approximation

  A new technique for ontology approximation was designed, implemented, and experimetally evaluated and compared to the one developed in Year 1.

- Mapping editing and visualization

  A new mapping catalogue and a fully integrated, build-in mapping editor have been developed, as well as a prototypical implementation of mapping visualisation. The prototypes have been demonstrated to IT experts at Use Case partners.

- Mapping analysis

  New algorithms for mapping analysis have been defined and implemented in a prototypical module for mapping analysis.

- Provenance of query results

  A first technique for handling provenance in OBDA context has been defined; the ontology and mapping bootstrapper has been extended to automatically support such an approach to provenance.

Task 4.1: Bootstrapping the ontology and mappings

► Task definition according to DoW

This WP4 task is devoted to the design and development of techniques for ontology and mapping *bootstrapping*, i.e., the automatic or semi-automatic generation of ontologies and mappings starting from existing meta-data (e.g., data source schemas, queries, etc.). The results of this task will be presented in Deliverable D4.2, "Techniques for ontology and mapping bootstrapping: final version" (month 24).

Progress

During the last 12 months we significantly improved the Ontology&Mapping bootstrapper, conducted extensive evaluation of the bootstrapper with the Statoil, Siemens, and other databases, and met the objectives of T4.1. More precisely the progress of the last 12 months is the following:

- O&M components of Year 1 implementation were improved and extended:
  - ontology alignment is improved with safety verification,
  - W3C test cases for the O&M bootstrapper were implemented,
  - the O&M bootstrapper is tightly integrated with the mapping editor,
  - installation GUI and wizards were improved,
- extensive experiments and evaluation of O&M bootstrapper were conducted on relational schemas from use-cases and other schemas,

- new techniques bootstrapping of direct mappings, called *layering*, were developed, implemented, and integrated in the Optique platform,
- preliminary techniques for bootstrapping of complex mappings, i.e., involving select, project, and join relational operators were developed,
- preliminary techniques for embedding provenance in bootstrapped mappings were developed,
- work on benchmarking O&M bootstrapping approaches has been started.
- We publish some of our techniques (see D4.2 for details) and submitted a demo paper to a prestigious international conference.

Meetings A number of meetings and teleconferences have been held in the reporting period:

- Hands-on meeting in Oxford between UOXF and UiO from March 10th to March 18th. O&M Bootstrapping applied to the Statoil use case (WP9).
- A number (25+) of teleconferences where different partners were involved, including UOXF, UiO, FOP, FUB and UNIROMA1 on the integration within the platform, the integration with the R2RML API, the discovery of implicit relational constraints, and the application to the Statoil use case.
- Visit of a UOXF representative (Ernesto Jimenez-Ruiz) to UNIROMA1 took place from June 25 till July 12.

No Significant Deviations from Annex I

All Objectives achieved according to Schedule

No Corrective Actions Required

Task 4.2: Ontology and mapping analysis and transformation

► Task definition according to DoW

This WP4 task is devoted to the design and development of techniques for ontology analysis and mapping analysis, and for transforming a set of mappings according to a predefined formal criteria. Analysis tasks over the ontology include consistency and redundancy checking. To ensure that the schema mappings enjoy appropriate properties, such as efficiency and scalability, suitable transformations (e.g., simplification) should be applied to them. Besides traditional equivalence preserving transformations, we will study more innovative transformation primitives.

The final results of this task will be presented in Deliverable D4.3, “Techniques for ontology and mapping analysis” (month 36) and in Deliverable D4.4, “Ontology and mapping analysis and evolution” (month 48).

Progress

- In the last year, we have studied ontology approximation. In particular, we have provided a general, parametric formal framework for ontology approximation. Such a framework generalizes the approach to ontology approximation that was defined during Year 1. Based on this framework, we have defined two specific notions of sound approximations of ontologies: global semantic approximation (GSA) and local semantic approximation (LSA). We have then defined algorithms for computing the GSA and LSA of OWL 2 ontologies in OWL 2 QL. We have implemented such techniques in a first prototype and have conducted an extensive experimental comparison of the two techniques over a large existing ontology repository.

- Then, we have implemented the above ontology approximation techniques within the Optique platform. More specifically, while an implementation of the LSA was provided in Year 1, in Year 2 we have implemented the GSA for OWL 2 QL. We have tested the new implementation and compared it with the previous one through an extensive experimental analysis. We have also tested the two ontology approximation techinques both in the Siemens use case and in the Statoil use case. The general outcome of these experiments is that these kinds of approximations are in practice very useful, since they provide very good approximations (in the vast majority of the cases, only very few axioms are missed by the approximations), and our algorithms are able to effectively comptute them. The GSA is more precise than the LSA, while it requires more computational resources to be constructed.

- The following paper reports the results of the above activities on ontology approximation:

  Marco Console, Jose Mora, Riccardo Rosati, Valerio Santarelli, Domenico Fabio Savo: Effective computation of maximal sound approximations of Description Logic ontologies. In *Proceedings of ISWC 2014*, 164-179, 2014.

- During the reporting period, we have also worked on mapping analysis. In particular, we have analysed and formally defined the main basic problems relative to mapping analysis, classifying such problems into local mapping analysys problems and global mapping analysis problems. We have also defined techniques for the verification tasks related to the above problems, in particular the global mapping analysis problems.

- The following paper reports the results of the above activities on mapping analysis:

  Domenico Lembo, Jose Mora, Riccardo Rosati, Domenico Fabio Savo, Evgenij Thorstensen: Towards mapping analysis in Ontology-based Data Access. In *Proceedings of RR 2014*, 108-123, 2014.

- Then, we have implemented the first version of a mapping analysis component within the Optique platform. This component implements the main functionalities formally defined at the previous point. The component is able to check the mapping in an OBDA specification, performing both syntactic checks and semantic checks over the mappings:
  - correctness with respect to the source schema signature;
  - correctness with respect to the ontology signature;
  - correctness with respect to the usage of mapping variables;
  - identification of inconsistent mapping assertions;
  - identification of redundant mapping assertions.

  We have also experimented such a component on the mappings developed within the Statoil use case.

- Finally, we have defined a technique for the visualization of mappings in the OBDA context. The visualization technique is based on the idea of presenting mappings according to a 5-layered structure, from the ontology (top layer) to the data sources (bottom layer). Three different representation modalities are defined: ontology-centered representation, mapping-centered representation, and source-centered representation. A prototype implementing the proposed visualization of mappings has been developed and presented at the DL 2014 international workshop, and has been demonstrated to IT experts at the Statoil User Evaluation Workshop. In addition, a new mapping catalogue and a fully integrated, build-in mapping editor have been developed, as reported in WP2.

- The following paper reports the results of the above activities on mapping visualization:

  Domenico Lembo, Riccardo Rosati, Marco Ruzzi, Domenico Fabio Savo, Emanuele Tocci: Visualization and Management of Mappings in Ontology-based Data Access (Progress Report). In *Proccedings of the International Workshop on Description Logics 2014*, 595-607, 2014.

Meetings

- A number (15+) of teleconferences (including the WP4 monthly teleconferences) where different partners were involved, including UOXF, UiO, FOP, FUB and UNIROMA1 on the integration within the platform, the mapping analysis functionalities, and the application to the Statoil use case.
- On March 26 we had a teleconference with FluidOps to coordinate the work on mapping analysis with the other mapping-related implementation and integration work within Optique.
- On April 17 we had a teleconference to coordinate the final set of tasks to (fully) meet the objectives of the Year-2 demo.
- Visit of a UiO representative (Evgenij Thorstensen) to UNIROMA1 took place from October 13 till October 26.

No Significant Deviations from Annex I

All Objectives achieved according to Schedule

No Corrective Actions Required

Task 4.3: Supporting ontology & mapping evolution

► Task definition according to DoW

This task provides the techniques for supporting the evolution and the maintenance of both the ontology and the mappings, thus addressing the dynamics of the OBDA specification. In order to support the required evolution of ontologies, a combination of different techniques will be used, including ontology alignment and ontology approximation.

The final results of this task will be presented in Deliverable D4.4, "Ontology and mapping analysis and evolution" (month 48).

Progress

- We have studied the basic problems that are relevant for ontology and mapping evolution in OBDA. In particular, we have produced a first set of problems and a partial classification of such problems with respect to some fundamental characteristics.
- The results of the above activity have been presented in an internal report, which we aim to submit to an international conference in 2015.
- In addition, we have started examining the interplay between ontology and mapping evolution and the activities conducted within the above Task T4.2: in particular, we would like to investigate the possible usage of ontology approximation in the context of the present task, i.e., evolution in OBDA.

Meetings

- During the WP4 monthly teleconferences where different partners were involved, including UOXF, UiO, FOP, FUB and UNIROMA1, we have discussed the issues related to ontology and mapping evolution functionalities, and, in the long term, their future integration with the mapping editor.

No Significant Deviations from Annex I

All Objectives achieved according to Schedule

No Corrective Actions Required

Task 4.4: Ontology and mapping management sub-system implementation and evaluation

► Task definition according to DoW

This task delivers the sub-system for the management of ontology and mappings for the Optique platform. The sub-system provides all functions described in the previous tasks, including bootstrapping the ontology and mappings, browsing and querying the OBDA specification, analysis and debugging of mappings, mapping transformations, and maintenance of mappings under the evolution of the OBDA system. This task also deals with evaluation and testing of successive prototypes of the ontology and mapping management sub-system.

Progress

- as already explained in Task T4.1, the ontology and mapping bootstrapping subsystem has been extended and integrated in the Optique platform;
- as already explained in Task T4.1, the ontology and mapping bootstrapping subsystem has been experimented within the project use cases;
- as already explained in Task T4.2, a new ontology approximation technique has been developed within the Optique platform and has been experimented within the project use cases;.
- as already explained in Task T4.3, a mapping analysis component has been developed within the Optique platform and has been experimented within the project use cases.

Meetings

- see the above descriptions of meetings in Task T4.1 and Task T4.2.

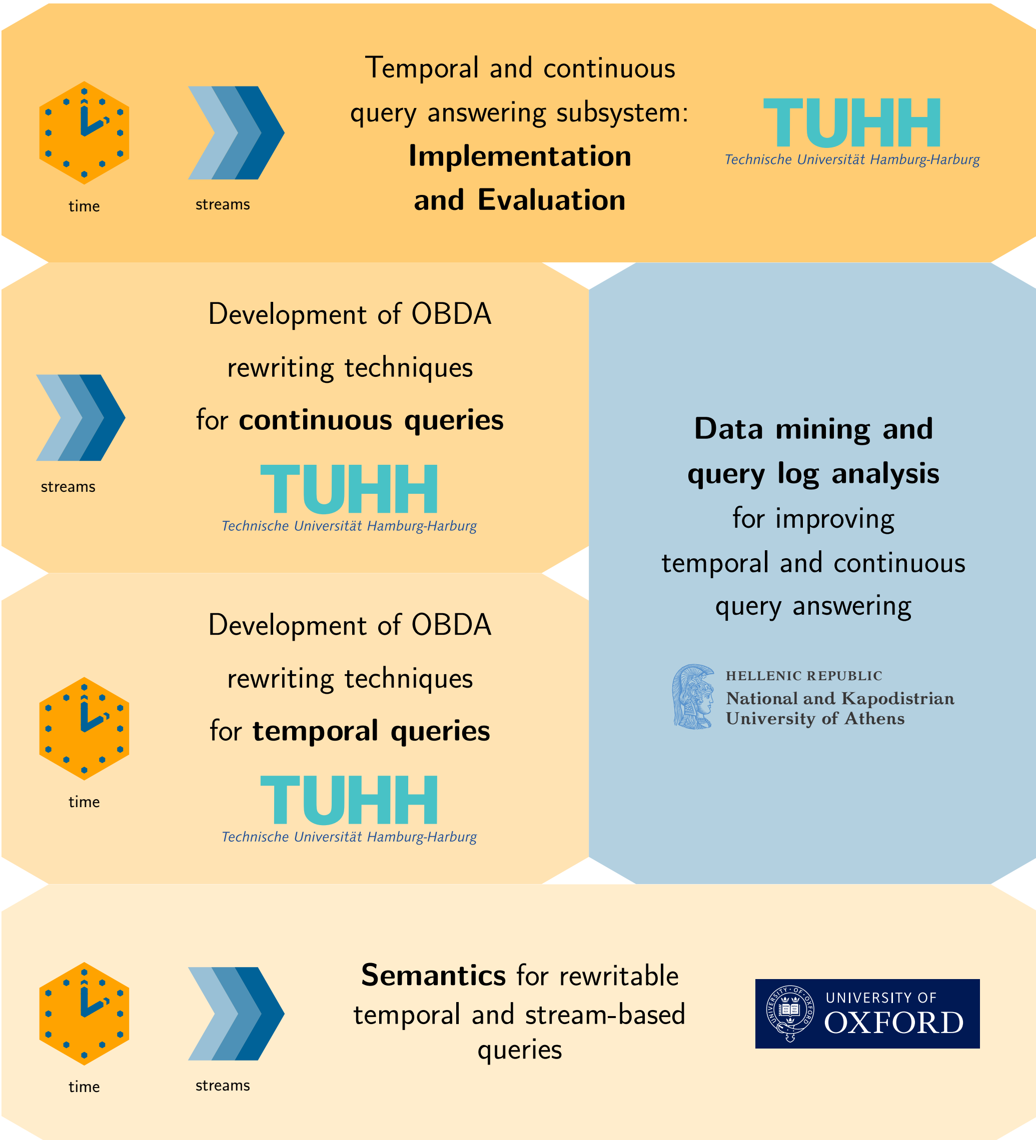
Deviations from Annex I

The task was planned to start in M20, but started in M13 in response to the reviewer recommendations for the Y1 review implementation of the mapping management subsystem was accelerated.

All Objectives achieved according to Schedule

No Corrective Actions Required

WP5: Time and Streams



Work Package 5: Time and Streams

WP5 Main Objectives (from Table 2.2) and progress made:

Demonstrate continuous query over streaming Siemens data

- Integrate streaming into query reformulation

  A rewriting and unfolding methodology for STARQL queries was developed such that any STARQL query can be transformed into SQL like queries. In particular, for stream reasoning with STARQL, a sound and complete algorithm for transforming STARQL queries into the relational stream query language CQL and the extended SQL-Lite query language provided by the ADP system was developed. Moreover, for temporal reasoning with STARQL, a transformation into the SQL variant provided by PostgreSQL was given.

- Show query execution via ontop and ADP as well as visualization of results

  A STARQL prototype was implemented and integrated into the Optique platform. The prototype uses onto mappings and realizes the transformation into queries of the ADP system.

- Prepare demonstrator with local data in case Siemens systems are inaccessible

  A demonstrator with Siemens data was installed in the Optique platform running on the servers of fluidOps; moreover the platform now runs on-site at Siemens and was demonstrated at the Siemens workshop at the end of October.

- Improve performance, test scalability

  The STARQL engine was tested on data provided by Siemens, a small public dataset and a larger anonymized dataset. Approx 70% of the queries from the query catalog of the reactive diagnosis scenario were formulated and run successfully in STARQL. First optimizations in case of temporal reasoning with STARQL were tested: The transformation from STARQL to PostgreSQL (ADP) produced many occurrences of subqueries some of which could be shown to be redundant.

Task 5.1: Semantics for rewritable temporal and stream-based queries

► Task definition according to DoW

This task defines a semantics for temporal and stream-based queries in such a way that many application requirements are considered and query rewriting can ensure scalable and efficient query execution. The syntax of the Optique query language(s) is based on extensions to SPARQL and full conjunctive queries, which will also be developed in this task. Although some suggestions to extend SPARQL with stream capabilities have been described in the literature, and although these have been also been implemented in some stream engines, they do not fit the needs for a full "streamified" OBDA as they do not involve real ontologies (such as OWL 2 QL ontologies) but at most RDF(S) knowledge bases. The definition of a stream-temporal query language that, on the one hand, is expressive enough for a large range of applications and, on the other hand, can be implemented efficiently for large scale data, provides the basis for the implementation of the stream-temporal components for the second prototype.

Progress

- The task has been completed in M12 successfully with the STARQL query language on the basis of a syntax specified by an extended context free grammar (see Deliverable D5.1)

Task 5.2: Development of OBDA rewriting techniques for temporal queries

► Task definition according to DoW

This task extends query rewriting in the sense of OBDA to temporal data such that scalable query answering over existing data sources is supported. Using an ontology and a set of mapping rules for mapping ontology notions mentioned in temporal queries to the nomenclature used in particular relational database schemas, both query formulation and execution can be simplified.

Based on work done in WP3, WP6, WP7, time in query languages will be embedded into query rewriting, data storage layout and indexing will be considered in this process.

Progress

- An instantiation of the STARQL query framework was developed that guarantees rewriting STARQL queries into ADP queries.
- For this we invented a safety mechanism that guarantees domain independence (and we chose DL-Lite as ontology language and unions of conjunctive queries as embedded conditions.)
- We investigated the requirements for a relational stream query language into which STARQL queries can be transformed.
- We formulated these requirements as technical requirements on the ADP system, resulting in a new extended SQLite language provided by the stream extended version of ADP.

Meetings

- Regular group telephone conferences between TUHH, UOXF, UoA, FUB, UiO, Siemens, and FOP, and further bilateral teleconferences as needed.
- Aspects of rewriting and unfolding STARQL into ADP were discussed in a workshop with colleagues from Siemens (WP8) and from UoA (WP7) on 8th of April 2014 in Athens.

No Significant Deviations from Annex I

All Objectives achieved according to Schedule

Corrective Actions

Task 5.3: Development of OBDA rewriting techniques for continuous queries

► Task definition according to DoW

This task extends query rewriting in the sense of OBDA to continuous queries such that scalable query answering over existing data sources is supported as well. Quantitative performance evaluations will make sure that many thousands of continuous queries can be registered while query answering time should keep pace with representative data rates found in the use cases (i.e., the velocity with which the “query window” moves over the stream).

Progress

Please see the progress for Task 5.2, as rewriting/unfolding for continuous STARQL queries were handled in the same way as for pure historical STARQL queries.

Task 5.4: Data mining and query log analysis for improving temporal and continuous query answering

► Task definition according to DoW

In this task, custom operators for data interpretation are developed that enable scalable query answering and avoid repetitive recognition of often used structures. Such operators, combined with a more expressive language, are needed to support more advanced temporal queries, as well as pre-processing using threshold and trend analysis. High-level event descriptions will be generated in an adaptive way (based on the query log, i.e., the queries the system regularly has to answer). High-level event descriptions are considered during query rewriting such that query answering can be optimised.

This task also provides extensible data analysis and mining techniques in order to discover and monitor common patterns both in data and queries, supporting reactive and predictive scenarios analysed in the Siemens use case, as well as, adaptive query planning and execution. Queries are automatically rewritten such that automatically generated explicit events are anticipated and query answering can scale. Feedback about induced event classes is given such that event classes can be named in the ontology and can also be referred to in future queries. We expect that these techniques will further improve the scalability of the Optique system, especially when deployed on elastic clouds using ADP.

Progress

Work on this task has not started yet.

Task 5.5: Temporal and continuous query answering sub-system: Implementation and Evaluation

► Task definition according to DoW

This task develops and evaluates the software components that integrate temporal and continuous query answering into the Optique implementation infrastructure, implementing the results of T5.1–T5.4.

Progress

- Implementation of a STARQL prototype that 1) uses a safety mechanism, 2) relies on ontop for the rewriting, and 3) uses an own unfolding strategy for unfolding STARQL queries into the extended SQLite query language of the stream extended ADP system.
- Successfully tested some of the Siemens queries identified in the WP1 requirements gathering process (see Deliverables D8.1 and D1.1).
- Identification and refinement of a stream software architecture for Optique into which the STARQL prototype can be embedded
- Implementation of a standalone STARQL prototype and testing it w.r.t. temporal reasoning on the standard DBMS PostgreSQL in comparison to ADP as backend data source.
- Extension of Ontop to support temporal datatypes in all components of the OBDA architecture (database, ontology, mapping, user queries, and generated SQL query). Specifically, the temporal datatypes supported are `xsd:date`, `xsd:time`, `xsd:gYear`, and `xsd:dateTime`.

Meetings

- Implementation issues regarding transformation of STARQL into ADP were discussed in diverse teleconferences with UoA and in the workshop with colleagues from Siemens (WP8) and from UoA (WP7) on 8th of April 2014 in Athens.
- Architecture for STARQL prototype and integration with Siemens platform were discussed with colleagues from Siemens in Hamburg on 23rd of January 2014.

- Diverse follow up telephone conferences regarding implementation and integration into Optique platform

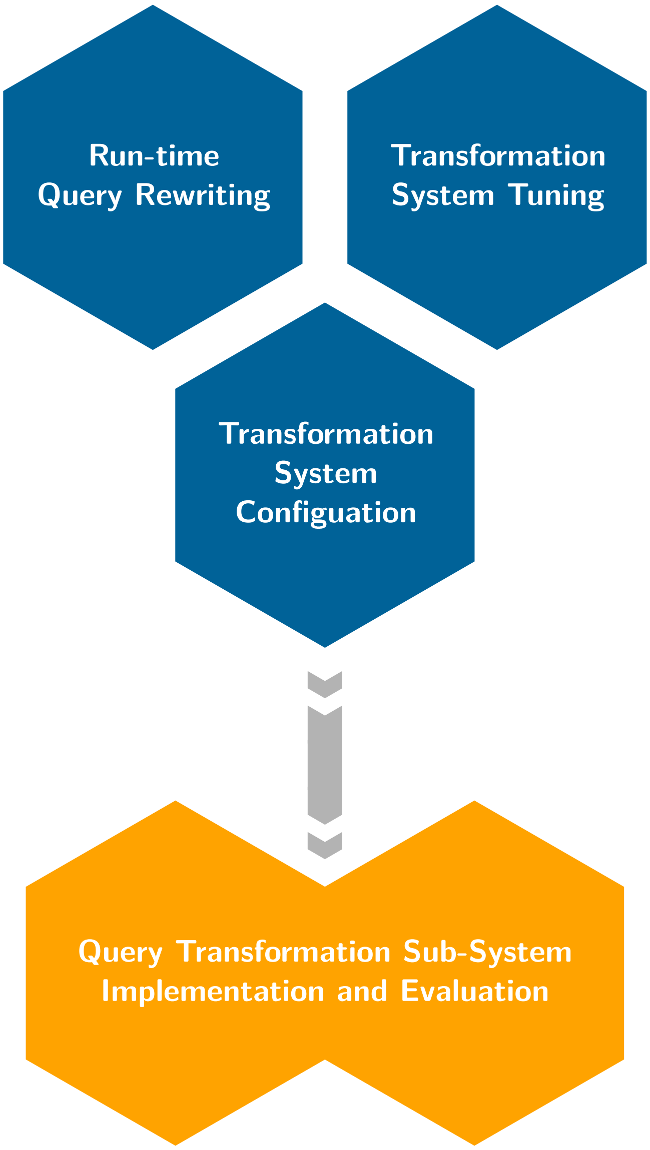
Deviations from Annex I

Task T5.3 was planned to start in M20, but started in M13 in response to the reviewer recommendations for the Y1 review.

Failure to Achieve Objectives/Lapses in Schedule

Corrective Actions

WP6: Query Transformation



Freie Universität Bozen
Libera Università di Bolzano
Free University of Bozen · Bolzano

Work Package 6: Query Transformation

WP6 Main Objectives (from Table 2.2) and progress made:

Evaluate realistic Statoil queries

- Analyze, optimize for problematic Statoil queries

  Development of novel tuning techniques was started to produce SQL queries that can be more efficiently handled by the underlying database engine. These techniques are based on the discovery that (1) the redundant self joins in generated SQL queries are often a consequence of the fact that functional constraints that hold in the database are actually not specified explicitly; (2) the ontology reasoning for a concept that is part of a hierarchy in the ontology can be safely disabled when the mapping for that concept is complete with respect to the ontology.

- Improve coverage of the query catalogues

  To improve the support for standard languages and libraries, which are used in the query catalogues, we implemented various extensions of the system functionalities: (a) support for the R2RML API of the standard R2RML mapping language, (b) support for extended SQL in the mapping language, (c) support for queries over multiple schemas, and (d) improved support for the SPARQL query language. The results are reported in Deliverable D6.2.

- Implement semi-automatic configuration and optimization

  We first concentrated on the role of mappings in the query answering process, and studied what affects the efficient translation of SPARQL queries formulated over (virtual) RDF data, to SQL queries. Then we addressed the challenges posed by the additional presence of an ontology formulated in the standard OWL 2 QL fragment. The results are reported in Deliverable D6.2.

Task 6.1: Transformation system configuration

► Task definition according to DoW

This task provides the techniques for the initial configuration of the various components of the query transformation system that are produced by application of the techniques developed in WP4. On the one hand, these techniques are able to exploit dependencies that are known to hold over the data layer, and on the other hand, they perform transformations of the mapping and the ontology layers so as to enforce desired dependencies over the derived virtual data.

Progress

In Task T6.1, we made significant progress towards the corresponding objectives. More specifically, withing the last 12 months we made the following progress and achieved the following results:

- We extended the fragment of the SQL language that we support in the queries in the mapping that are issued over the relational data source, by implementing a new SQL parser. As opposed to the previous version of the parser, which was a custom development, the new parser is based on the Java library JSQLParser. This makes it much more robust than before, and able to handle additional SQL queries that were not supported before.

- We extended the protege plugin of *Ontop* to support *meta-mappings*, which are mappings with variables in the position of properties or classes. Allowing for such kinds of mappings simplifies the task of the OBDA system designer, since it becomes possible to group together several traditional mappings, thus providing a more concise and compact representation of the mappings.

- We set up automatic testing of *Ontop* through Github and Travis-CI. As a consequence, every time we push changes to Github, all the JUnit tests (∼2000 tests) are run on Travis-CI automatically. This greatly improves the overall correctness of the developed codebase, and thus the stability and robustness of the system.

- Towards the wider and easier adoption of open standards and tools:
  - we integrated the new R2RML API developed in the Optique project to better support the RDB2RDF mapping language R2RML developed by the W3C.
  - we released *Ontop* under the open-source Apache license version 2. This license is permissive and friendly to both business and academic users.
  - We have deployed the *Ontop* packages in the central maven repository. Now users can integrate *Ontop* as maven dependencies into their Java applications easily.

We further observe that, during the reporting period, we released four stable versions of *Ontop*: v1.10, v1.11, v1.12, and v1.13.[1]

Meetings

- Several teleconference meetings with Fluidops to discuss release plans for *Ontop* and issues related to configuration and integration of *Ontop* within the Optique platform.

No Significant Deviations from Annex I

All Objectives achieved according to Schedule

No Corrective Actions Required

Task 6.2: Runtime Query Rewriting

► Task definition according to DoW

This task develops the techniques for rewriting, at runtime, end-user queries into queries over the data sources. The objective is on the one hand to produce queries over the data sources that are optimised towards an efficient execution, and on the other hand to keep the rewriting process as simple and as efficient as possible. The developed techniques will take into account data dependencies that hold in the system or that are derived from the mappings, the infrastructure developed in WP7, and the extensions of the query language as those considered in WP3.

Progress

During the reporting period, a significant effort has been put in understanding and taking into account the requirements coming from the Optique use cases. In particular, we analysed the queries provided by Statoil users, and the needs of other project members, and extended accordingly the formalisms adopted within Optique, and the underlying reasoning and query answering algorithms.

- We developed techniques for supporting in *Ontop* various forms of *aggregates* in end-user queries over the ontology. Specifically, we adhere to the SPARQL 1.1 standard, which has introduced aggregation operators in the SPARQL query language. The aggregation operators that *Ontop* currently supports are SUM, COUNT, AVERAGE, MIN, and MAX.
- We developed techniques for supporting the SPARQL OPTIONAL construct combined with UNION in the second argument. Observe that the OPTIONAL construct (unlike JOIN) is not distributive with respect to UNIONs, and since the process of unfolding the query with respect to the mappings introduces UNIONs, correctly dealing with OPTIONAL and UNION together required quite significant changes in the unfolding process.
- We developed techniques for supporting in *Ontop* rule based ontology formalisms, such as SWRL. *SWRL* is a rule language that is used extensively in the context of the Semantic Web and whose expressive power is incomparable to that of OWL 2 QL. For example, using SWRL, we can express

[1] `https://github.com/ontop/ontop/wiki/ObdalibPluginChangeLog`

Datalog style rules, which cannot be expressed using OWL 2 QL. In particular, we support linear recursive rules using recursive Common Table Expressions (CTEs). This extension may turn out to be useful also for the work carried out in work-packages WP4 and WP9, since in many cases they allow one to avoid expressing explicitly long chains of individuals in the query interface.

- We developed techniques for supporting the *OWL 2 QL SPARQL Entailment Regime*, which allows users to query TBox axiom templates with variables standing for classes and object/data properties of the ontology. Under OWL 2 QL SPARQL Entailment Regime one can retrieve not only domain elements but also class and property names using second-order variables.

In addition to the improvements in the *Ontop* system aimed at increasing its expressive power towards the requirements of the Statoil and Siemens Use Cases, we also developed techniques for efficient directed acyclic graph (DAG) based TBox reasoning. The new algorithm is significantly more efficient in classifying large OWL 2 QL TBoxes.

Meetings

- Nov. 2013: Research visit of Martin Rezk to Roman Kontchakov and Michael Zakharyaschev (Birkbeck, University of London, U.K.) to work on query rewriting under the OWL 2 QL SPARQL Entailment Regime. There has also been a sequence of follow-up teleconferences, including also Mariano Rodriguez-Muro (IBM T.J. Watson Research Center). The research outcome of the visit and the teleconferences was a publication that has been submitted and now accepted at ISWC 2014.
- Dec. 2013 and July 2014: Two research visits of Dag Hovland (University of Oslo) to Bolzano to work on query optimisation and the Statoil Use Case.
- June 2014: Research visit of Martin Rezk to IBM T.J. Watson Research Center to work on OBDA Optimisations. During this period he gave a talk titled "Answering SPARQL Queries under the OWL 2 QL Entailment Regime with Databases". The research outcome of this visit was submitted to the Journal of Web Semantics.
- Several teleconferences with the University of Athens, Fluidops, University of Oslo, and University of Hamburg to discuss several issues related to query rewriting and optimisation.
- Oct. 2014: Research visit of Roman Kontchakov and Michael Zakharyaschev (Birkbeck, University of London, U.K.) to work on techniques for aggregation in SPARQL in OBDA.

No Significant Deviations from Annex I

All Objectives achieved according to Schedule

No Corrective Actions Required

Task 6.3: Transformation System Tuning

► Task definition according to DoW

This task addresses the aspects related to the evolution over time of the query transformation system, considering updates to each of the system layers (ontology, mappings, and structure or availability of data sources), as well as query execution statistics and performance indicators produced by the query execution engine developed in WP7. The techniques developed in the context of this task focus on minimizing the cost of handling these updates. This includes efficient reorganization of the data back-end, and efficient re-computation of pre-computed inferences.

Progress

During the reporting period, we started the development of novel tuning techniques to produce SQLs that can be more efficiently handled by the underlying database engine. Specifically:

- We discovered that redundancy joins (in particular self joins) in the produced UCQs are often due to the fact that functional constraints (e.g., primary keys and foreign keys) that hold in the database are actually not specified explicitly. To be able to exploit such constraints for optimization, we developed techniques for users to supply such constraints manually to *Ontop*.
- We noticed that a large number of unions in the generated UCQs are often due to the deep concept hierarchy in the ontology. On the other hand, when the mapping for a concept is complete with respect to the ontology, ontology reasoning for that concept can be disabled. We developed technique to exploit this property.

Meetings

- Oct. 2014: Martin Rezk and Guohui Xiao visited Statoil to work with Dag Hovland and Martin G. Skjæveland on tuning techniques for optimizing query answering in OBDA and in paticular for the Statoil use case. The results obtained during the visit will form the basis for a submission to an international conference.

Corrective Actions

According to the DOW, Task 6.3 and the associated activities for transformation system tuning carried out in the reporting period, were planned to start at M24. However, since the performance of query answering is crucial to the project, we started these activities already at M20.

Task 6.4: Query Transformation Sub-System Implementation and Evaluation

► Task definition according to DoW

This task addresses the implementation, evaluation and testing of the sub-system for the rewriting of end-user queries into queries over the data sources. The delivered sub-system will consist of two separate components: one that deals with the transformation system configuration optimisation and tuning, and one that deals with the proper query rewriting process done at runtime. In both phases, data dependencies are used to optimize both the rewriting process and the query produced as a result of the rewriting.

Progress

Most of the extensions of the formalisms and the underlying techniques identified in Task 6.2 have been implemented in Task 6.4. More specifically, we implemented the following features:

- the support for aggregates in SPARQL queries. This extension is currently maintained in a separate GitHub Branch, and will then be integrated in Version 2.0 of *Ontop*.
- the support for SPARQL OPTIONAL with UNIONs in the second argument.
- the support for SWRL Ontologies. Also this extensions is currently in a separate GitHub Branch.
- the support for OWL 2 QL SPARQL Entailment Regime Also this extensions is currently in a separate GitHub Branch.

We also extended the support of the mapping language and query languages to handle broader mappings and queries.

- We implemented the possibility of using multi-schema queries in the query that constitutes the SQL part of the mapping assertions.
- We implemented the support for the SQL keywords IN, BETWEEN, and LIKE in the SQL part of the mapping assertions.

- We implemented the support for the BIND construct in SPARQL queries, allowing a value to be assigned to a variable from a basic graph pattern.
- We extended the support for datatypes in the database, ontology, mapping, SPARQL user query, and generated SQL query by (1) supporting more datatypes required by the use cases (specifically, temporal datatypes – see also WP05), and (2) carefully handling the situation when the datatypes for a concept defined in the mapping and ontology are different.

Finally, in line with the requirement for better performance:

- we implemented the new DAG based TBox reasoning technique;
- we improved the tree-witness query rewriting by exploiting the new DAG based TBox reasoner.

We also observe that we extended (and are constantly extending) the online documentation in the Github Wiki and project website.

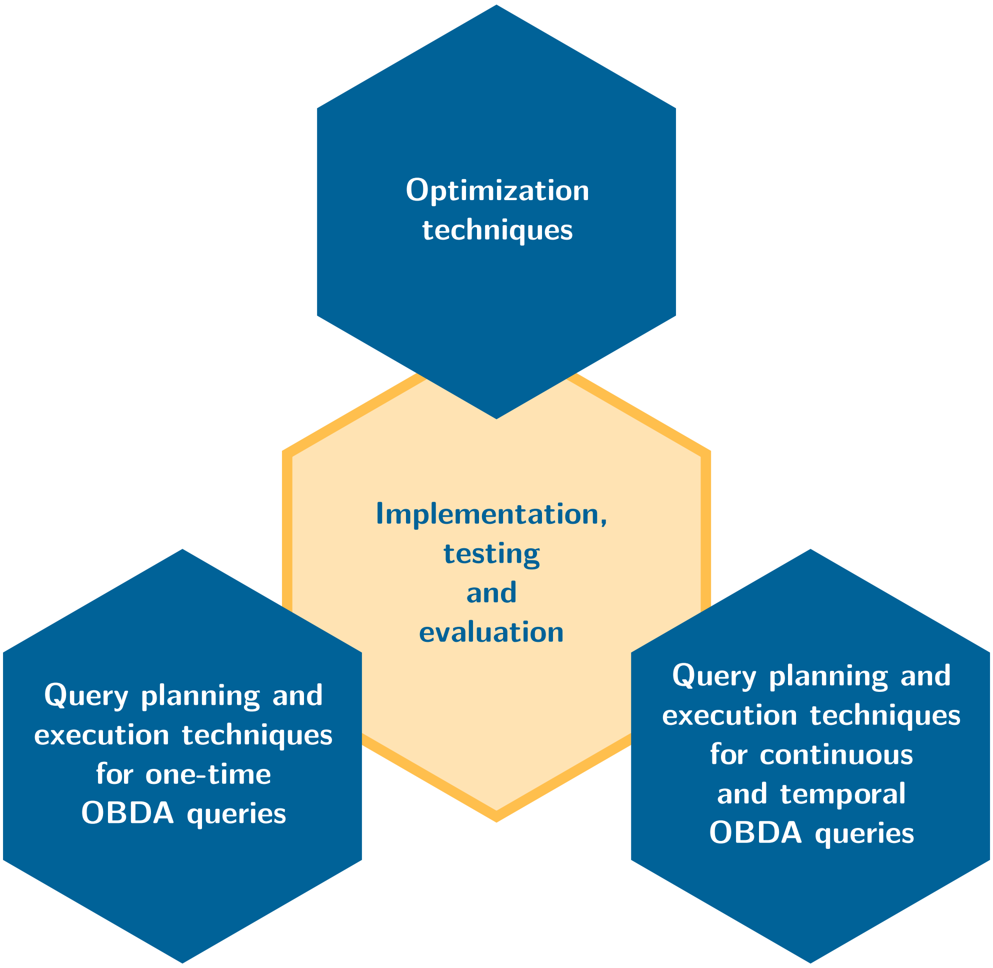
Meetings

- Dec. 2013 and July 2014: During the research visits of Dag Hovland (University of Oslo) to Bolzano reported under Task T6.2, we implemented also multi-schema queries and tuning techniques.
- Several teleconference meetings with Fluidops to discuss implementation issues, in particular R2RML.
- June 2014 and Oct. 2014: Research visit of Guohui Xiao and Diego Calvanese to OKKAM (Trento) and meeting in Bolzano with OKKAM management (Paolo Bouquet and Andrea Molinari), to discuss about possible research and development collaborations. A demo of the *Ontop* system was given to OKKAM staff.
- Sep. 2014: Research Visit of Martin Rezk and Guohui Xiao to Athens, to collaborate with Dimitris Bilidas, on the federation of OBDA using *Ontop* and ADP.
- Oct. 2014: During the ISWC 2014 conference (Riva del Garda, Italy), several meetings were organised for the core developers of *Ontop*, including Martin Rezk, Benjamin Cogrel, Guohui Xiao, Roman Kontchakov (Birkbeck, University of London), and Mariano Rodriguez-Muro (IBM), to discuss the code quality and the future development of *Ontop* within Optique.
- Oct. 2014: Research visit of Roman Kontchakov to work on the optimization of TBox reasoning implementation in *Ontop* to improve the performance of the classifications of OWL 2 QL ontologies.

Corrective Actions

According to the DOW, the implementation activity for the techniques developed in the reporting period in Task 6.2 was planned to start at M18. However, we started early already at M14 to complement the theoretical developments reported for Task 6.2 with implementation work on the query transformation sub-system. This allowed us to get useful early feedback about the practical feasibility and implementability of the developed techniques, which had a positive impact on the theoretical development.

WP7: Distributed Query Execution



HELLENIC REPUBLIC
National and Kapodistrian University of Athens

Work Package 7: Distributed Query Execution

WP7 Main Objectives (from Table 2.2) and progress made:

Demonstrate federated query evaluation over multiple Statoil sources, and streaming query evaluation over Siemens data

- Improve the integration of ADP into the Optique platform

  A special kind of federated data source for ADP has been created in the Optique platform. In this data source the user can define other existing data sources as endpoints. Also, a stream data source has been created in the platform, that can communicate with the streaming version of ADP through a REST API.

- Support streaming sources, continuous queries

  ADP has been extended to provide stream processing functionality and enable users to define streams that can be queried and manipulated in a manner similar to regular tables. Operators have been defined for connecting to external stream sources and streaming versions of relational operators were implemented in order deal with streams that (unlike relational tables) are infinite. Queries that combine static data and streams are possible, and a REST interface was implemented for registering queries and fetching stream data.

- Install ADP at Statoil, Siemens; simplify the installation process

  A console for users was created, that will automate the installation and monitoring process. Statoil has agreed to install the system in a cluster of 8 machines.

- Execute test queries over federated sources at Statoil

  Simple federated queries were tested, that combine information from the SLEGGE and the NPD data sources.

- Execute test queries over streaming sources at Siemens

  An initial version of the OPC driver was implemented, that periodically queries sensors (e.g., every 10 seconds) and outputs the values as tables in ADP. To be able to use the anonymized data provided by Siemens, sensor simulators were developed that generate streams based on values taken from files. The functionality of the system and the queries will be the same when the real sensors will be used.

- Optimize and adapt ADP (e.g., for Oracle)

  The ADP query optimizer was developed, that incorporates distributed statistics, cost-based optimization techniques and network caching in order to optimize execution of complex star-join and analytical queries in the three available modes of execution: In normal mode database tables reside in ADP and all the processing takes place there. In federated mode tables are stored in external databases. Some processing primitives are pushed in the corresponding systems. In multi-query optimization mode common sub-expressions are exploited in different unions or nested queries.

Task 7.1: Query planning and execution techniques for one-time OBDA queries

► Task definition according to DoW

This task develops the basis of the Optique query planning and execution techniques by studying what extensions are necessary to ADP in the context of Optique. Specifically, the task targets the efficient execution of typical queries produced by OBDA translation components (SQL queries with nested subqueries and user defined functions), also taking into account the requirements on data reorganization and inference materialization coming from the translation system configuration component developed in Task 6.1.

Progress

The main objectives of this task were fulfilled during the first year of the project. Work related with this task is extended in the context of optimization techniques (see Task 7.3)

Meetings As this task was active during the first year of the project, we mention the WP7 related meetings in Tasks 7.2 and 7.3

No Significant Deviations from Annex I

All Objectives achieved according to Schedule

No Corrective Actions Required

Task 7.2: Query planning and execution techniques for continuous and temporal OBDA queries

► Task definition according to DoW

This task develops the second version of the query planning and execution techniques, following Task 7.1. ADP will be properly integrated with the query translation components, including the ability to provide feedback to the translation system tuning process of Task 6.3, and preliminary techniques for continuous and temporal queries.

Progress

The progress that has been made in the context of this work task can be summarized as follows:

- **Streams**. During the reporting period and through our close collaboration with TUHH from which we received requirements and feedback, we implemented a number of extensions to the ADP for the efficient execution of stream queries inside ADP. The most significant of these extensions are the following:
  - The "sliding window" operator. This operator defines a sliding window, by assigning a window id to a -fixed- amount of tuples each time
  - The "time sliding window" operator. This operator defines a sliding window by assigning a window id to a group of tuples whose timestamp is inside a time range. Each window is an ABox sequence, and the ABox index is also printed. By default, the tuples belong to the same ABox if they share the same timestamp. The user can also opt for alternative sequencing functions.
  - Joining multiple dynamic streams. Joining two or more dynamic stream is not a trivial task. One of the challenges that we encountered was the fact that the streams should be synchronized. Another challenge was how null values (timestamps with no information) would be handled. According to our current implementation, the user can now perform even a full outer join between two streams. To perform a full outer join on a time column we had to implement two separate operators, named wcache and tmerge respectively. The wcache operator performs a left outer join with an extra information on tmerge. The tmerge operator is a row operator that takes the result of wcache and produces the full outer join.

Meetings

- Participation at the regular WP5 conference calls as far as requirements for stream and temporal queries are concerned
- Organization of a technical meeting in Athens on April 8 and 9. Apart from UoA, colleagues from TUHH and Siemens attended the meeting and we discussed technical aspects of unfolding STARQL queries and the use of ADP in this context.

No Significant Deviations from Annex I

All Objectives achieved according to Schedule

No Corrective Actions Required

Task 7.3: Optimization techniques

► Task definition according to DoW

This task develops query optimization techniques for the ADP extensions designed in Task 7.1 and 7.2 for one-time, continuous and temporal queries. These include techniques for query rewriting, operator implementation choice, scheduling, adaptive execution and recovery, and multi-query optimization.

Progress

This Task was scheduled for the third year of the project but some development was necessary during the reporting period (See corrective actions below). The main activities were as follows:

- Implementation of an ADP component that is responsible for federated query execution.
- We have implemented and integrated into the platform a version of the ADP JDBC driver, that can use other relational data sources already defined in the platform as endpoints, for the federated query execution.
- We started implementing a caching over the network component that decides which temporary data can be reused in order to reduce network traffic. This component employs algorithms for replacement of temporary data in ADP that take into consideration the requirements for data freshness of the original data sources, as well as parameters that affect the time needed for importing the data.
- We started implementing a join reordering optimizer based on histograms. This optimizer reorders the joins in the SELECT-FROM-WHERE parts of each query that we receive from the query transformation component (usually a union of conjunctive queries). We also group together joins that can be executed completely in parallel and push them in the local nodes of the system to be executed together.

Meetings

- We have participated to the related conference calls (WP2 monthly calls, federation working group calls and calls regarding the use cases of WP8 and WP9)

Corrective Actions

This task was scheduled to start in the third year of the project, but during the collection of the user requirements, it became evident that working on federated query processing should start earlier. In order to achieve this, we started working on federation and considering related optimizations. These optimizations were on a different level for what we were targeting in the first year. During the first year our main focus was in the data-flow optimization, whereas now we are examining the optimizations that needs to be done prior to the creation of the data-flow and are about the ordering and grouping of parts of the SQL query (pushing as much processing as possible to the original data sources in the federated scenario, join ordering to reduce intermediate results, grouping of joins that can be executed together, pushing of filters and projections as early as possible etc)

Task 7.4: Implementation, testing and evaluation

► Task definition according to DoW

This task comprises the implementation, testing and evaluation of all the functionalities developed in T7.1–T7.3 as parts of the integrated Optique platform.

Progress

- This task has not officially started yet, but we already have integrated the ADP JDBC drivers inside the Optique platform and we have set up an ADP instance to work in fully distributed mode inside the FluidOps infrastructure.

No Significant Deviations from Annex I

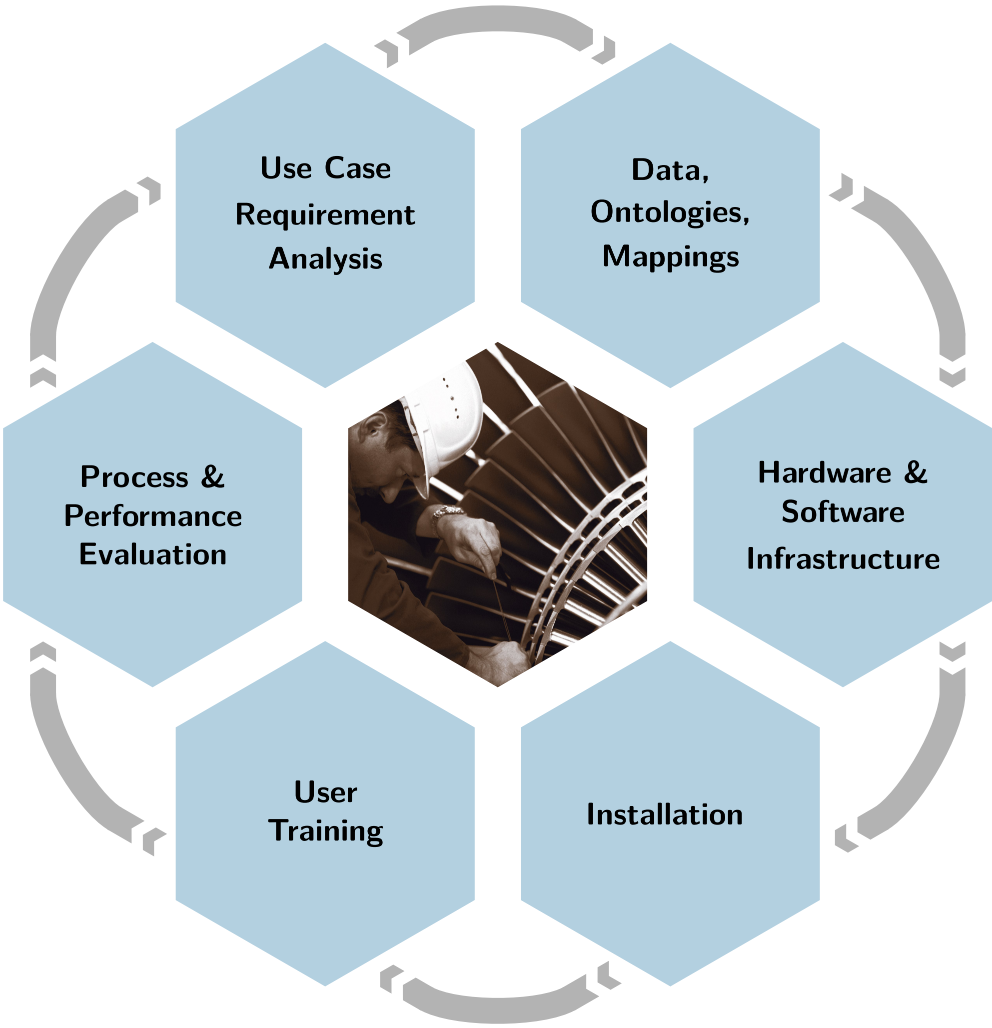
All Objectives achieved according to Schedule

No Corrective Actions Required

WP8: Siemens Use Case

SIEMENS

Siemens Energy Services runs service centers for power plants, each responsible for remote monitoring and diagnostics of many thousands of gas/steam turbines and associated components.



Work Package 8: Siemens Use Case

WP8 Main Objectives (from Table 2.4) and progress made:

Demonstrate an integrated platform running at Siemens and answering realistic streaming queries

- Install the Optique platform on Siemens premises

  In collaboration with software developers from Fluid Operations, Siemens has succeeded in customising and installing the Optique platform within the Siemens IT infrastructure. More importantly, Optique runs on a server within the highly secured de-militarised zone and has access to real-world turbine data. This means that, for the first time, engineers have the ability to access their 'usual' data through the Optique platform and its respective web front-end.

- Prepare a test installation of streaming data

  In addition to providing access to 'historic' data retrospectively, the Optique platform at Siemens is capable of consuming, processing, and rendering streamed data. This could be demonstrated using streams generated from actual turbine data taken from the Remote Diagnostci Centres.

- Initial evaluation of the platform at Siemens

  The query-catalog has been set up and tested in the Optique platform at Siemens. Preliminary results indicate that the queries in the catalog could be evaluated successfully. In order to meet Siemens' use-case, bespoke queries for fault detection have been added. Evaluations were conducted at Siemens in Munich and during the end-user workshop at Siemens in Oslo in October.

Task 8.1: Infrastructure

► Task definition according to DoW

The purpose of this task is to set-up, operate, and maintain the hardware and software infrastructure for providing test data sets to the partners, and testing the Optique prototypes in a use case specific environment.

Progress

- *Hosting Optique at Siemens*
  For the Year-two prototype Siemens has allocated space on a more powerful machine – providing 4 Intel Xeon processors and 512GB of RAM. More importantly, the machine is situated in a secure zone within the Siemens intranet and has access to actual turbine data. Hence, the original signal and event message data becomes accessible to Optique.

- *Better test data generation*
  Siemens has updated the work-flow for generating test data of a kind that can be sent to project partners without violating Siemens' security regulations. This simplifies the creation of larger data sets and hence provide a basis for realistic performance testing.

- *Large test data set*
  Utilising the improved work-flow mentioned above, Siemens has created a larger data set of anonymised turbine data for testing by project partners. The new data set is about 2 Terabytes in size and hence represents an amount of data of the same order of magnitude as the actual database used by the Remote Diagnostic Centres for gas turbine diagnostics.

Meetings

- *Meetings with Siemens IT department*
  The main purpose of these meetings has been to find and implement viable options for running the Optique platform at Siemens in such a way that access to actual turbine data would be

possible without violating Siemens' guidelines for data security. As the data resides within a highly secured zone, several straightforward options (e.g., opening ports or moving the relevant databases) were rejected.

No Significant Deviations from Annex I

All Objectives achieved according to Schedule

No Corrective Actions Required

Task 8.2: Siemens Use Case Requirements Analysis

► Task definition according to DoW

This task analyses requirements specific for the Siemens use case, thereby providing input for the joint requirements analysis in work package WP1. Additionally, the task encompasses sub-tasks that ensure a thorough understanding of user needs for query formulation as well as methodologies currently employed for query formulation, and the current state of the art in query formulation tools.

Progress

- *Further analysis of use-case "predictive analysis"*
  Siemens CT has conducted additional interviews with engineers from the Remote Diagnostic Centres to ensure that the use-case developed within the Optique project is aligned with the actual work of the diagnostic engineers. The key task of their predictive work is to anticipate major malfunctions before they happen. This involves analysing both short-term (24 hours) historical data as well as life data streams.

- *Definition of Failure Mode Scenarios*
  The insight gained from direct collaboration with diagnostic engineers has led to the definition of some core scenarios that reflect common cases of the engineer's daily diagnostic work. Examples of such failure modes are: discrepancies between pairs or groups of signals, unexpectedly high rates of change of signals, or certain patterns in the event message queue.

- *Update of query catalogue*
  In order to provide effective assistance for the diagnosis of the failure modes mentioned above, Siemens CT has updated the query catalog accordingly.

Meetings

- *Internal meetings with end-users*
  Siemens CT has reached out to diagnostic engineers at the Remote Diagnostic Centres in Lincoln, UK, in order to capture the requirements for failure mode analysis as mentioned above. In addition to a better understanding of failure-mode-driven diagnostics, the meetings have provided valuable requirements for the user-interface of the optique demonstrator.

No Significant Deviations from Annex I

All Objectives achieved according to Schedule

No Corrective Actions Required

Task 8.3: Data Models

► Task definition according to DoW

Siemens has a collection of informal and semi-formal models currently used for turbine analysis. Based on this collection, the following steps are performed in this task: Characterisation of data sources and associated data models, characterisation of existing models, provisioning of an ontology for turbine diagnostics, mapping to pre-existing models. As models may evolve during the project runtime, this task requires several iterations.

Progress

- *Finalised mappings*
  As part of the set-up of the Optique platform at Siemens, the relevant mappings between ontologies and the internal Siemens database has been finalised. This has enabled Siemens, for the first time, to access actual turbine data from the Optique platform.
- *Access to extended schema*
  While access to actual 'production data' of Siemens gas turbines has been a major step for the set-up of Optique at Siemens, that data is being provided in a simplified schema. This has limited the utility of algorithms developed by Optique partners for bootsrapping the domain ontology from the database schema. Siemens CT has now succeeded to obtain the original database schema of the actual production system and has started testing of bootstrapping on the baseis of the extended schema.

Meetings

- *Coordination with technical work packages*
  In particular, find a workflow in which bootstrapping experiments can be conducted at Siemens without violating Siemens' security rules. The bootstrapping algorithms need to be applied to the extended schema of the production system which is classified as confidential.

No Significant Deviations from Annex I

All Objectives achieved according to Schedule

No Corrective Actions Required

Task 8.4: Installation preparation and tool development

► Task definition according to DoW

In order to enable diagnosis engineers to use the Optique platform developed in WP2, a seamless integration with existing Siemens diagnostic tools and processes is mandatory. This requires the infrastructure set up in 8.1 to be prepared and an overall system architecture for the use case has to be developed. Preparation of the infrastructure includes the installation of required software packages, granting access rights to selected users and ensuring connection between the Optique system and the diagnostic databases. Development of the use case software architecture comprises design and implementation of software components that provide interface adaptors to the existing tool chain and add-ons for use case specific functionalities that support query formulation and result presentation.

Progress

- *Stream generator*
  In order to utilise the stream reasoning capabilities of Optique, a stream generator based on real historic data has been implementated and integrated into the Optique platform at Siemens. This allows Siemens CT to develop and refine solutions for stream-reasoning before deploying them in the domain of Sienens Energy who operates the Remote Diagnostic Centres where, among other things, data streams are being processed.

- *Use-case specific visualisation widgets*
  In collaboration with Fluid Operations, Siemens has enhanced the data visualisation widgets used by the Optique platform at Siemens. The main challenges here were related to the large amount of data generated by turbine control systems that had to be visualised meaningfully.

Meetings

- See Task 8.5 below.

No Significant Deviations from Annex I

All Objectives achieved according to Schedule

No Corrective Actions Required

Task 8.5: Installation

▶ Task definition according to DoW

This task sets up the use case demonstrator based on the prototype provided by WP2 and the architecture specified in task 8.4. Upon availability, versions of the Optique software system are installed, and interfacing with existing Siemens systems is ensured. If models are updated, integration with the running system is ensured. Periodically, the data provided for the test-bed by Siemens may be updated (e.g. on significant changes or additions).

Progress

- *Year-2 Optique platform successfully installed*
  The single most significant achievement for Year 2 from a Siemens perspective has been the deployment of the current Optique platform at Siemens and its connection to actual turbine data from the Remote Diagnostic Centres. This has, for the first time in the Optique project, allowed Siemens to test the Optique platform on realistic data.

- *Initial evaluation of query catalog*
  Installing its own instance of the Optique platform with access to realistic data is a key requirement for a an evaluation of Optique under realistic conditions. Siemens has started to conduct systematic performance tests in order to evaluate whether the relevant queries from the query catalog can be evaluated on realistic data sets.

Meetings

- One two-day and two one-day workshops at Siemens in Munich together with Fluid Operations AG for specifying, configuring, and installing the Optique platform at Siemens. Moreover, these meetings were used for gathering requirements related to the visual presentation of turbine data by the Optique platform.

No Significant Deviations from Annex I

All Objectives achieved according to Schedule

No Corrective Actions Required

Task 8.6: User Training

► Task definition according to DoW

In this task, diagnosis engineers and modelling experts practice using the new tool chain. They apply Optique to a selection of concrete cases, re-expressing familiar queries in the ontology language (Task 8.3). They execute the queries in the Optique client to feed their current portfolio of analytic software. The targeted applications and data sources include Siemens Service Dashboard, trend analyser, and sensor and event databases. The training done in this task is a prerequisite for the performance evaluation conducted in task 8.8. It furthermore serves to test the Optique query formulation tools (WP3) on real cases.

Progress

Task for year 2 not started.

Meetings No meetings.

No Significant Deviations from Annex I

All Objectives achieved according to Schedule

No Corrective Actions Required

Task 8.7: Performance Evaluation

► Task definition according to DoW

The query catalogue established in Task 8.2 serves as a basis for a thorough evaluation of system performance focusing on temporal and continuous queries. The correctness of the integrated system is tested in this task against specific Siemens use case scenarios and huge data sets using evaluation and test frameworks such as those developed, for instance, in the SEALS project12. Since some query results are used to provide support for inhibiting certain decisions, completeness is another issue to be evaluated in Optique. For this purpose, for the Siemens query catalogue the expected answers are recorded and used for automatic comparison. In addition to correctness and completeness, performance is another important issue to be analysed w.r.t. several dimensions and Siemens data.

Progress

Task for year 2 not started.

Meetings No meetings.

No Significant Deviations from Annex I

All Objectives achieved according to Schedule

No Corrective Actions Required

Task 8.8: Process Evaluation

► Task definition according to DoW

Process evaluation done in this task comprises evaluating user performance and satisfaction as well as result quality. The query catalogue established in Task 8.2 serves as a basis for this evaluation. User performance is determined by comparing overall query formulation time of experts (including human interaction) using the old system with experts using the Optique platform.

Progress

- *End-user workshop*
  Siemens CT has conducted the end-user workshop at Siemens AS in Oslo, Norway, on 30 October 2014. For the first time, engineers from the Remote Diagnostic Centres both in Lincoln, UK, and Finspang, Sweden, have evaluated the Year-2 demonstrator. This evaluation involved performing realistic diagnostic tasks on actual trubine data provided by Siemens.

- *Feed-back from End-user workshop*
  At the end-user workshop in Oslo, the diagnostic engineers at Siemens have used the Optique platform under the supervision of Optique partners and, additionally, have responded to two questionnaires about their impression of the work with the Optique platform. The results of both the questionnaire and the observations made during the trials are currently being evaluated.

Meetings No meetings.

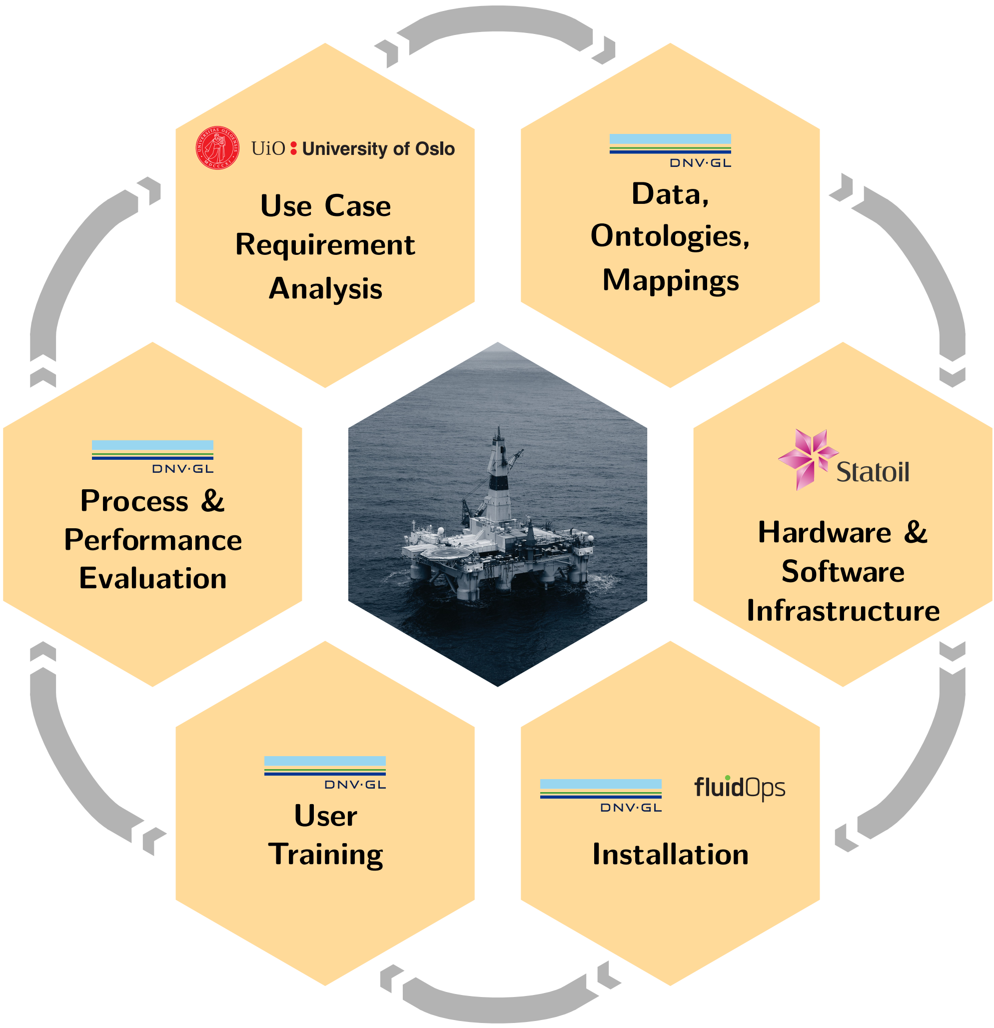
No Significant Deviations from Annex I

All Objectives achieved according to Schedule

No Corrective Actions Required

WP9: Statoil Use Case

At **Statoil Exploration,** experts in geology and geophysics develop stratigraphic models of unexplored areas on the basis of data acquired from previous operations at nearby geographical locations.



Work Package 9: Statoil Use Case

WP9 Main Objectives (from Table 2.4) and progress made:

Demonstrate an integrated platform running at Statoil and answering realistic queries over EPDS and NPD FP

Queries can be formulated, including the use of a new map widget, and run from the VQS in the Optique platform installed on Statoil, towards the EPDS database. Results can be exported as KML files from the platform, and these can be imported into the tool ArcGIS. The administration of the platform includes data source and repository setup, ontology and mappings bootstrapping and the management of these, including users and queries.

- Test and evaluate bootstrapping component on EPDS

  WP4 bootstrapping software has been run on EPDS. The resulting ontology and mappings have been evaluated towards the query catalogue and aligned with other ontologies.

- Study usability of VQS in Statoil

  The VQS loaded with an ontology for well and stratigraphic data has been evaluated in the end-user evaluation meeting with Statoil in September.

- Evaluate and optimise performance of query execution over EPDS

  The queries in the query catalogue have been used extensively for evaluation of performance and tuning of query answering in OBDA in cooperation with WP6.

- Test federated queries over EPDS and NPD FP

  Both SPARQL-level federation with FedEx, and database-level federation with Oracle's multi-schema, have been used for simple queries. database-level federation using ADP is in development.

Task 9.1: Infrastructure

► Task definition according to DoW

The purpose of this task is to provide an operational infrastructure for the Statoil use case. This task starts with setting up the requirements for both the hardware and the software.

Progress

- Optique platform installation at Statoil is updated periodically. This is done in cooperation with WP2. The Optique platform with WP6 software now works on Statoil infrastructure.
- Initial work on querying with geospatial primitives is done in cooperation with WP7.
- Making query results from the Optique platform available in ArcGIS with the KML format.
- We have improved the technical coordination of the use of the Statoil terminal servers, especially the transfer of software, ontologies and mappings from the Optique development server at FOP to the Statoil server.
- Work has started on setting up a network of 8 servers for efficient deployment of the WP7 software, and also 2 separate servers for running of the Optique platform.

Meetings Communication mainly via email and video conferencing.

No Significant Deviations from Annex I

All Objectives achieved according to Schedule

No Corrective Actions Required

Task 9.2: Statoil Use Case Requirement Analysis

► Task definition according to DoW

This recurring task performs a requirement analysis for the Statoil use case at the beginning of every phase and serves as input for the Requirement Analysis Framework as defined in D1.2 in Work Package WP1.

Progress

- Meetings with domain experts at Statoil have clarified some domain questions, and also provided a new question pattern.
- A member of the UiO team has been given permanent office space at Statoil in Stavanger. This has lead to increased informal contact and better understanding.
- The query catalogue has been collected and analyzed. Keywords have been extracted, and a catalogue containing both natural text, SPARQL, generated SQL, and run times has been compiled.
- We have discovered that database objects which do not have constraints (e.g., views) are important for the use case.
- We have improved our understanding of the workflows into which a Statoil installation of the Optique platform may fit.
- We have discovered that integration with the tool Petrel from Schlumberger would be important for end-users to see the relevance of the Optique platform.
- Our understanding of the ProSource tool at use in Statoil has improved.

Meetings

- Video meeting, April 3 2014, Statoil, UiO
- Video meeting, April 9 2014, Statoil, UiO
- Oslo, January 30–31, 2014, UOXF, UiO
- Oxford, March 10–14, 2014, UOXF, UiO
- Video meeting, May 8 2014, Statoil, UiO
- Video meeting, June 18 2014, Statoil, UiO

No Significant Deviations from Annex I

All Objectives achieved according to Schedule

No Corrective Actions Required

Task 9.3: Data models

▶ Task definition according to DoW

Statoil uses slightly different data models for the databases that are relevant for the use case. The data models are based on industry standards like the ones provided by POSC Epicentre and Professional Petroleum Data Management (PPDM). The different choices made in the implementation of the data models are in fact irrelevant for the exploration expert with a specific data need. He/she typically formulates his request in terms of some stratigraphic model that is relevant for his/her need and that model may be different from the models used for the data sources. This task will bridge the gap between domain and data source models.

Progress

- Bootstrapped ontologies and mappings from the EPDS schema have been constructed.
- The manually created ontology for the query catalogue has been developed in cooperation with WP4, and is now quite stable, while the mappings are undergoing work.
- We have merged the manually created and bootstrapped ontologies and mappings for use in the Optique platform.
- WP9 has taken part in development of the *International Stratigraphic Chart Ontology*, which we wish to include in the Optique platform at Statoil.

Meetings

- November 6 2013, UNIROMA1, UiO: ontology development
- November 13 2013, UNIROMA1, UiO: ontology development
- November 22 2013, UNIROMA1, UiO: ontology development
- November 29 2013, UNIROMA1, UiO: ontology development
- December 17 2013, UNIROMA1, UiO: ontology development
- February 17 2014, UNIROMA1, UiO: ontology development
- May 28 2014, UNIROMA1, UiO: ontology development

No Significant Deviations from Annex I

All Objectives achieved according to Schedule

No Corrective Actions Required

Task 9.4: Installation preparation and tool development

▶ Task definition according to DoW

In order to allow exploration experts at Statoil to use the Optique platform developed by WP2, it needs to be seamlessly integrated with existing Statoil exploration tools and processes like for instance the visual analysis tool ArcGIS. To this end, the infrastructure which has been set up in task 9.1 has to be prepared and an overall system architecture for the use case has to be developed. The preparation of the infrastructure involves the installation of required software packages, granting access rights to selected users, ensuring connection to the diagnostic databases, etc. The development of the use case software architecture requires to design and implement software components that provide interface adaptors to the existing tool chain and add-ons for use case specific functionalities that support query formulation and result presentation.

Progress

- Cooperation with WP2 and WP7 on exporting data to GIS tools.
- We have been given access to use ArcGIS Desktop at Statoil premises.
- We have been explained how to use the custom spatial functions in EPDS.
- Accessing several database schemas in EPDS has been made possible with the WP6 software.
- We have started investigating how to build an Optique plugin to *Petrel* from Schlumberger.
- We have experimented with providing the WP6 software with constraints that are not present in the database metadata (e.g., uniqueness constraints for columns in views that are based on primary keys).

Meetings

- Bolzano, December 9–13 2013: FUB, UiO
- Bolzano, June 30–July 4 2014: FUB, UiO
- Stavanger, October 1–October 10: FUB, UiO

No Significant Deviations from Annex I

All Objectives achieved according to Schedule

No Corrective Actions Required

Task 9.5: Installation

► Task definition according to DoW

The purpose for this recurring task is to install the solutions from the Implementation Infrastructure Work Package, WP2, on the Statoil use case testbed as prepared by Task 9.4. Upon availability, updated versions of the Optique software system are installed, including an integration with existing Statoil tools and processes. If models are updated, integration with the running system are done here as well.

Progress

- WP2 has made available a build of the Optique platform on the internal Optique wiki. This simplifies installation at Statoil greatly, and new builds can now be tested with little effort.
- A patch for Ontop which makes it better at accessing data which is stored in views (common in EPDS) has been developed.

Meetings Communication mainly via email and video conferencing.

No Significant Deviations from Annex I

All Objectives achieved according to Schedule

No Corrective Actions Required

Task 9.6: User Training

▶ Task definition according to DoW

In this task, explorationists and IT-experts practice with the Optique tool chain. They apply Optique to a selection of concrete cases by re-expressing familiar queries in the ontology language (task 9.3). For the exploration experts, this training will focus on using the Optique client to feed their current visual analysis tool ArcGIS. For the IT-experts, the training will also include the tools for model maintenance.

Progress

- Presentation for Statoil it staff in October 2014 with live demonstrations of the platform and Protégé with Ontop.

Meetings

- Stavanger, October 3 2014: FUB, UiO, Statoil

No Significant Deviations from Annex I

All Objectives achieved according to Schedule

No Corrective Actions Required

Task 9.7: Performance Evaluation

▶ Task definition according to DoW

Performance testing determines the actual runtime and memory consumption of the Optique system for static queries in the Statoil use case. The query catalogue established in task 9.2 serves as a basis for this evaluation.

Progress

- Reports on automated test runs have been compiled and are updated regularly.
- Comparisons of different tuning and optimisation techniques in cooperation with WP6.

Meetings

- Stavanger, April 22–24 2014

The three meetings under Task 9.4 were also considering performance evaluation.

No Significant Deviations from Annex I

All Objectives achieved according to Schedule

No Corrective Actions Required

Task 9.8: Process Evaluation

▶ Task definition according to DoW

Process evaluation comprises both evaluating user performance and satisfaction, and result quality. It will also include the usability of the tools by both exploration experts and IT-experts. The query catalogue established in task 9.2 serves as a basis for this evaluation.

Progress

- Results of some of the SPARQL queries run with Ontop have been compared with existing tools and discussed with Statoil.
- The visual query system was studied in the end-user evaluation at Statoil in September, 2014.

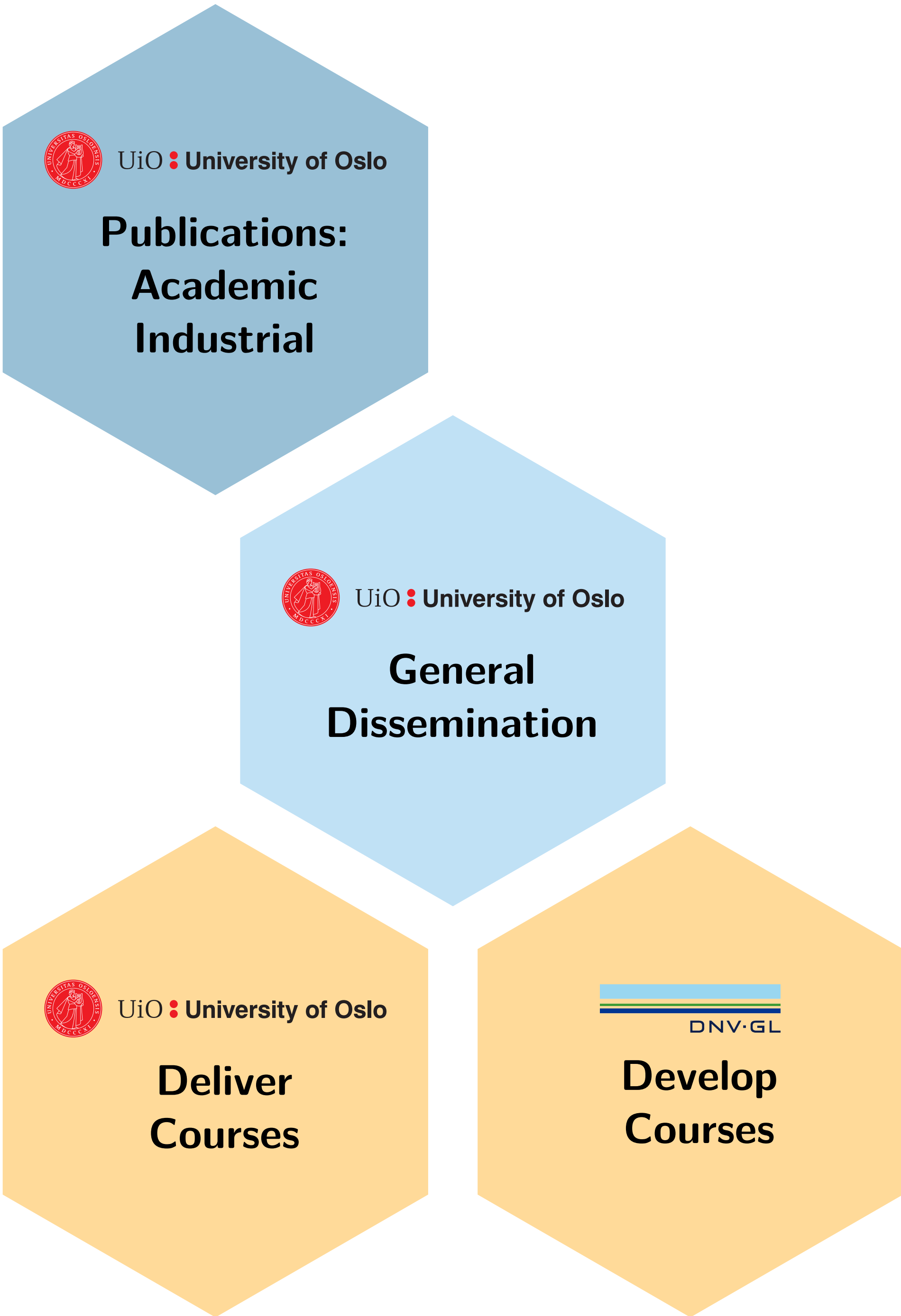
Meetings

No Significant Deviations from Annex I

All Objectives achieved according to Schedule

No Corrective Actions Required

WP10: Dissemination



Work Package 10: Dissemination

WP10 Main Objectives (from Table 2.5) and progress made:

Enhance visibility of Optique in academia, industry, and to the general public

- Continue academic publishing and presentation activities

  In Year 2 the Optique consortium published more than 50 refereed publications, gave more than 50 presentations, and gave 5 keynote talks.

- Continue presentation activities for the industrial consortium partners and targeted industry

  In Year 2 the Optique consortium gave 25 presentations to the industrial consortium partners, 30 on site presentations to external companies, and 6 presentations at industry conferences.

- Establish the Optique Partner Programme and prepare training workshops for end users

  The Optique Partner Programme with the first event organized by DNV GL. The first training workshop is scheduled at the beginning of Year 3 for a member of the Partner Programme.

- Create white papers and other user-oriented documentation

  An executive white paper produces, several others are by the end of Year 2 in progress.

- Enhance Optique's web presence, including providing public access to the Public showcase

  The project has produced 9 YouTube videos with in total 566 views in Year 2. The web page had 15.000 visits. The public showcase has been installed with most of the features openly available. The source code of Ontop has been migrated to a public Git repository hosted on Github; only in the last two weeks, the Ontop Github site received +2000 hits.

Task 10.1: Publication

► Task definition according to DoW

Optique will publish its results extensively as:

- Academic publications
- Industry publications

Progress

A significant number of manuscripts have been produced for international conferences, with key performance indicators below in Table 3.2 and in Table 3.3. For more details, please refer to deliverable D10.5, Chapters 2 and 3.

Table 3.2: Visibility to the research community.

Measure	Y2	Y1	Total
Publication count	54	51	105
Journal publications	2	4	6
Conference Workshop publications	52	47	99
Keynotes at Scientific Events	5	1	6
Workshop and Conference Presentations	39	49	88
Other presentations for the research community	13	0	13

Table 3.3: Visibility to the target industry.

Measure	Y2	Y1	Total
Presentations on-site for Industrial Consortium Partners	25	25	50
Presentations on-site for External Companies	30	10	40
Workshop and Conference Presentations for Industry	6	9	15
Presentations at Optique arranged Conferences	1	0	1

Task 10.2: General Dissemination

► Task definition according to DoW

This task targets the visibility of the project to a broad public, beyond scientific peers and immediate industrial adopters, and covers:

- Project Fact sheet
- Media Visibility
- Graphical profile
- Website
- Dissemination through online channels
- Public Showcase
- Clustering and Concertation

Progress

For key performance indicators for dissemination to the general public, cf. Table 3.4. For more details, please refer to deliverable D10.5, Chapter 4.

Table 3.4: Visibility to the general public.

Measure	Y2	Y1	Total
Newspapers and Magazine Articles	0	7	7
Miscellaneous Talks and Presentations	3	13	16
Website visits	15174	13000	28174
Unique website visitors	6834	5365	12199
YouTube videos	4	5	9
YouTube views	566	174	740

Task 10.3: Course development

► Task definition according to DoW

Development of the Optique training programme is the largest single effort of WP10. The WP will deliver a curriculum for the Optique methodology, providing guidance on organization, work processes, and data management that targets several layers of the enterprise, from management to domain experts and IT staff.

Progress

The Optique *Training Programme* is described in deliverable D10.5, chapter 3.7. Year 2 activities have focused on developing white papers and providing good learning resources. Optique is reaching out to selected industry users, initially in Norway, in order to build to the Optique Partner Network of early adopters.

Development of documentation and training material during Year 2 has targeted a general audience, in line with the target of recruitment and wider outreach.

With the use case partners Siemens and Statoil, some training activities have accompanied the development tasks, although the project has been in a very active development phase. End users have been engaged in particular in building query catalogues and user interfaces.

Task 10.4: Deliver courses, in workshops and online

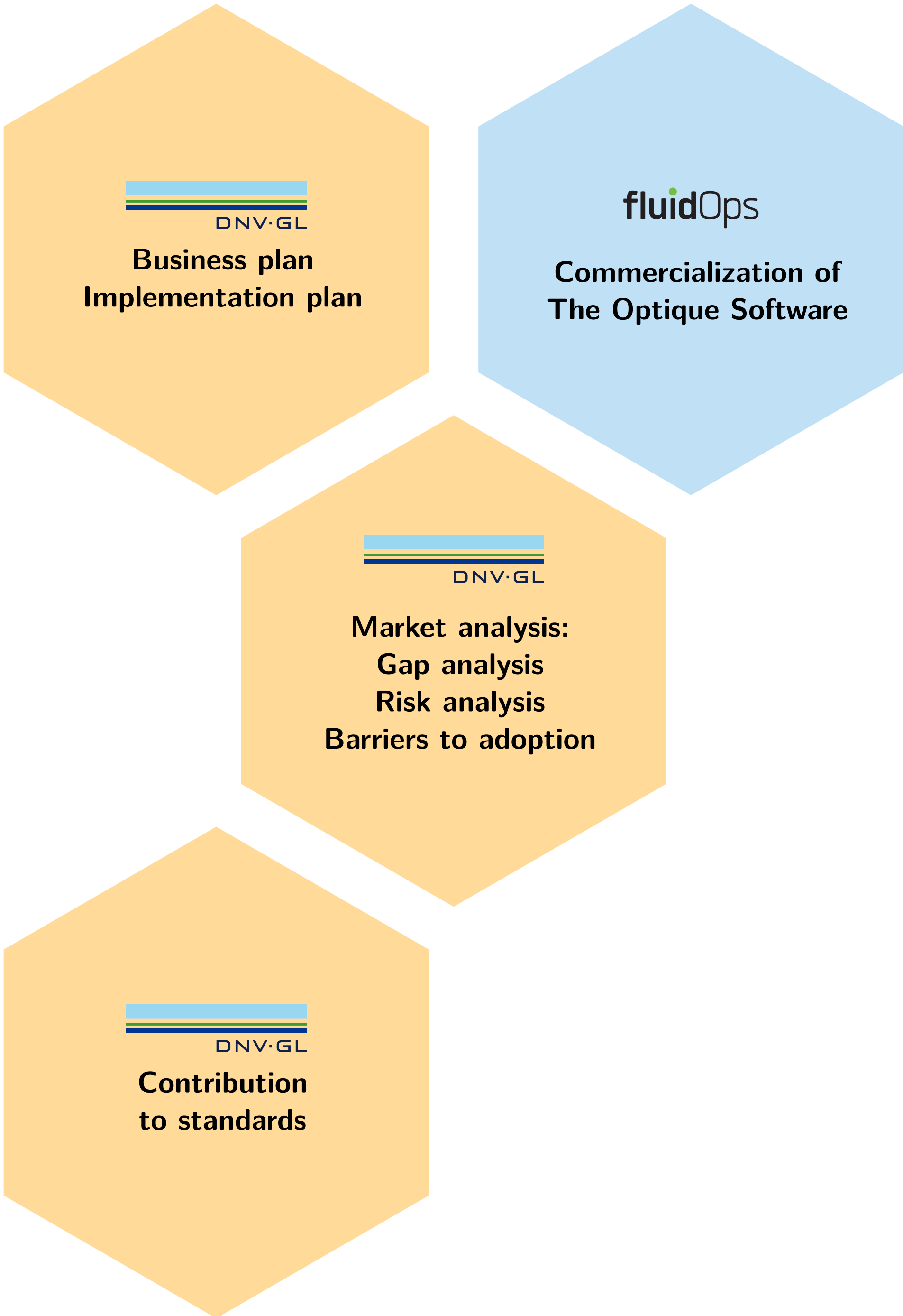
► Task definition according to DoW

Training for industry will target different roles, including: Domain experts, who use Optique to express queries in their specialist terminology; data modellers who work with experts to capture engineering domains in ontology; ICT staff specialised in corporate data stores, who set up mappings from databases to ontologies.

Progress

This task is scheduled to start with initial workshops during Spring 2015.

WP11: Exploitation



Work Package 11: Exploitation

Year 2 objectives:

- Create the Optique Alpha package for pilot customers

  For the planned March 2015 project partner network meeting the intention is to hand-out an Optique demonstrator to get a first "hands-on-experience" to the participants. Work to specify and build the "Optique Alpha package" is ongoing. A predefined example database will be included in the package accompanied with Optique introductory material.

- Implement a test installation at an Optique Pilot Partner

  An Optique prototype and the end-to-end-process are currently being tested with a DNV GL operated system, OREDA, in parallel with the WP8 and WP9 activities. Effects of using Optique as an enabler for integrating OREDA and other DNV GL systems and tools are ongoing. The possible positive effects on existing work processes and possible new opportunities are being discussed. Additional pilot installations are also planned and will be performed during Y3.

- Test the Optique messaging with customers

  Optique plans and preliminary results have been presented more than 60 times to external parties, including different business units of the consortium partners; the project has received a lot of positive and valuable feedback that will feed back into business plans.

Task 11.1: Market Analysis

► Task definition according to DoW

This task will find out which companies are the most likely to benefit from implementing Optique, estimate the business potential, and identify risks and characteristic barriers to adoption.

Progress

A report was prepared and delivered as part of the Y1 delivery (Deliverable D11.1 Optique Initial Exploitation Plan).

Task 11.2: Business and Implementation Plans

► Task definition according to DoW

The task will produce a generic business plan that allows companies to estimate commercial value and obtain a general picture of the viability of adopting Optique. The implementation plan will describe how an Optique system is introduced.

Progress

Work with the Market Analysis and implementation plans have been accelerated. More effort will be given to these tasks during 2015 when Optique results are materialized, can be demonstrated and evaluated by Partner Program participants.

Year 2 activities focus on building a partner network, engaging industrial users for feedback, advice, and exploration of implementation potential. The partner network will be used actively in recruiting users for training and business plan workshops. The work with the generic business plan is ongoing and will be accelerated when input form partners can be incorporated. Results obtained so far are documented in the D11.2 report.

A long-list of Optique partner programme candidates has been established, mainly comprising public and private enterprises within Norway and Sweden representing a variety of branches. Several presentations of Optique to key marketing personnel in DNV GL have led to number of requests for information and presentations. An article in DNV GL's autumn-2014 issue of customer magazine Perspectives presents Optique as ground breaking technology enabling engineers to "work smarter". The magazine is distributed to 30.000 DNV GL customers worldwide during November-December 2014 will probably result in requests for more information about Optique and be a good test of the Optique messaging.

To obtain a detailed understanding of the market and validate the business proposition, fluid Operations (FOP) has included Optique in its sales and marketing activities already at an early stage. The primary objectives, principles and decisive advantages of the Optique project are a central part of many presentations with prospective customers and with almost every presentation with multipliers such as reselling partners of FOP.

The partner program build-up preparations have started primarily involving private and public enterprises in Norway and Sweden. Year 3 becomes an important year since the first pilot installations will be performed, specific training and introductory documentation will be required and also support and assistance to get both the pilots and test sites up and running.

During the reporting period, the project has seen considerable advances in the use cases explored in WP8 and WP9. Looking to the wider market, Optique finds increased interest from industrial and governmental user communities. Recruitment of partners to the program is pursued under the Optique Partner program, with a Norwegian partner network in active development as a pilot, building the Optique ecosystem of vendors and users.

Specific market activities performed by FOP and DNV GL:

FOP

FOP can build on established relations with customers where trusted advisory status are held and have fostered an extensive partner network with resellers, system integrators and managed service providers across all of Germany, where the automotive industry is traditionally very strong. Hence it is not particularly surprising that many of those appointments had been conducted in this very industry and its suppliers. However, we managed to position the Optique solution framework across virtually all verticals and FOP observed notable interest with telecommunication companies, service providers and financial institutions such as banks, insurances and clearing agencies.

DNV GL

Several presentations to DNV GL key personnel within the Maritime, Oil & Gas and Energy areas have been given. This has led to several requests for additional detailed information and to presentations to customers. In addition DNV GL has identified potential own take-up projects. So far use of Optique with the Oreda system within Oil & Gas and with the Compass project within DNV GL maritime division is being evaluated. Oreda is a system used for reliability analysis of data while the Compass project is a research project focusing on condition monitoring of ships. Within the DNV GL energy division there are contacts with power supply enterprises that have shown interest for using Optique for applications within management of complicated power grid systems.

It is also expected that the Optique article published in the DNV GL customer magazine will attract additional companies and persons with an interest in Optique.

Generic Business Plan status

A separate report has been delivered (Deliverable D11.2). The report outlines:

- A first identification of stakeholders (stakeholder analysis)
- A preliminary gap analysis, discussing Optique versus competing solutions in the Big Data landscape
- A first value proposition, discussing game-changing benefits that can be achieved with implementing Optique

- A high-level discussion of barriers to adoption

Further details are found in report D11.2 Exploitation report: Market Analysis- Opportunities, Barriers and Risks.

The intention is to further develop this report into a more comprehensive Business Plan and Implementation guide during 2015 and to deliver a more comprehensive D11.3 report in month 36. The project has found it beneficiary to move some of the project resources from D11.2 to D11.3 since this is a more resource demanding activity.

Task 11.3: Commercialisation of Optique Software

► Task definition according to DoW

In the task we will plan and manage the activities needed to develop the R&D results from prototypes into marketable products.

Progress

The current Optique system offers a full set of OBDA components, with database connections, mapping management and a graphical query interface. That means, substantial parts of the Optique end-to-end integration are already available today. During year 2, robustness of existing features has been pursued so that a reliable core of functionality can be offered on commercial terms, while new functionality is being incorporated continuously.

A key enabler for the ability to make results available commercially as early as possible is the fact that the integration of components is realized as extensions to an established, commercially proven software platform, the fluidOps Information Workbench. The integration and productization of Optique can thus be done in increments, rather than as a downstream activity after the development of all components has been completed.

For the components that have and will be integrated, proper license agreements have been defined, which on the one hand are aligned with the commercialization plans within the Optique consortium and on the other hand also support adoption and exploitation in wider contexts. For example, Ontop has been made available under Apache licence, which allows commercialising as part of the Optique platform (with additional value adds) as well as usage as standalone component.

- A first demonstration version of the Optique software ("The Alpha package") will be distributed free of charge as part of the partner program. The use of this demo version will be limited and not intended for commercial use.
- Participation in the Partner program is free of charge but there will be a price for activities that require mobilization of Optique specialists to perfume specific activities for a partner (e.g. training, consultancy, support). This price tag is agreed on a case to case basis between the seller and receiver of the service.
- A suitable set of commercial and open source licenses, for Optique models as well as the platform as a whole is in place.

Task 11.4: Standards

► Task definition according to DoW

This task will promote active contributions to developing and advancing standards in the area of Optique. We will closely monitor ongoing standardisation activities within standardization bodies, and contribute to these activities with the results of the project.

During Year 2, development of the Optique system maintains its commitment to standards. This includes, for the software system, existing standards for databases, ontologies, and mappings; and for the use cases, Energy and O&G industry standards. Optique is committed to submitting relevant innovations as proposed extensions to existing standards.

1. ICT standards: As a generic data access platform, Optique is being developed for full compliance with data formats and protocols.
2. Industry standards: the use case work packages apply ontology representations of industry standards pertinent to the problem domain.

For the Energy domain,

- The Siemens use case (WP8) applies standards DIN EN 13306, IEC 62682, ISO 13372, ISO 13379, ISO 17359, and ISO 3977-9.
- Contact has been initiated with the CIM Users Group, targeting power transmission and distribution. A demonstration of Optique as front-end to CIM data is being planned.

For the O&G domain,

- The Statoil use case (WP9) applies: The International Stratigraphic Chart (ISC) ontology, the NORLEX vocabulary, and the Concise Geologic Time Scale for stratigraphy, in addition to ontologies and standards BFO and GeoSPARQL.QUDT, RDF Data Cube Vocabulary, and others have been identified as relevant. Additionally, Norwegian national de-facto database Petrobank and NPD have been demonstrated extensively with the Optique platform.

General project management standards,

- Implementation plans will be based on recommendations from Project Management Institute (PMI)
- Risk management is performed according to ISO-31000

4 Deliverables and Milestones Tables

Deliverables (excluding the periodic and final reports)

Del. no.	Deliverable name	WP no.	Lead beneficiary	Nature	Dissemination level	Delivered	Actual/ Forecast delivery date	Comments
D10.1	Project Fact Sheet	10	1	O	PU	yes	1/1	
D10.2	Press Release	10	1	O	PU			
D10.3	Project Web site	10	1	O	PU	yes	2/1	
D2.1	Specification of the Architecture	2	2	R	PU	yes	6/6	
D12.1	Quality Assurance and Risk Management Plan	12	1	R	CO	no	7/6	Delayed until after M6 project plenary meeting
D1.1	Joint Phase 1 Evaluation and Phase 2 Requirement Analysis	1	1	R	PU	yes	13/12	
D2.3	First Prototype of the Core Platform (incl. documentation)	2	7	R	PU	yes	12/12	
D3.1	Year 1 WP3 progress report	3	2	R	PU	yes	12/12	
D4.1	WP4 Year 1 progress report	4	4	R	PU	yes	12/12	
D5.1	A semantics for temporal and stream-based query answering in an OBDA context	5	2	R	PU	yes	12/12	
D6.1	WP6 Year 1 progress report	6	5	R	PU	yes	12/12	
D7.1	Techniques for distributed query planning and execution: one-time queries	7	6	R	PU	yes	12/12	
D8.1	Interim Siemens use case report 1	8	8	R	CO	yes	13/12	
D9.1	Interim Statoil use case report 1	9	10	R	CO	yes	13/12	

D10.4	Year 1 Dissemination Report	10	1	R	PU	yes	13/12	
D11.1	Exploitation Report: Initial Exploitation Plan	11	10	R	PU	yes	13/12	
D12.2	Intermediate project report, year 1	12	1	R	CO	yes	13/12	
D1.2	Requirement Analysis and Evaluation Framework	1	6	R	PU	yes	15/15	

Table 4.1: Deliverables until end of reporting period.

All Y1.5 deliverables are summarised in Table 4.1. All scientific deliverables and the first prototype of the core platform were delivered on time. A preliminary version of the project web site was created for the project kick-off meeting; the full web site was in place in M02. D12.1 (Quality Assurance and Risk Management Plan) was finalised after the M07 plenary meeting. The deadlines for D1.1, D8.1 and D9.1 were extended in order to have adequate time after the end user evaluation workshops. The deadlines for D10.4, D11.1 and D12.2 were extended in order to report on the full first-year project period. All deliverables were ready before the Y1 project review.

Milestones

Milestone no.	Milestone name	WP no.	Lead beneficiary	Achieved (yes/no)	Actual/ Forecast achievement date	Comments
MS1	Online Collaboration Platform	12	1	yes	3/1	Delay due to necessary hardware procurement
MS2	Ph. 1: Joint requirements analysis finished	1	1	yes	1/1	First-year requirements as described in DoW
MS3	Ph. 1: Development environment in place	2	7	yes	3/3	
MS4	Ph. 1: Requirement analysis framework finished	1	6	yes	3/3	Draft version of D1.2 available
MS5	Ph. 1: Architecture aligned with use case partners	2, 8, 9	7	yes	3/3	
MS6	Ph. 1: Components ready for integration	2, 4, 6, 7	7	yes	8/8	
MS7	Ph. 1: System, ontologies and mapping ready for installation	2, 8, 9	7	yes	10/10	

MS8	Ph. 1: Use case testings and evaluation finished	8, 9	8	yes	12/11	Due to organisational constraints, the annual Siemens end user workshops will take place in October
MS9	Ph. 1: Joint evaluation finished	1, 2	1	yes	12/11	Delay due to delayed MS8
MS10	Ph. 2: Joint requirements analysis finished	1, 8, 9	1	yes	13/11	Delay due to delayed MS8
MS11	Ph. 2: Architecture aligned with use case partners	2, 8, 9	7	yes	14/14	

Table 4.2: Milestones until end of reporting period.