Project N°: **FP7-318338**

Project Acronym: **Optique**

Project Title: **Scalable End-user Access to Big Data**

Instrument: **Integrated Project**

Scheme: **Information & Communication Technologies**

Deliverable D10.5
Year 2 Dissemination Report

Due date of deliverable: (T0+24)

Actual submission date: November 15, 2014

Start date of the project: **1st November 2012** Duration: **48 months**

Lead contractor for this deliverable: **UiO**

Dissemination level: **PU** – Public

Final version

Executive Summary:

Year 2 Dissemination Report

This document summarises deliverable D10.5 of project FP7-318338 (Optique), an Integrated Project supported by the 7th Framework Programme of the EC. Full information on this project, including the contents of this deliverable, is available online at `http://www.optique-project.eu/`.

Work package 10 on dissemination has three objectives:

O10.1 Visbility to the research community

O10.2 Visibility to targeted industry

O10.3 Visibility to the general public

This report summarises the dissemination activities conducted during the second year of the Optique project.

With respect to the research community, the project has continued to publish widely and to give presentations at relevant international conferences and workshops. With respect to targeted industry, the project has made significant advances; in particular the visibility of Optique within the three industry partners has greatly increased. We have also set up the Optique Partner Programme with two major events scheduled for 2015, and we chaired the European Data Forum 2014. With respect to the general public, we have enhanced the web site, produced YouTube videos, released new versions of the public showcase, and released open source code.

List of Authors

Vidar N. Klungre (UiO)
Tore Hartvigsen (DNV)
Arild Waaler (UiO)
Martin G. Skjæveland (UiO)
Ahmet Soylu (UiO)

Contents

Chapter 1

Introduction

Optique dissemination targets three distinct groups reflected in the three objectives of Work Package 10:

O10.1 Be visible to the research community

O10.2 Be visible to targeted industry

O10.3 Be visible to the general public

As the project progresses, the Optique platform takes shape and the project gains experience from the use cases. At the same time the opportunities for making impact become clearer. For this reason the Optique project continually revises its dissemination strategy, so as to be able to concentrate its dissemination efforts as efficiently as possible. 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.

Below we have summarised the dissemination strategy as of end of Year 2. In Section 1.1 we review the main target groups for Optique dissemination and summarise the impact goals relative to each of them. In Year 2 the Optique project is about to leave the early phase; the specific priorities for the dissemination eforts that have been taken in Year 2 are summarised in Section 1.2. In Section 1.3 we summarize how the strategy will be adjusted once the project enters a more mature phase.

1.1 End of project impact goals for key target groups

Objective O10.1 Optique aims to impact the future research agenda of the broad research community working on Big Data access by triggering researchers to explore Optique results in their own projects. For this to happen, Optique should become *widely known* within this research community, as well as becoming known *in depth* by key researchers.

Objective O10.2 The Optique partners from industry (i.e. DNVGL, Siemens and Statoil) are especially important as targets for dissemination. By virtue of their being Optique partners, we know that they have units that actively participate in Optique. In these units they build up knowledge about Optique and experience from deploying the platform; Siemens and Statoil also have installations of Optique at their premises that run against their own data and that address data access challenges identified by their own experts. We also know that these units have a direct interest in the problems that Optique addresses and the solution that the project provides, since it was this interest that brought them into the project in the first place.

Although this gives Optique a very strong starting point, it is by no means preordained that Optique will achieve high visibility in these companies. DNVGL, Siemens and Statoil are large enterprises with many internal projects and initiatives competing for attention. Also, business strategies for these companies are constantly subject to revision and will in all likelihood change in the course of the 4 year project period of Optique.

To create lasting impact in DNVGL, Siemens and Statoil, Optique will exploit our broad access to internal dissemination channels in these companies in order to target three different kinds of stakeholders: the management level, the corporate technology units, and the end user business units.

The impact goal for the corporate management level is that Optique becomes part of the strategic discussion related to future technology investments. In order to achieve this, it is important that the Optique project uses every opportunity to present itself to managers in key positions. But this is in itself not enough; Optique needs to be recommended internally both by corporate technology units and some end user business units. The impact goal for these units is precisely to reach the point where they support the project to the degree that they actively point to Optique solutions internally and also recommend Optique to other units in the companies.

For enterprises other than the three Optique industry partners, the Optique strategy is to channel dissemination activities through the Optique Partner Programme. The impact goal for Partner Programme members is that they set up Optique pilots and build up sufficient internal expertise to successfully use such pilots.

Objective O10.3 With respect to the general public (Objective O10.3), the goal is to make Optique widely known in communities with an interest in big data in general, and data access technologies in particular. This includes all kinds of stakeholders, including stakeholders at the political level. For the Optique software components released as open source, the goal is to foster active communities around the open source projects that include developers as well as users.

1.2 Priorities in Year 2

The Optique project started in November 2012, with the starting point being several existing software components—especially the Information Workbench of FOP (WP2), the Ontop project of FUB (WP6), the ADP project of UoA (WP7)—and a number of innovative new concepts and ideas for research and development. In Year 1 of the Optique project the consortium was thus able to disseminate a prototype implementation and elaborations of key concepts using appropriate dissemination measures for each of the key target groups.

At the start of Year 2, the Optique project was still at an early phase. In particular, the new and novel components resulting from research and development in Year 1 existed only as early prototypes, with little testing or exposure to real users. This situation has gradually changed during the course of Year 2, with the development of a more robust integrated platform, and the conclusion of successful end user workshops. Thus, by the start of Year 3, the project will be ready to leave the early phase of dissemination.

Objective O10.1 In order to impact the research community Optique has used the opportunities to present early stage research results offered by workshops and conferences. Such events may have relatively low industry impact, but they are nevertheless important in order to make researchers working in the same areas aware of the project. These researchers are especially important because they are the ones who are most likely to take up ideas from Optique and exploit them further in new directions, and in this way create even stronger scientific impact from Optique.

In Year 2, as in Year 1, Optique has disseminated research results to the scientific community mainly through participation in conferences and workshops. Work on publications with higher expected impact has already started, and will continue with high priority throughout the remainder of the project.

Objective O10.2 Year 2 has seen significant advances in the maturity of the use case installations at Siemens and Statoil. It has been a main priority of the Optique project to demonstrate these advances internally to corporate technology units, business units and management level staff. It has been too early to set up proper user training events in Siemens and Statoil, but training of evaluators has been given and preparation of professional training material has began. The annual end user evaluations and user requirements events have also been important as dissemination events within Siemens and Statoil.

At DNV GL, the project has actively promoted the Optique platform internally. As a consequence of this, two internal DNV GL projects are currently planning to install and test the Optique platform by setting up their own pilots.

In Year 2 fluid Operations have actively promoted the Optique platform to a large number of customers and companies in their own network. In line with this DNV GL has established a community of interested stakeholders in Sweden and Norway, organising this within the Optique Partner Programme. The Optique project has been actively recruiting new partner members and is planning major Partner Programme events in Year 3. Production of White Papers and training material has also begun.

The Optique Advisory Board has also been a central target for dissemination to industry, with a full day meeting with high level representatives from Halliburton Landmark, IBM, Oracle and Schlumberger Information Solutions.

Objective O10.3 In the early states of the project, Optique has planned to reach the general public through the web site, through releases of open source code, and through developing and actively demonstrating the Optique public showcase.

Dissemination towards the general public has not been a main priority in Year 2. This is natural since the project in Year 2 is at a transition stage between "flashlight" from the project kick-off and the first "breaking news" results. For this reason there has been no new press releases in Year 2, but the project is constantly considering when this strategy should change. In Year 2 the project has prioritised video production for YouTube, release of open source code of Ontop, and improvements of the Public showcase. We foresee a much more aggressive strategy towards public dissemination in the later phase of the project.

1.3 Dissemination measures in the later phases of the project

When a mature implementation of the Optique platform is in place and we can point to specific achievements in the use case installations, the project is positioned for high impact activities. The dissemination strategy will hence be adjusted so as to focus on opportunities for creating impact that goes far beyond the project period.

Objective O10.1 Dissemination to the research community will focus more on publications in high impact journals, and publications, tutorials and workshops at major conferences.

Objective O10.2 In the industry partners DNVGL, Siemens and Statoil, Optique will focus on training activities so as to give professionals in these companies deep knowledge about the Optique platform, as well as experience with using it. The project will also work proactively towards setting up pilots of the Optique platform in other units in these companies than those that are actively participating in the Optique project.

All kinds of training and dissemination material will be distributed to members of the Optique Partner Programme, and the project will seek every opportunity to assist its members in setting up pilots on their own data and to share their experiences.

The Optique Partner Programme will be organised in several nodes, where each node is managed by an Optique project partner. In particular DNV GL will lead the Norwegian based node, and fluid Operations will set up and lead a node in Germany.

Objective O10.3 In the later phases of the project we expect that a number of opportunities will arise for promoting project results to the general public. We will proactively work for visibility in the public space, and the web page will be used more actively as a way of promoting project results. The project plans to extend the release of open source code to parts of the Optique platform.

Chapter 2

Visibility to the research community (O10.1)

Optique has been very active in Year 2 in disseminating its results to the research community. In line with the priorities described in Sect. 1.2 we have been present at a number of conferences and published in refereed conference proceedings for a number of events, summarized in Table 2.2. Besides giving oral presentations at these events, we have given several invited keynotes and stand-alone presentations to the community.

A special priority of Work Package 10 has been to position the Optique project in the big data arena. The first serious step in this direction has been to submit an article to a call from the IEEE Computer magazine for a special issue on "Big data: management and application". The article, now accepted for publication, is included in Appendix A in this report.

2.1 Key performance indicators

Table 2.1: 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

2.2 Articles in Journals

[1] Diego Calvanese, Thomas Eiter, and Magdalena Ortiz. "Answering Regular Path Queries in Expressive Description Logics via Alternating Tree-Automata". In: *Information and Computation* 237 (2014), pp. 12–55. ISSN: 0890-5401. DOI: 10.1016/j.ic.2014.04.002.

[2] Giuseppe De Giacomo, Riccardo De Masellis, Marco Grasso, Fabrizio Maria Maggi, and Montali Marco. "Monitoring Business Metaconstraints Based on LTL & LDL for Finite Traces." In: *Journal on Data Semantics* (2014), pp. 1–25.

2.3 Refereed Conference Papers

[1] Shqiponja Ahmetaj, Diego Calvanese, Magdalena Ortiz, and Mantas Simkus. "Managing Change in Graph-structured Data Using Description Logics". In: *Proc. of the 28th AAAI Conf. on Artificial*

Table 2.2: Summary of the contributions to major venues in 2014.

Venue	Papers	Presentations
ISWC[a]	5	6
DL workshop[b]	2	3
WWW[c]	2	2
OWLED[d]	1	2
KR[e]	0	2
ECAI[f]	0	1

[a] International Semantic Web Conference 2014
[b] Description Logic workshop 2014
[c] International World Wide Web Conference
[d] OWL: Experiences and Directions Workshop
[e] International Conference on Principles of Knowledge Representation and Reasoning 2014
[f] European Conference on Artificial Intelligence 2014

Intelligence (AAAI 2014). AAAI Press, 2014, pp. 966–973. ISBN: 978-1-57735-661-5. URL: http://www.aaai.org/ocs/index.php/AAAI/AAAI14/paper/view/8238.

[2] Albin Ahmeti, Diego Calvanese, and Axel Polleres. "SPARQL Update for Materialized Triple Stores under DL-Lite$_{RDFS}$ Entailment". In: *Proc. of the 27th Int. Workshop on Description Logics (DL 2014)*. Vol. 1193. CEUR Electronic Workshop Proceedings, http://ceur-ws.org/. 2014, pp. 8–20. URL: http://ceur-ws.org/Vol-1193/paper_7.pdf.

[3] Albin Ahmeti, Diego Calvanese, and Axel Polleres. "Updating RDFS ABoxes and TBoxes in SPARQL". In: *Proc. of the 13th Int. Semantic Web Conf. (ISWC 2014)*. Vol. 8796. Lecture Notes in Computer Science. Springer, 2014, pp. 441–456. ISBN: 978-3-319-11963-2. DOI: 10.1007/978-3-319-11964-9_28. URL: http://link.springer.com/chapter/10.1007/978-3-319-11964-9_28.

[4] Marcelo Arenas, Bernardo Cuenca Grau, Evgeny Kharlamov, Sarunas Marciuska, and Dmitriy Zheleznyakov. "Enabling Faceted Search over OWL 2 with SemFacet". In: *11th OWL: Experiences and Directions Workshop*. 2014.

[5] Marcelo Arenas, Bernardo Cuenca Grau, Evgeny Kharlamov, Sarunas Marciuska, and Dmitriy Zheleznyakov. "Towards Semantic Faceted Search". In: *23rd International World Wide Web Conference, WWW '14, Seoul, Republic of Korea, April 7-11, 2014, Companion Volume*. 2014, pp. 219–220. DOI: 10.1145/2567948.2577381. URL: http://doi.acm.org/10.1145/2567948.2577381.

[6] Marcelo Arenas, Bernardo Cuenca Grau, Evgeny Kharlamov, Sarunas Marciuska, Dmitriy Zheleznyakov, and Ernesto Jiménez-Ruiz. "SemFacet: semantic faceted search over yago". In: *23rd International World Wide Web Conference, WWW '14, Seoul, Republic of Korea, April 7-11, 2014, Companion Volume*. 2014, pp. 123–126. DOI: 10.1145/2567948.2577011. URL: http://doi.acm.org/10.1145/2567948.2577011.

[7] Marcelo Arenas, Bernardo Cuenca Grau, Evgeny Kharlamov, Šarūnas Marciuška, and Dmitriy Zheleznyakov. "Faceted Search over Ontology-Enhanced RDF Data". In: *Proc. ACM International Conference on Information and Knowledge Management (CIKM)*. Acceptance rate: 21%. 2014.

[8] Timea Bagosi, Diego Calvanese, Josef Hardi, Sarah Komla-Ebri, Davide Lanti, Martin Rezk, Mariano Rodriguez-Muro, Mindaugas Slusnys, and Guohui Xiao. "The Ontop Framework for Ontology Based Data Access". In: *Proc. of the 8th Chinese Semantic Web Symposium & Web Science Conference (Posters and Demos)*. Wuhan, China, 2014. URL: http://www.ghxiao.org/publications/2014-csws-ontop.pdf.

[9] Meghyn Bienvenu, Diego Calvanese, Magdalena Ortiz, and Mantas Simkus. “Nested Regular Path Queries in Description Logics”. In: *Proc. of the 14th Int. Conf. on the Principles of Knowledge Representation and Reasoning (KR 2014)*. AAAI Press, 2014. ISBN: 978-1-57735-657-8.

[10] Meghyn Bienvenu, Diego Calvanese, Magdalena Ortiz, and Mantas Simkus. “Nested Regular Path Queries in Description Logics (Extended Abstract)”. In: *Proc. of the 8th Alberto Mendelzon Int. Workshop on Foundations of Data Management (AMW 2014)*. Vol. 1189. CEUR Electronic Workshop Proceedings, http://ceur-ws.org/. 2014. URL: http://ceur-ws.org/Vol-1189/paper_34.pdf.

[11] Diego Calvanese. “Query Answering over Description Logic Ontologies”. In: *Proc. of the 14th European Conf. on Logics in Artificial Intelligence (JELIA 2014)*. Vol. 8761. Lecture Notes in Computer Science. Springer, 2014. ISBN: 978-3-319-11557-3. DOI: 10.1007/978-3-319-11558-0_1. URL: http://link.springer.com/chapter/10.1007/978-3-319-11558-0_1.

[12] Diego Calvanese, Wolfgang Fischl, Reinhard Pichler, Emanuel Sallinger, and Mantas Simkus. “Capturing Relational Schemas and Functional Dependencies in RDFS”. In: *Proc. of the 28th AAAI Conf. on Artificial Intelligence (AAAI 2014)*. AAAI Press, 2014, pp. 1003–1011. ISBN: 978-1-57735-661-5. URL: http://www.aaai.org/ocs/index.php/AAAI/AAAI14/paper/view/8487.

[13] Diego Calvanese, Wolfgang Fischl, Reinhard Pichler, Emanuel Sallinger, and Mantas Simkus. “Expressive Identification Constraints to Capture Functional Dependencies in Description Logics”. In: *Proc. of the 27th Int. Workshop on Description Logics (DL 2014)*. Vol. 1193. CEUR Electronic Workshop Proceedings, http://ceur-ws.org/. 2014, pp. 416–420. URL: http://ceur-ws.org/Vol-1193/paper_48.pdf.

[14] Diego Calvanese, Davide Lanti, Martin Rezk, Mindaugas Slusnys, and Guohui Xiao. “A Scalable Benchmark for OBDA Systems: Preliminary Report”. In: *Proc. of the 3rd Int. Workshop on OWL Reasoner Evaluation (ORE 2014)*. CEUR Workshop Proceedings. 2014.

[15] Xi Chen, Weiguo Xia, Ernesto Jimenez-Ruiz, and Valerie Cross. “Extending an ontology alignment system with BioPortal: a preliminary analysis”. In: *Poster paper at International Semantic Web Conference (ISWC)*. 2014.

[16] Cristina Civili. “Query Answering over Ontologies Specified via Database Dependencies”. In: *In Proceedings of the 2014 SIGMOD PhD symposium*. 2014, p. 36–40.

[17] Marco Console, Jose Mora, Ricardo Rosati, Valerio Santarelli, and Domenico Fabio Savo. “Effective computation of maximal sound approximations of Description Logic ontologies”. In: *The Semantic Web - ISWC 2014 - 13th International Semantic Web Conference*. 2014, pp. 164–179.

[18] Bernardo Cuenca Grau, Evgeny Kharlamov, Sarunas Marciuska, Dmitriy Zheleznyakov, and Yujiao Zhou. “Querying Life Science Ontologies with SemFacet”. In: *Semantic Web Applications and Tools for Life Sciences (SWAT4LS)*. 2014.

[19] Giuseppe De Giacomo, Riccardo De Masellis, and Montali Marco. “Reasoning on LTL on Finite Traces: Insensitivity to Infiniteness.” In: *In Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*. AAAI Press, 2014, pp. 1027–1033.

[20] Giuseppe De Giacomo, Yves Lespérance, Fabio Patrizi, and Stavros Vassos. “LTL Verification of Online Executions with Sensing in Bounded Situation Calculus”. In: *ECAI 2014 - 21st European Conference on Artificial Intelligence*. 2014, pp. 369–374.

[21] Giuseppe De Giacomo, Yves Lespérance, Fabio Patrizi, and Stavros Vassos. “Progression and verification of situation calculus agents with bounded beliefs”. In: *International conference on Autonomous Agents and Multi-Agent Systems*. 2014, pp. 141–148.

[22] Riccardo De Masellis, Domenico Lembo, Marco Montali, and Dmitry Solomakhin. “Semantic Enrichment of GSM-Based Artifact-Centric Models”. In: *Principles of Knowledge Representation and Reasoning: Proceedings of the Fourteenth International Conference, KR*. AAAI Press, 2014.

[23] Thomas Eiter, Patrik Schneider, Mantas Šimkus, and Guohui Xiao. "Using OpenStreetMap Data to Create Benchmarks for Description Logic Reasoners". In: *Proc. of the 3rd Int. Workshop on OWL Reasoner Evaluation (ORE 2014)*. CEUR Workshop Proceedings. 2014.

[24] Daniel Faria, Ernesto Jiménez-Ruiz, Catia Pesquita, Emanuel Santos, and Francisco M. Couto. "Towards Annotating Potential Incoherences in BioPortal Mappings". In: *The Semantic Web - ISWC 2014 - 13th International Semantic Web Conference, Riva del Garda, Italy, October 19-23, 2014. Proceedings, Part II*. Vol. 8797. Lecture Notes in Computer Science. Springer, 2014, pp. 17–32. DOI: 10.1007/978-3-319-11915-1_2. URL: http://dx.doi.org/10.1007/978-3-319-11915-1_2.

[25] Bernardo Cuenca Grau, Evgeny Kharlamov, Dmitriy Zheleznyakov, Marcelo Arenas, and Sarunas Marciuska. "On Faceted Search over Knowledge Bases". In: *Informal Proceedings of the 27th International Workshop on Description Logics, Vienna, Austria, July 17-20, 2014*. 2014, pp. 153–156. URL: http://ceur-ws.org/Vol-1193/paper_89.pdf.

[26] Simen Heggestøyl, Guillermo Vega-Gorgojo, and Martin Giese. "Visual Query Formulation for Linked Open Data: The Norwegian Entity Registry Case". In: ed. by Tom Heine Nätt. 2014.

[27] Ernesto Jiménez-Ruiz, Bernardo Cuenca Grau, Weiguo Xia, Alessandro Solimando, Xi Chen, Valerie Cross, Yuan Gong, Shuo Zhang, and Anu Rekha. "LogMap family results for OAEI 2014". In: *Proc. of the 9th Ontology Matching Workshop*. 2014. URL: http://www.dit.unitn.it/~p2p/OM-2014/oaei14_paper6.pdf.

[28] Evgeny Kharlamov, Nina Solomakhina, Oezgur Oezcep, Dmitriy Zheleznyakov, Thomas Hubauer, Steffen Lamparter, Mikhail Roshchin, and Ahmet Soylu. "How Semantic Technologies can Enhance Data Access at Siemens Energy". In: *Proc. International Semantic Web Conference*. 2014.

[29] Roman Kontchakov, Martin Rezk, Mariano Rodriguez-Muro, Guohui Xiao, and Michael Zakharyaschev. "Answering SPARQL Queries over Databases under OWL 2 QL Entailment Regime". In: *Proc. of International Semantic Web Conference (ISWC 2014)*. Lecture Notes in Computer Science. Springer, 2014.

[30] Davide Lanti, Martin Rezk, Mindaugas Slusnys, Guohui Xiao, and Diego Calvanese. "The NPD Benchmark for OBDA Systems". In: *Proc. 10th International Workshop on Scalable Semantic Web Knowledge Base Systems (SSWS 2014)*. CEUR-WS.org, 2014.

[31] Domenico Lembo, Jose Mora, Riccardo Rosati, Domenico Fabio Savo, and Evgenij Thorstensen. "Towards Mapping Analysis in Ontology-Based Data Access". In: *Web Reasoning and Rule Systems - 8th International Conference, RR*. 2014, pp. 108–123.

[32] Domenico Lembo, Riccardo Rosati, Marco Ruzzi, Domenico Fabio Savo, and Emanuele Tocci. "Visualization and Management of Mappings in Ontology-based Data Access (Progress Report)." In: *Informal Proceedings of the 27th International Workshop on Description Logics*. 2014, pp. 595–607.

[33] Maurizio Lenzerini and Marco Console. "Data Quality in Ontology-based Data Access: The Case of Consistency". In: *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*. 2014, pp. 1020–1026.

[34] Maurizio Lenzerini and Marco Console. "Reducing global consistency to local consistency in Ontology-based Data Access." In: *ECAI 2014 - 21st European Conference on Artificial Intelligence*. 2014, pp. 219–224.

[35] Maurizio Lenzerini and Marco Console. "Reducing Global Consistency to Local Consistency in Ontology-based Data Access - Extended Abstract." In: *Informal Proceedings of the 27th International Workshop on Description Logics*. 2014, pp. 496–499.

[36] Maurizio Lenzerini, Lorenzo Lepore, and Antonella Poggi. "Practical Query Answering over HI(DL-LiteR) Knowledge Bases". In: *Informal Proceedings of the 27th International Workshop on Description Logics*. 2014, pp. 608–619.

[37] Marco Montali, Diego Calvanese, and Giuseppe De Giacomo. "Verification of Data-Aware Commitment-Based Multiagent System". In: *In Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*. 2014, pp. 157–164.

[38] José Mora, Riccardo Rosati, and Óscar Corcho. "kyrie2: Query Rewriting under Extensional Constraints in ELHIO". In: *The Semantic Web - ISWC 2014 - 13th International Semantic Web Conference*. 2014, pp. 568–583.

[39] Ö. L. Özçep and R. Möller. "Ontology Based Data Access on Temporal and Streaming Data". In: *Reasoning Web. Reasoning and the Web in the Big Data Era*. Ed. by M. Koubarakis, G. Stamou, G. Stoilos, I. Horrocks, P. Kolaitis, G. Lausen, and G. Weikum. Vol. 8714. Lecture Notes in Computer Science. 2014.

[40] Özgür L. Özçep, Ralf Möller, and Christian Neuenstadt. "A Stream-Temporal Query Language for Ontology Based Data Access". In: *KI 2014*. Ed. by Carsten Lutz and Michael Thielscher. Vol. 8736. LNCS. Springer International Publishing Switzerland, 2014, pp. 183–194.

[41] Özgür L. Özçep, Ralf Möller, and Christian Neuenstadt. "A Stream-Temporal Query Language for Ontology Based Data Access". In: *Proceedings of the 7th International Workshop on Description Logics (DL-2014)*. Ed. by Meghyn Bienvenu, Magdalena Ortiz, Riccardo Rosati, and Mantas Simkus. 2014.

[42] Fabio Patrizi and Stavros Vassos. "Action Theories over Generalized Databases with Equality Constraints". In: *Logics in Artificial Intelligence - 14th European Conference, JELIA*. 2014, pp. 472–485.

[43] Christoph Pinkel, Carsten Binnig, Peter Haase, Clemens Martin, Kunal Sengupta, and Johannes Trame. "How to Best Find a Partner? An Evaluation of Editing Approaches to Construct R2RML Mapping". In: *ESWC*. 2014.

[44] Alessandro Solimando, Ernesto Jiménez-Ruiz, and Giovanna Guerrini. "Detecting and Correcting Conservativity Principle Violations in Ontology-to-Ontology Mappings". In: *The Semantic Web - ISWC 2014 - 13th International Semantic Web Conference, Riva del Garda, Italy, October 19-23, 2014. Proceedings, Part II*. 2014, pp. 1–16. DOI: `10.1007/978-3-319-11915-1_1`. URL: `http://dx.doi.org/10.1007/978-3-319-11915-1_1`.

[45] Alessandro Solimando, Ernesto Jimenez-Ruiz, and Christoph Pinkel. "Evaluating Ontology Alignment Systems in Query Answering Tasks". In: *Poster paper at International Semantic Web Conference (ISWC)*. 2014. URL: `/brokenurl#PosterpaperatInternationalSemanticWebConference(ISWC)`.

[46] Alessandro Solimando, Ernesto Jiménez-Ruiz, and Giovanna Guerrini. "A Multi-strategy Approach for Detecting and Correcting Conservativity Principle Violations in Ontology Alignments". In: *Proceedings of the 11th International Workshop on OWL: Experiences and Directions (OWLED 2014)*. 2014, pp. 13–24. URL: `http://ceur-ws.org/Vol-1265/owled2014_submission_2.pdf`.

[47] Ahmet Soylu, Martin Giese, Ernesto Jimenez-Ruiz, Evgeny Kharlamov, Dmitriy Zheleznyakov, and Ian Horrocks. "Towards Exploiting Query History for Adaptive Ontology-based Visual Query Formulation". In: *Proceedings of 8th Metadata and Semantics Research Conference (MTSR 2014)*. Ed. by Sissi Closs, Rudi Studer, Emmanouel Garoufallou, and Miguel-Angel Sicilia. Vol. 478. CCIS. Springer, 2014, pp. 107–119.

[48] Ahmet Soylu, Martin Giese, Ernesto Jimenez-Ruiz, Evgeny Kharlamov, Dmitriy Zheleznyakov, and Ian Horrocks. "Why not simply Google?" In: *Proceedings of the 8th Nordic Conference on Human-Computer Interaction (NordiCHI 2014)*. ACM, 2014, pp. 1039–1042.

[49] Ahmet Soylu, Evgeny Kharlamov, Dmitriy Zheleznyakov, Ernesto Jimenez-Ruiz, Martin Giese, and Ian Horrocks. "OptiqueVQS: Visual Query Formulation for OBDA". In: *Proceedings of the 27th International Workshop on Description Logics (DL 2014)*. Vol. 1193. CEUR-WS.org, 2014, pp. 725–728.

[50] Guohui Xiao, Martin Rezk, Mariano Rodriguez-Muro, and Diego Calvanese. "Rules and Ontology Based Data Access". In: *Proc. 8th International Conference on Web Reasoning and Rule Systems (RR 2014)*. Ed. by Marie-Laure Mugnier and Roman Kontchakov. Lecture Notes in Computer Science. Springer, 2014.

[51] Martin G. Skjæveland and Espen H. Lian. "Benefits of Publishing the Norwegian Petroleum Directorate's FactPages as Linked Open Data". In: *Norsk informatikkonferanse (NIK 2013)*. Tapir, 2013.

[52] Fabio Patrizi and Stavros Vassos. In: *Principles of Knowledge Representation and Reasoning: Proceedings of the Fourteenth International Conference, KR.*

2.4 Keynotes at Scientific Events

Description/Title	Presenter	Venue	Date
Invited talk at AIIA 2014: Ontology and mapping management	Maurizio Lenzerini	Turin, Italy	2013-12-05
Invited talk at CILC 2014: Tractable Approaches to Consistent Query Answering in Ontology-Based-Data Access	Riccardo Rosati	Turin, Italy	2014-06-17
Invited talk at RuleML 2014: Efficient mapping rules in OBDA	Arild Waaler	Prague, Czech Republic	2014-08-20
Keynote at AIMSA 2014: Scalable End-User Access to Big Data	Diego Calvanese	Varna, Bulgaria	2014-09-11
Keynote at JELIA 2014: Query Answering over Description Logic Ontologies	Diego Calvanese	Madeira, Portugal	2014-09-24

2.5 Presentations at Conferences and Workshops

Description/Title	Presenter	Venue	Date
Invited Talk at 11th Open Excellence Workshop in RELATE : Time/Money Tradeoffs in Big Data Processing on the Cloud	Herald Kllapi	Athens, Greece	2014-03-21
EU Projects Track at EDBT/ICDT 2014: Optique Project	Yannis Ioannidis	Athens, Greece	2014-03-27
Poster Presentation at WWW'14: Towards semantic faceted search	Evgeny Kharlamov	Seoul, Republic of Korea	2014-04-10
Demonstration at WWW'14: SemFacet: Semantic Faceted Search over Yago	Evgeny Kharlamov	Seoul, Republic of Korea	2014-04-10
Oral Presentation at AAMAS 2014: Action Theories over Generalized Databases with Equality Constraints	Yves Lespérance	Paris, France	2014-05-07
Oral Presentation at AMW 2014: Nested Regular Path Queries in Description Logics (Extended Abstract)	Magdalena Ortiz	Cartagena, Colombia	2014-06-05
Oral Presentation at DL 2014: SPARQL Update for Materialized Triple Stores under DL-Lite$_{RDFS}$ Entailment	Magdalena Ortiz	Cartagena, Colombia	2014-06-05

continued ...

Description/Title	Presenter	Venue	Date
Oral Presentation at ORE 2014: A Scalable Benchmark for OBDA Systems: Preliminary Report	Davide Lanti	Vienna, Austria	2014-07-13
Oral Presentation at ORE 2014: Using OpenStreetMap Data to Create Benchmarks for Description Logic Reasoners	Guohui Xiao	Vienna, Austria	2014-07-13
Oral Presentation at DL 2014: Practical Query Answering over HI(DL-LiteR) Knoledge Bases	Lorenzo Lepore	Vienna, Austria	2014-07-17
Poster presentation at DL 2014: A Stream-Temporal Query Language for Ontology Based Data Access	Özgür L. Özçep and Christian Neuenstadt	Vienna, Austria	2014-07-18
Oral Presentation at DL 2014: Expressive Identification Constraints to Capture Functional Dependencies in Description Logics	Wolfgang Fischl	Vienna, Austria	2014-07-19
Tutorial at at International Conference on Principles of Knowledge Representation and Reasoning (KR 2014): Query Answering and Rewriting in Ontology-Based Data Access	Riccardo Rosati	Vienna, Austria	2014-07-20
Oral Presentation at KR 2014: Nested Regular Path Queries in Description Logics	Meghyn Bienvenu	Vienna, Austria	2014-07-21
Oral Presentation at KR 2014: Action Theories over Generalized Databases with Equality Constraints (Extended Abstract)	Stavros Vassos	Vienna, Austria	2014-07-22
Presentation at HDMS 2014: Time is Money but How Much? Optimal Tradeoffs in Dataflow Processing on the Cloud	Herald Kllapi	Athens, Greece	2014-07-24
Presentation at HDMS 2014: TRIREME: Sailing through Flows of Big Data	Lefteris Stamatogiannakis	Athens, Greece	2014-07-24
Oral Presentation at AAAI 2014: Managing Change in Graph-structured Data Using Description Logics	Shqiponja Ahmetaj	Quebec City, Quebec, Canada	2014-07-30
Oral Presentation at AAAI 2014: Capturing Relational Schemas and Functional Dependencies in RDFS	Wolfgang Fischl	Quebec City, Quebec, Canada	2014-07-30
Oral Presentation at ECAI 2014: Action Theories over Generalized Databases with Equality Constraints	Yves Lespérance	Madeira, Portugal	2014-08-22
Tutorial at RW 2014: Ontology Based Data Access on Temporal and Streaming Data	Özgür L. Özçep	Athens, Greece	2014-09-11
Oral Presentation at RR 2014: Rules and Ontology Based Data Access	Guohui Xiao	Athens, Grece	2014-09-15
Oral Presentation at RR 2014: Towards Mapping Analysis in Ontology-based Data Access	Jose Mora	Athens, Grece	2014-09-15
Oral Presentation at JELIA 2014: Action Theories over Generalized Databases with Equality Constraints	Fabio Patrizi	Madeira, Portugal	2014-09-26
Oral presentation at KI 2014: A Stream-Temporal Query Language for Ontology Based Data Access	Christian Neuenstadt	Stuttgart, Germany	2014-09-26

continued ...

Description/Title	Presenter	Venue	Date
Demo Presentation at OWLED'14: Enabling Faceted Search over OWL 2 with SemFacet	Sarunas Marciuska	Riva del Garda, Italy	2014-10-19
Oral Presentation at OWLED'14: Enabling Faceted Search over OWL 2 with SemFacet	Evgeny Kharlamov	Riva del Garda, Italy	2014-10-19
Oral Presentation at SSWS 2014: The NPD Benchmark for OBDA Systems	Davide Lanti	Riva del Garda, Italy	2014-10-20
Oral Presentation at OM'14: Ontology Alignment Evaluation (OAEI) 2014 Results	Ernesto Jimenez-Ruiz	Riva del Garda, Italy	2014-10-20
Oral Presentation at ISWC 2014: Effective computation of maximal sound approximations of Description Logic ontologies	Valerio Santarelli	Riva del Garda, Italy	2014-10-21
Poster Presentation at ISWC'14: Extending an ontology alignment system with BioPortal: a preliminary analysis	Valerie Cross and Ernesto Jimenez-Ruiz	Riva del Garda, Italy	2014-10-21
Poster Presentation at ISWC'14: Evaluating Ontology Alignment Systems in Query Answering Tasks	Alessandro Solimando and Ernesto Jimenez-Ruiz	Riva del Garda, Italy	2014-10-21
Oral Presentation at ISWC 2014: Answering SPARQL Queries over Databases under OWL 2 QL Entailment Regime	Roman Kontchakov	Riva del Garda, Italy	2014-10-23
Oral Presentation at ISWC 2014: Updating RDFS ABoxes and TBoxes in SPARQL	Albin Ahmeti	Riva del Garda, Italy	2014-10-23
Oral Presentation at ISWC'14: Detecting and Correcting Conservativity Principle Violations in Ontology-to-Ontology Mappings	Alessandro Solimando and Ernesto Jimenez-Ruiz	Riva del Garda, Italy	2014-10-23
Oral Presentation at ISWC'14: Towards annotating potential incoherences in BioPortal mappings	Daniel Faria and Ernesto Jimenez-Ruiz	Riva del Garda, Italy	2014-10-23
Oral Presentation at ISWC'14: How Semantic Technologies can Enhance Data Access at Siemens Energy	Evgeny Kharlamov and Nina Solomakhina	Riva del Garda, Italy	2014-10-23
Oral Presentation at CIKM'14: Faceted Search over Ontology-Enhanced RDF Data	Dmitriy Zheleznyakov	Shanghai, Peoples Republic of China	2014-11-05
Oral Presentation at DL'14: On Faceted Search over Knowledge Bases	Bernardo Cuenca Grau	Vienna, Austria	2014-19-07

2.6 Other Presentations for the Research Community

Description/Title	Presenter	Venue	Date
Seminar Talk at 'University of Oxford': Evaluating Ontology Matching and Mapping Repair Systems with Large Biomedical Ontologies	Ernesto Jimenez-Ruiz	Rome, Italy	2013-12-03
Course within the Master of Science in Computer Science at the Free University of Bozen-Bolzano: Ontology and Database Systems	Diego Calvanese and Werner Nutt	Bolzano, Italy	2014-03-03 to 2013-05-30
Invited seminar at the Free University of Bolzano: Tractable Approaches to Consistent Query Answering in Ontology-Based-Data Access	Riccardo Rosati	Bolzano, Italy	2014-04-11
Seminar Talk at TU-Dresden: Challenging Data Access with Optique Platform	Ian Horrocks	Dresden, Germany	2014-05-21
Course within the Erasmus Mundus Master's Programme in Information Technologies for Business Intelligence (IT4BI) at the Universitat Politècnica de Catalunya: Description Logics for Information Modeling and Data Access	Diego Calvanese	Barcellona, Spain	2014-06-09 to 2014-06-20
Seminar Talk at 'La Sapienza': Bootstrapping meets Ontology Alignment	Ernesto Jimenez-Ruiz	Rome, Italy	2014-06-27
Tutorial at the 8th Chinese Semantic Web & Web Science Conference: Ontology-based data access – Theory and Practice	Guohui Xiao	Wuhan, China	2014-08-10
Seminar Talk at Dagstuhl: On OBDA	Ian Horrocks	Dagstuhl, Germany	2014-08-12
Panel member at SOMER-project Final Workshop disseminating Optique	Ernesto Jimenez-Ruiz	Lisbon, Portugal	2014-09-17
Demo Presentation at DBOnto meeting: Optique Project	Ernesto Jimenez-Ruiz	Oxford, UK	2014-09-28
Oral Presentation at DBOnto meeting: Optique Project	Ian Horrocks	Oxford, UK	2014-09-28
Demo Presentation at DBOnto meeting: Semantic Faceted Search	Sarunas Marciuska	Oxford, UK	2014-09-28
Oral Presentation at DBOnto meeting: Semantic Faceted Search	Evgeny Kharlamov	Oxford, UK	2014-09-28

Chapter 3

Visibility to targeted industries (O10.2)

Optique has been very active in Year 2 in disseminating its results to targeted industies, in line with the priorities described in Sect. 1.2.

The project has given in total 25 presentations to business units at Siemens, Statoil and DNV GL. Optique members have at several occasions presented the project at high executive levels in these companies. The Optique project is by now known in these companies, far beyond the units that contribute actively to the project. This activity is summarized in Section 3.2. Section 3.3 summarizes the presentations for other companies; Section 3.4 summarizes our contributions to events that targets industry, while Section 3.5 reports the first Optique-arranged partner event. fluid Operations has promoted Optique widely to targeted companies, cf. Section 3.6. Optique took a major responsibility for the European Data Forum conference in 2014, addressed in Section 3.7.

An article in DNV GL's customer magazine PERSPECTIVES, distributed to more than 30.000 DNV GL customers worldwide, introduces Optique as an enabler to work smarter. The article can be found in Appendix B.

The last two sections review activities that are pointing towards Year 3 of the project, and that establish the basic infrastructure for the Optique ecosystem: Progress on the Optique Training Programme is reviewed in Section 3.8; the Optique Partner Programme is reviewed Section 3.9. Production of dissemination material has stared; an Executive Whitepaper is included in Appendix C.

3.1 Key performance indicators

Table 3.1: 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

3.2 On-site Presentations for Industrial Consortium Partners

3.2.1 Presentations at Siemens

Description/Title	Presenter	Venue	Date
Optique overview for Siemens Energy Oil&Gas Diagnostic Solutions	Mikhail Roshchin	Nuremberg, Germany	2013-11-27
Optique presentation for Siemens Energy, Transformer Services	Steffen Lamparter	Munich, Germany	2013-12-11
Optique requirements workshop at Siemens Energy Oil&Gas	Mikhail Roshchin	Finspong, Sweden	2013-12-18
Optique presentation for Siemens Industry: Diagnostic Solutions for Customer Service	Mikhail Roshchin	Nuremberg, Germany	2014-01-20
Optique presentation for Siemens Industry: Diagnostic Solutions for Electrical Networks	Mikhail Roshchin	Nuremberg, Germany	2014-01-20
Optique presentation for Siemens Corporate Technology Vienna	Steffen Lamparter and Thomas Hubauer	Vienna, Austria	2014-02-26
Optique presentation for Siemens Energy Oil&Gas, Software Diagnostic Solution for KPI calculations group	Mikhail Roshchin	Finspong, Sweden	2014-02-27
Optique presentation for Siemens Corporate Research	Steffen Lamparter	Princeton, NJ, USA	2014-03-19
Optique requirements workshop for Siemens Energy Oil&Gas, Software Diagnostic Solution for KPI calculations	Mikhail Roshchin	Finspong, Sweden	2014-03-24
Optique presentation for Siemens Industry Metal Technologies	Steffen Lamparter	Erlangen, Germany	2014-04-10
Integration of Optique Platform into Diagnostic Agent Architecture	Mikhail Roshchin	Lincoln, UK	2014-05-14/15/16
Integration of Optique Platform into Diagnostic Agent Architecture	Mikhail Roshchin	Finspong, Sweden	2014-07-08/09/10
The Optique Approach: End-user access to Big Data	Mikhail Roshchin	Muelheim, Germany	2014-07-15
Optique Contributions to Global Diagnostic System for Energy	Mikhail Roshchin	Muelheim, Germany	2014-08-12, 2014-08-13
Optique presentation for Siemens Industry Product Lifecycle Management	Steffen Lamparter	Teleconference	2014-09-03
Optique presentation at the Siemens CT Information Day for Energy	Mikhail Roshchin and Sebastian Brandt	Munich, Germany	2014-09-22
Rule-based Approach for Diagnosis: important technological building blocks	Mikhail Roshchin	Erlangen, Germany	2014-10-02

3.2.2 Presentations at Statoil

Description/Title	Presenter	Venue	Date
Presentation at Statoil Corporate IT: Optique status report and way forward	Arild Waaler and Martin G. Skjæveland	Stavanger, Norway	2014-09-29
Optique. Scalable End-user Access to Data	Martin Giese	Stavanger, Norway	2014-09-29
Optique. Scalable End-user Access to Data	Martin Giese, Marco Ruzzi, Evgeny Kharlamov and Dag Hovland	Stavanger, Norway	2014-09-30
Presentation at Statoil: The Optique Project	Martin G. Skjæveland and Dag Hovland	Stavanger, Norway	2014-10-03

3.2.3 Presentations at DNV

Description/Title	Presenter	Venue	Date
Optique presentation for DNV GL	Martin Giese and Arild Waaler	Høvik, Norway	2014-06-18
Optique- presentation to DNV GL Energy divison	Tore Hartvigsen	Høvik, Norway	2014-08-08
Optique- presentation to DNV GL Maritime divison	Tore Hartvigsen	Høvik, Norway	2014-08-14
Optique- presentation to DNV GL Oil & Gas divison	Tore Hartvigsen	Høvik, Norway	2014-09-04

3.3 On-site Presentations to External Companies

Description/Title	Presenter	Venue	Date
Optique presentation for a global IT company	Thomas Grauer	Schwalbach, Germany	2013-11-04
Optique presentation for a global financial services company	Thomas Grauer	Zurich, Switzerland	2013-12-12
Optique presentation for Schlumberger Information Solutions	Martin Giese and Arild Waaler	Aberdeen, Scotland	2014-01-14
Optique presentation for a global ERP vendor	Thomas Grauer	Frankfurt, Germany	2014-01-21
Optique presentation for an IT service provider	Thomas Grauer	Stuttgart, Germany	2014-02-04
Optique presentation for a consulting company	Thomas Grauer	Zurich, Switzerland	2014-02-14
Optique presentation for an automotive supplier	Steffen Bischoff	Herzogenaurach, Germany	2014-02-19
Optique presentation for a pharmaceutical company	Thomas Grauer and Ralf Scharly	Basel, Germany	2014-03-19

continued . . .

Description/Title	Presenter	Venue	Date
Optique presentation for an IT company	Thomas Grauer and Andreas Eberhart	Wallldorf, Germany	2014-03-21
Optique presentation for Gaian DB team of IBM	Martin Giese	teleconference	2014-04-10
Optique presentation for an IT company	Thomas Grauer and Ralf Scharly	Böblingen, Germany	2014-04-10
Optique presentation for an insurance company	Thomas Grauer and Andreas Eberhart	Zurich, Switzerland	2014-04-24
Optique presentation for a business process and IT outsourcing company	Thomas Grauer and Andreas Eberhart	Zurich, Switzerland	2014-04-24
Optique presentation for an industry engineering and services company	Thomas Grauer and Steffen Bischoff	Mannheim, Germany	2014-05-07
Optique presentation for an IT Service Provider	Thomas Grauer and Tobias Mathaess	Halle, Germany	2014-05-22
Invited Seminar at the IBM T.J. Watson Research Center: SPARQL-OWL Entailment Regime	Martin Rezk	Yorktown Heights, NY, USA	2014-06-02
Optique presentation for international Schlumberger audience	Martin Giese and Arild Waaler	Webinar	2014-06-23
Optique presentation at partner roadshow	Francesco Incorvaia	Dietzenbach, Germany	2014-07-02
Optique presentation at partner roadshow	Francesco Incorvaia	Dietzenbach, Germany	2014-07-09
Optique presentation for an university	Thomas Grauer and Steffen Bischoff	Frankfurt, Germany	2014-07-24
Optique- presentation to Norwegian Defence Estates Agency	Tore Hartvigsen	Høvik, Norway	2014-08-11
Optique presentation for an insurance company	Thomas Grauer and Andreas Eberhart	Aachen, Germany	2014-08-29
Optique presentation for an insurance company	Thomas Grauer and Andreas Eberhart	Aachen, Germany	2014-08-29
Optique presentation for an IT company	Thomas Grauer and Andreas Eberhart	Walldorf, Germany	2014-09-18
Optique- presentation to SICS (Swedish Institute of Computer Science	Tore Hartvigsen	Stockholm, Sweden	2014-09-19
Optique presentation at a partner event	Peter Haase	Mönchen, Germany	2014-09-28

continued ...

Description/Title	Presenter	Venue	Date
Optique presentation for an industry services company	Thomas Grauer and Steffen Bischoff	Bad Mergentheim, Germany	2014-10-02
Optique- presentation to Australian Oil& GAsComapny Woodside	Tore Hartvigsen	Høvik, Norway	2014-10-02
Optique presentation for a automotive company	Thomas Grauer and Christian Hötter	Stuttgart, Germany	2014-10-14
Optique- presentation to Norwegian public Railroad (NSB)	Tore Hartvigsen and Jarl Magnusson	Oslo, Norway	2014-10-23

3.4 Presentations at Industry Conferences and Workshops

Description/Title	Presenter	Venue	Date
Booth presentation at the Best in Cloud event	Thomas Grauer and Ralf Scharly	Frankfurt, Germany	2013-10-23/24
Individual meetings and Optique presentations at CIO event	Ralf Scharly	Hamburg, Germany	2013-11-28
MASTRO STUDIO: Ontology-based Data Access (in italian). (the talk has been repeated three times in the context of the Italian Forum for the Pubic Administration)	Domenico Lembo and Maurizio Lenzerini	Rome, Italy	2014-05-27
Booth presentation at SemTech	Thomas Grauer and Peter Haase	San Francisco, USA	2014-08-19/21
Invited talk at the First Business Development Platform between Italy and the German-Speaking Markets (PONTI 2014): Optique: Scalable End-user Access to Big Data at Siemens and Statoil	Diego Calvanese	Merano, Italy	2014-09-29
Individual meetings at a partner CXO event	Thomas Grauer	Athens, Greece	2014-10-07/08

3.5 Presentations at Optique arranged Conferences and Workshops

Description/Title	Presenter	Venue	Date
Project Optique Partner Program - Project briefing	Jarl Magnusson, Tore Hartvigsen, Arild Waaler and Martin Giese	Høvik, Norway	2014-06-25

3.6 fluidOps dissemination activities to targeted industries

Since the start of the Optique project, it is featured on the research overview page. Optique also has its own special landing page. The page includes a project description, benefits, partner logos, links to collateral and to Optique related websites.

With the relaunch of the new website in September 2014, fluidOps launched an AppGallery. The AppGallery provides a platform for partners. Using the AppGallery fluidOps features its own data management and data center and cloud management apps as well as partner apps. For Optique, fluidOps created a special app page including app description, listing of functionalities, link to the demo system and app factsheet. Tags allow users to easily find Optique.

In addition the Optique use case is described in the data management success stories section. A PDF is available for download.

3.7 European Data Forum

Figure 3.1: European Data Forum 2014, Flyer and organizing committee.

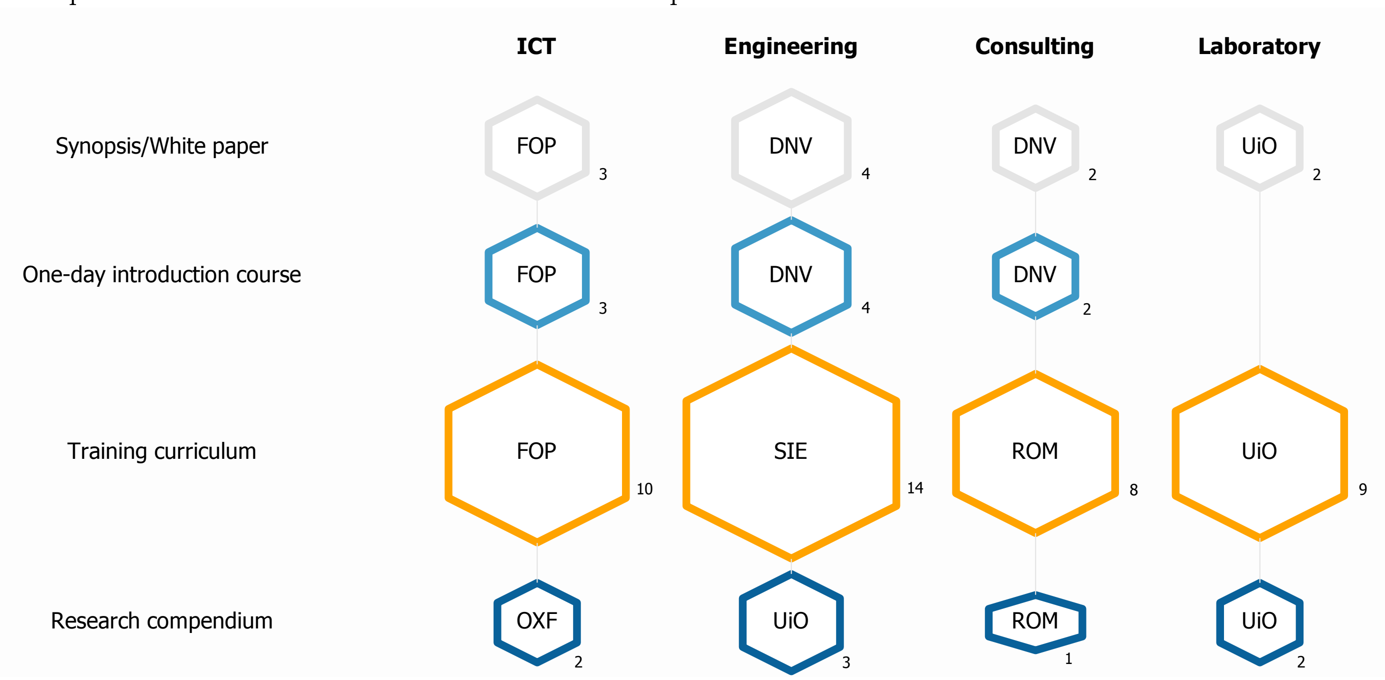
The European Data Forum 2014 (EDF) was organized March 19-20 in Athens with support from 22 partners (cf. the EDF2014 Final Report), including several EC projects like BIG, LOD2, PlanetData, Optique and EUCLID.

Optique contributed to the European Data Forum 2014 (EDF) by having two representatives in the organizing committee. Manolis Koubarakis (National and Kapodistrian University of Athens, Greece), contributed by being one of two conference chairs, and Peter Haase (fluidOps, Germany) in the role as the Industry Sponsoring Chair.

The Optique project was a major sponsor for the event and took part in the exhibition, where a special demo was given to Commisioner Kroes. Participation at EDF2014 allows Optique to keep up the attention of main representatives of the European Big Data community.

3.8 The Optique Training Programme

The Optique training programme is included in the deliverable D10.7 due project month 42. Current status on the ongoing work to prepare this deliverable is described below: The first Optique training course is scheduled for December 2014. Participants will be ICT personnel from Norwegian engineering company AIBEL and from DNV GL. The materials used in this first course, will be preliminary but will give good feed-back for preparing the remaining documentation for the coming courses to be arranged. The package of training documentation will be continuously prepared throughout the remaining project to reflect current development and also user needs. Courses will be arranged on an as-needed basis throughout the project. A plan for regular courses to be arranged after the completion of the project will be established as part of the Exploration activities as described in the D11.2 report.



3.8.1 Target audiences

The training programme is primarily designed to meet the needs of industrial enterprises preparing to implement Optique, targeting three verticals:

- ICT professionals, responsible for installation and running Optique services
- Engineers, end- users need for accessing data
- Consultants, internal or external, facilitating adoption of OBDA

In the figure above the "horizontals" illustrates categories documents to be produced while the "verticals" indicates the category of personnel that will be subject for training. It is expected to be a degree of overlap in the documentation. I.e. documents or parts of documents produced for one category personnel may be the same as included in the documentation for other. Work with preparing the training documentation is on track as specified below.

3.8.2 "Train the trainer"

A separate training programme will be required to "train the trainer"- that is material needed to train the personnel who shall arrange and give courses, perform more in depth customer support and advisory, arrange user workshops, etc. This can be arranged in different ways and has not been concluded yet. One model to be discussed is to arrange trainee posts where the trainers work closely with the developers for a period to obtain detailed knowledge of the various parts of Optique. Comprehensive training of the trainers can also be done by developing a set of cases under the supervision of Optique experts.

3.8.3 Training material to be produced

We have grouped the training material into four categories as illustrated in the figure on the previous page. The Optique partner assigned the responsibility for each particular category of training material and per target audience, is illustrated in the figure.

Synopsis/White paper

Optique has a stated goal of providing industry with better ways to access enterprise data. Recognizing that executives are vital to the adoption of new methods in the enterprise, the Optique training program needs to address the needs of high-level management as well as the ICT professionals, consultants that will be engaged in the implementation of an Optique solution and the End-users. White papers are intended to give a first introduction to Optique. The focus in these papers is possible effects of implementing Optique, risks and possible barriers. The White papers are on an introductory level, non-academic and targeting the specified personnel category. Work is currently ongoing on the ICT Synopsis/ White paper and on the Engineering White Paper. The preparation of the Consulting and Laboratory White papers will be initiated when the two first are completed.

One-day introduction course

The one-day introductory ICT course will be targeted to those who are responsible to install and prepare the Optique system within an organization (ICT personnel). A basic ICT background is expected from the attendees. After a full day course the attendees should be able to install and prepare the system for use within his organization. The course documentation shall include examples, technical prerequisites to successfully install and operate Optique, technical risks and barriers, etc. The deliverables under WP2 will give a good basis for specifying ICT training material. The Engineering one-day course will enable the Engineer to evaluate the potential of Optique and give training in navigating the Visual Query Interface. This will be done by accessing pre-defined example data bases e.g. the NPD fact database. The users should be able to define and run own queries and to evaluate the potential in using Optique with own data. Consulting one-day course will give the consultants sufficient knowledge to plan for implementing and integrating an Optique solution within an enterprise. This course will be an extract of the ICT and Engineering one-day courses and give additional information for the consultant. Existing material produced in the project will give a good basis for preparing the

Training curriculum

Training curriculum is special designed training material for those wanting more comprehensive and detailed knowledge about Optique. The curriculum should be used as a complement to hands-on training to be performed by the various categories personnel and reflecting the various needs of these groups. The idea is to develop courses that allow the users to work with real cases from their own company and are-of-interest. The curriculum will be developed by combining original Optique developed content with material from external sources. The curriculum will be suitable for in-depth study and may serve as reference material for implementation project. An Optique tutorial – getting ready to use has been prepared by FOP and is a good start on the training curriculum for ICT. The work with the other training curriculum has not been started.

Research compendium

Includes an index to (or a collection of) academic and scientific papers to be used for additional in-depth studies. No specific additional documents are planned prepared under this category in WP10. The index may include references to available articles, academic papers, etc.

3.9 Optique Partner Programme

Optique has initiated the work with establishing the Optique Partner programme. Please refer to WP11 Exploitation deliverable D11.2 for a detailed report. Included in the work with WP11 some more business executive focused documentation has been developed: A first meeting to recruit participants to the Partner Programme was arranged at DNV GL, Høvik Norway 25 June 2014. Participants were invited by a direct mail campaign to Norwegian companies expected to have “Big data” challenges. Material presented at the meeting was mainly slides and on-line demonstration of the NPD-fact pages VQI. A separate “Optique Executive White paper” was prepared for and handed out at the meeting. The first major Partner programme summit is currently being planned held in Early 2015. Work to identify companies/persons to invite are ongoing (ref. report D11.2 for details) and work is ongoing to specify the contents of an “Optique alpha package” with a sneak preview and hands-on examples of preliminary Optique software. The software will need to be accompanied with proper documentation/training material in order to be useful.

An article in DNV GL’s customer magazine PERSPECTIVES, distributed to more than 30.000 DNV GL customers worldwide, introduce OPTIQUE as an enabler to work smarter. Contact information is given in the article to the Optique project and engagement in the Partner Programme is encouraged. The hope is that this article will contribute to recruit companies within the Oil & Gas, Energy and Maritime industries, target group for this magazine, to the Optique partner programme. The article can be found in appendix B.

Chapter 4

Visibility to the general public (O10.3)

This chapter presents and discusses general dissemination activities beyond scientific and immediate industrial adopters to increase visibility to the general public.

Dissemination activities that have been realized in the second year are classified according to four main categories, namely website (see Section 4.3), social media (see Section 4.4), Optique public showcase (see Section 4.5), and news papers and magazine articles (see Section 4.6).

In this chapter we first, in Section 4.1, present key performance indicators to establish a quantitative basis for the evaluation of the activities conducted, followed by a general overview in Section 4.2. Then each category is presented in detail in the subsequent sections.

4.1 Key performance indicators

A set of performance indicators has been defined and calculated as shown in Table 4.1 for the second year, first year, and total. The key indicators suggest a considerable increase on the general visibility of the project.

Table 4.1: 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

4.2 Overview of dissemination activities

In the following we provide an overview of the activities toward the general public and suggest improvements for the coming project years.

The Optique website continues to be one of the primary instruments for reaching out to the broad public and has served as a digital face and identity of the project. Therefore, we believe that statistics that quantify the website traffic is of crucial importance to measure the success of website. Google analytics plugin has been installed to the system in its early phases. The website had around 13,000 page visits and attracted more than 4,000 unique visitors. Our YouTube channel is kept up-to-date and acquired more than 500 views – 3 times more compared to the first year.

Towards the end of second year a demo, which includes all the main components, was produced and made available on the YouTube. This has been a very valuable practice for the project towards the realization of a public showcase. The public showcase will be further developed over the entire project period and will, in its final form, be delivered in M48.

In the third year we expect increase number of social channels and encourage project partners to increase their dissemination activities towards the general public even more. This seems to be quite a realistic goal since the project will reach a higher level of maturity both with respect to the increasing amount of scientific results and the availability of a public showcase.

4.3 Website

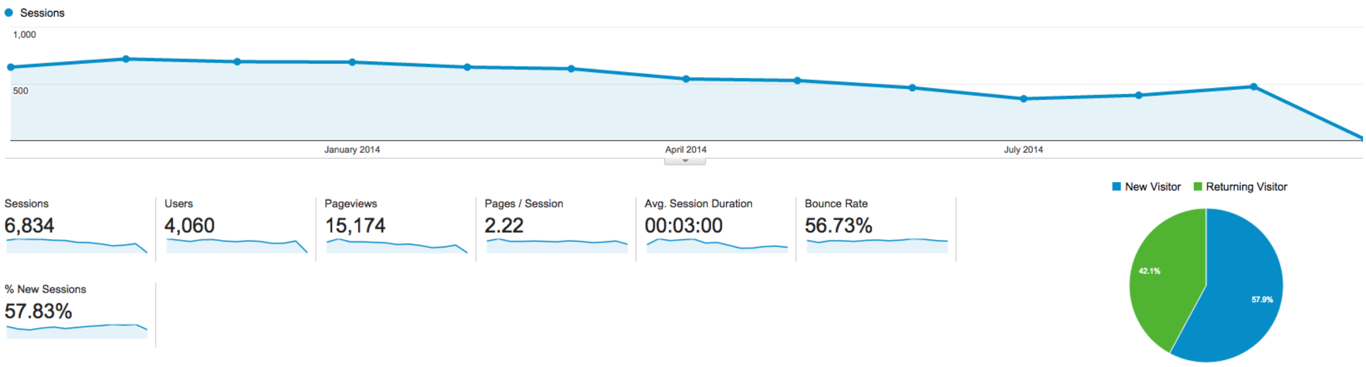


Figure 4.1: Optique website traffic between Nov 1, 2013 and Nov 1, 2014.

Optique website traffic is being monitored by Google Analytics tool.

Fig. 4.1 present an overview of website traffic and monthly distribution, while Fig. 4.2 provides an overview of the geographical distribution between Nov 1, 2013 and Nov 1, 2014. As the figures suggest, a broad community has been reached in the second year as well.

4.4 Social media

The Optique YouTube channel has 4 new videos in the second year – see Figure 4.3. This videos are meant to demo the second year version of the Optique platform. The channel can be accessed from `http://www.youtube.com/user/optiqueproject/`.

The videos uploaded to YouTube during Year 2 are listed in the following table:

Video Title	Description
Towards Ontology Based Data Access for Statoil: Part 1, Introduction	Overview of the solution
Towards Ontology Based Data Access for Statoil: Part 2, Visual Query Interface	Overview of OptiqueVQS
Towards Ontology Based Data Access for Statoil: Part 3, Installation Module	Overview of the installation module
Towards Ontology Based Data Access for Statoil: Part 4, Ontology Layering and Mapping Editor	Overview of the ontology layering component and of the mapping editor

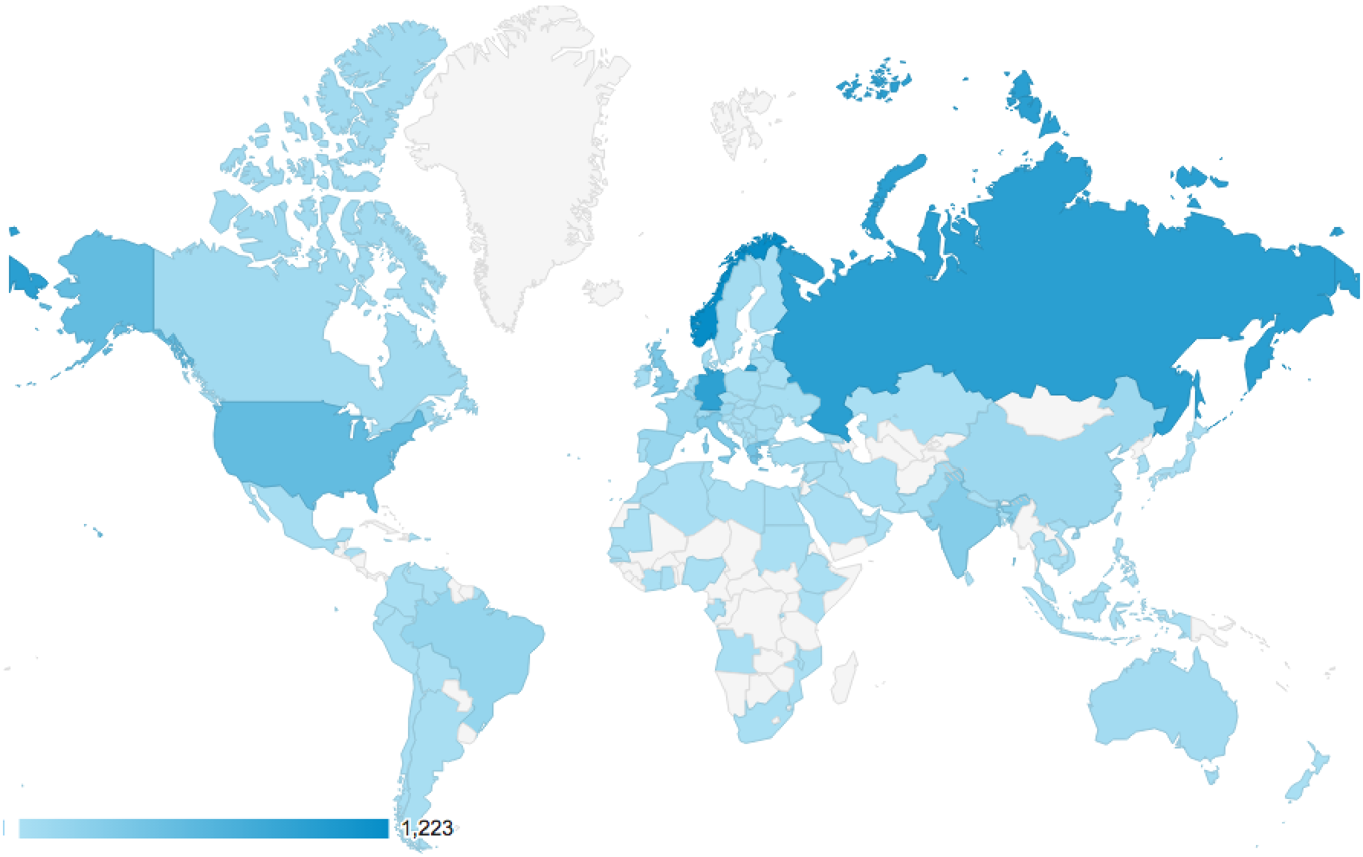


Figure 4.2: Geographical distribution of Optique website traffic between Nov 1, 2013 and Nov 1, 2014.

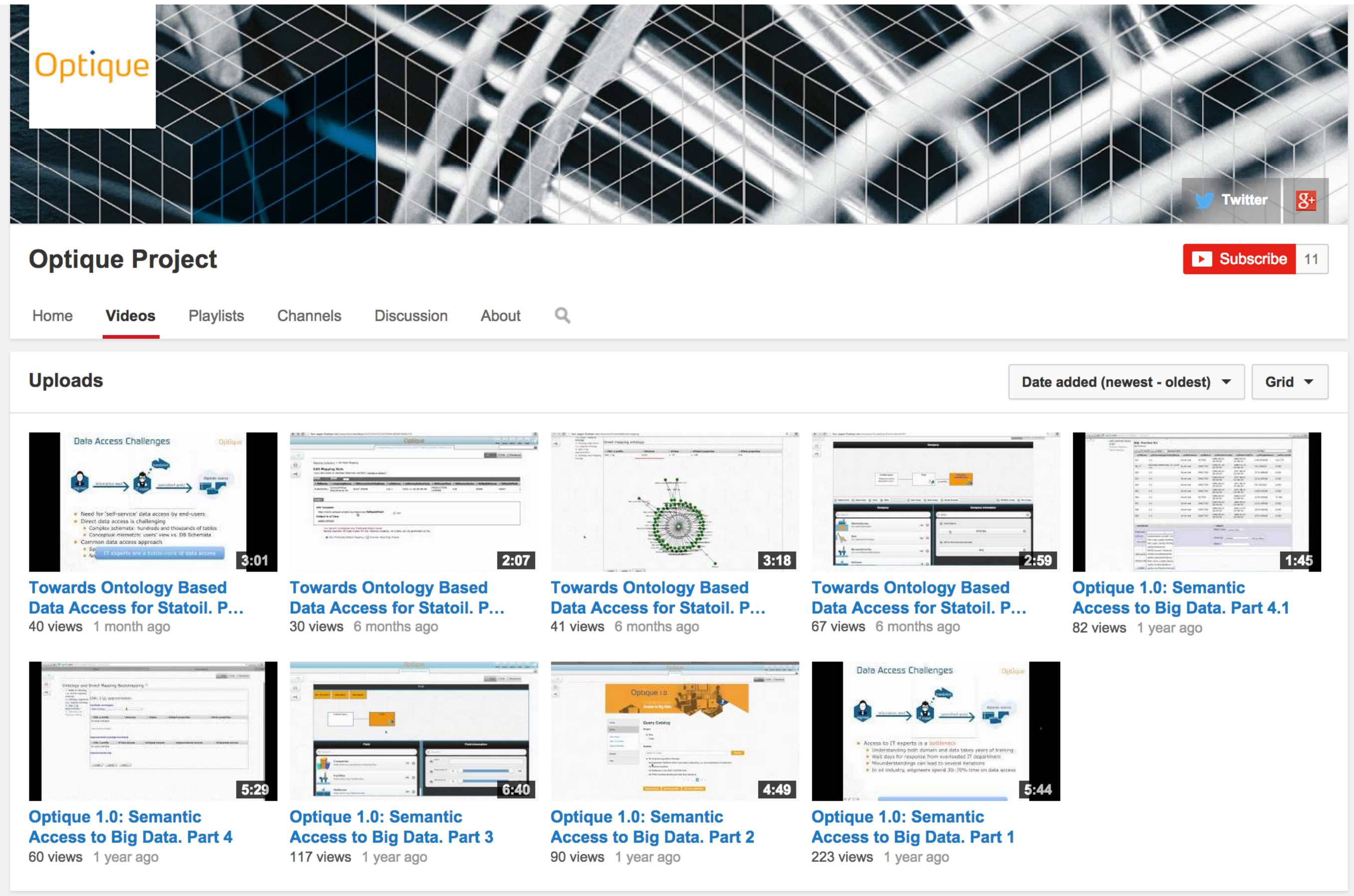


Figure 4.3: YouTube channel for Optique.

4.5 Steps towards the Optique Public showcase

The *Optique public showcase* is a publicly available installation of the Optique platform whose purpose is to allow the general public and the Commission to test and learn the features of the platform by directly using it in the same form as it is developed for the Siemens and Statoil use cases. The public showcase should include the same functionality as for the use case platform installations, but will offer access to different datasets, as the use case datasets are confidential and also probably too complex to fit as introductory example material for the general public.

The public showcase platform instance is installed at `http://fact-pages.fluidops.net` and contains a recent version of the Optique platform which is set up to access two datasets: the *NPD FactPages* and the open music encyclopedia *MusicBrainz*. Using public showcase platform is it possible to query the two datasets with the visual query system, the *OptiqueVQS*, and to administer the installation by, e.g., bootstrapping ontologies and mappings from data sources, aligning and layering ontologies, and browsing the installed ontologies and mappings. For security reasons some of these features are not openly available in the current installation, but require users to logon.

The use of the platform should be as self-explanatory as possible, but the project has also prepared demonstration papers with accompanying videos that explain and exemplify the features of the platform in general and the public showcase in particular. The aforementioned YouTube channel has been used to disseminate videos of the current showcase.

In the continuation of the project the public showcase will be used as the common testing use case for, to the extent possible, all technical and theoretical developments in the project. This will ensure that the public showcase will feature all developments of the Optique platform. We will also set up separate installations of the public showcase to serve the different needs that we have identified for a public showcase:

1. A stable installation that the project can use to demonstrate the Optique platform at dissemination events

2. A fully open installation that allows the general public to test all, and potential harmful, features of the platform without disturbing the stable version.

3. Possibly a trunk/unstable version showing the bleeding edge developments of the project.

We also intend to package the public showcase in a form that allows it to be distributed without, to the extent possible, requiring any installation of additional software. Additionally, we are investigating the possibility of extending the datasets in the public showcase with one or more datasets that are well-known and immediate understandable by larger parts of the targeted audience of the Optique platform.

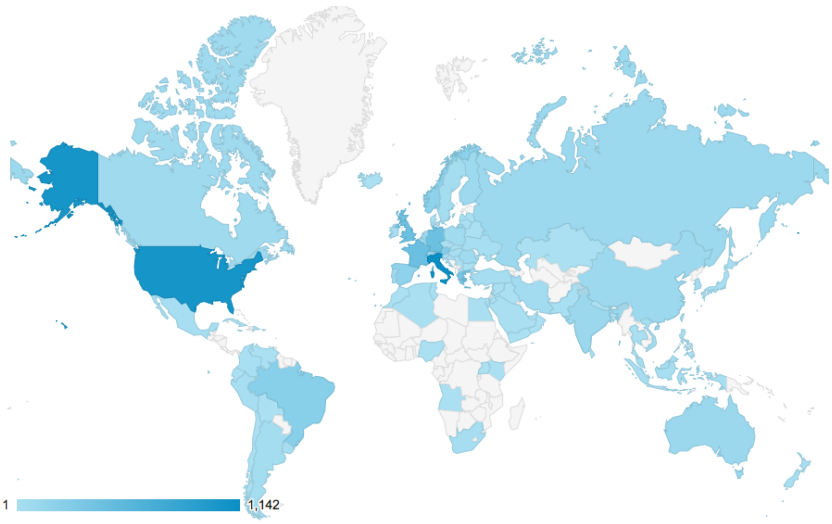
4.6 Presentations to the general public

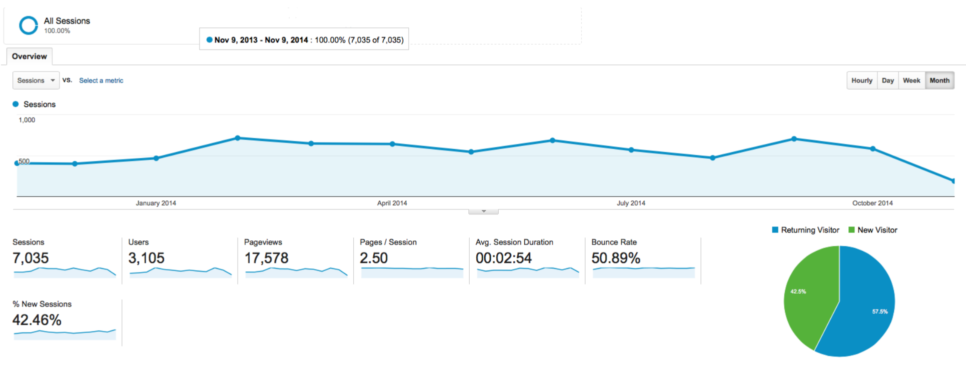
Description/Title	Presenter	Venue	Date
Seminar on H2020 arranged by The Norwegian Association of Higher Education Institutions	Arild Waaler	Oslo, Norway	2013-11-22
Seminar on Big Data. The Norwegian Ministry of Local Government and Modernisation	Arild Waaler	Oslo, Norway	2014-04-07
Press conference led by the Norwegian Minister of Education and Research, launching of the Government's new strategy for research `http://www.uio.no/om/aktuelt/arrangementer/andre/2014/juni/lanserer-ny-eu-strategi-for-forskning.html`	Arild Waaler and Martin Giese	Oslo, Norway	2014-06-05

4.7 Dissemination of *Ontop*

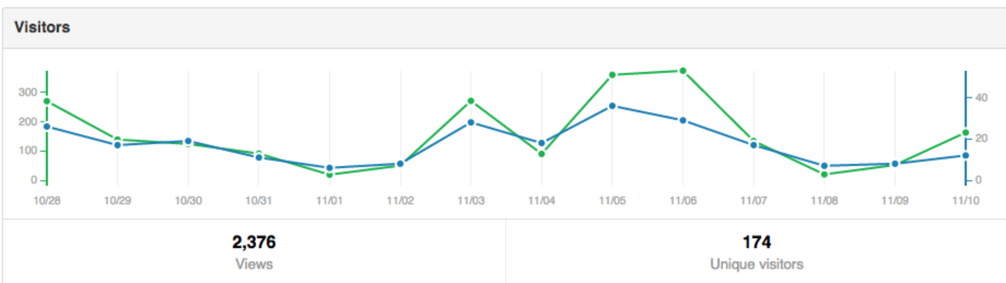
Ontop is the query translation component of the Optique Platform. During the reporting period, we have carried the following activities on opensourcing the system and building the user community to disseminate *Ontop* in the context on Optique.

- In agreement with the Optique consortium, we released *Ontop* under the open-source Apache license version 2. This license is permissive and friendly to both business and academic users.
- We keep the *Ontop* webpage updated and monitor its website traffic by the Google Analytics tool. We received more than 7000 hits per year, of which around 3000 are new visitors.



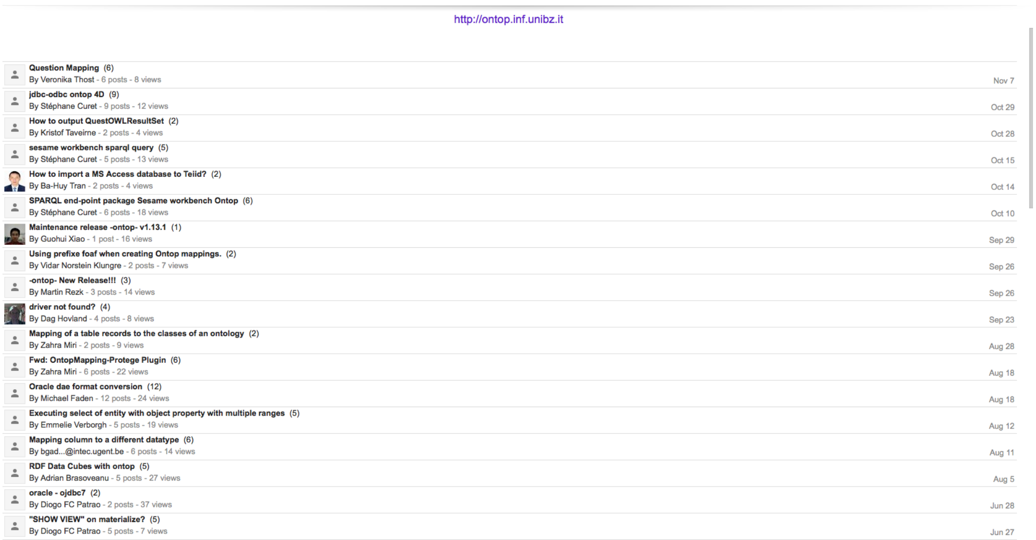


- The source code of *Ontop* has been migrated from an internal svn repository to a public Git repository hosted on Github.[1] Only in the last two weeks, the *Ontop* Github site received +2000 hits.

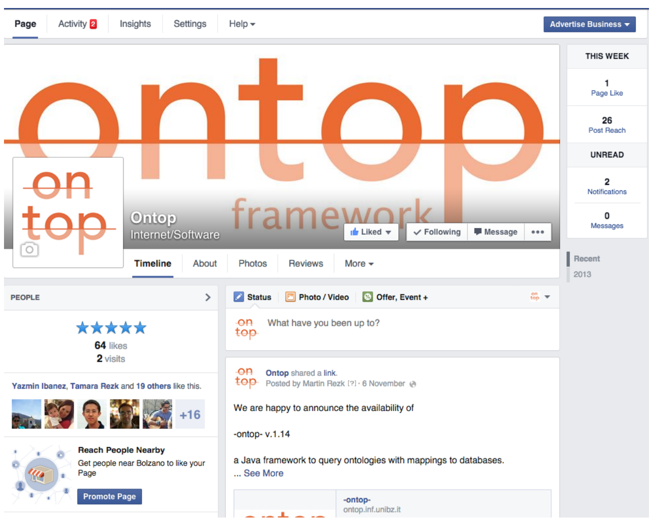


[1] `htpps://github.com/ontop/ontop`

- We have deployed the *Ontop* packages in the central Maven repository. Now users can easily integrate *Ontop* as Maven dependencies into their Java applications. We observe that this facilitates also the integration of *Ontop* into the Optique Platform.
- We created a Google group[2] on Aug 1, 2013 for the community of *Ontop* users. Users have actively posted 115 topics in the group about the usage, and have provided suggestions and bug reports about *Ontop*.



- We created a Facebook page for *Ontop*.[3]



- We announce every release of *Ontop* in several major mailing lists, such as: DBWorld, SemanticWeb, DL, and Protege-users.

[2] https://groups.google.com/forum/#!forum/ontop4obda
[3] https://www.facebook.com/obdaontop

Glossary

DL
Description Logics

EDBT
International Conference on Extending Database Technology

ESWC
Extended Semantic Web Conference

ISWC
International Semantic Web Conference

KI
German Conference on Artificial Intelligence

MEDES
International Conference on Management of Emergent Digital EcoSystems

NPD
Norwegian Petroleum Directorate

OBDA
Ontology-based Data Access

ORE
OWL Reasoner Evaluation

OWL
Web Ontology Languange

OWLED
OWL Experiences and Directions Workshop

RR
International Conference on Web Reasoning and Rule Systems

SPARQL
SPARQL Protocol and RDF Query Language

SQL
Structured Query Language

Appendix A

Article for the IEEE Computer magazine: "Optique - Zooming in on big data access"

Computer magazine is the IEEE Computer Society's flagship publication, covering all aspects of computer science, computer engineering, computing technology and applications. For the "Big data: management and application" special issue call, Computer solicits papers in big data management and applications, including the topics of big data infrastructures and distributed interoperability. As one of the leading research consortiums in big data, Optique partners have submitted a contribution to this call, titled "Optique - Zooming in on big data access". We aim to position the Optique project in the big data arena, so we present here how the Optique platform serves to unlock the data access bottleneck present in data-intensive industrial settings. The manuscript has been accepted for publication, subject to minor revisions.

Optique – Zooming In on Big Data Access

Martin Giese*, Peter Haase‡, Ernesto Jiménez-Ruiz§, Davide Lanti†, Özgür Özçep‖, Martin Rezk†, Riccardo Rosati**, Ahmet Soylu*, Guillermo Vega-Gorgojo*, Arild Waaler*, Guohui Xiao†

*University of Oslo, Norway {martingi, ahmets, guiveg, arild}@ifi.uio.no
†Free University of Bozen-Bolzano, Italy {davide.lanti@stud-, mrezk@, xiao@}inf.unibz.it
‡Fluid Operations AG, Walldorf, Germany peter.haase@fluidops.com
§University of Oxford, UK ernesto@cs.ox.ac.uk
‖Hamburg University of Technology, Germany oezguer.oezcep@tu-harburg.de
**Sapienza Università di Roma, Italy rosati@dis.uniroma1.it

***Abstract*—Despite the dramatic growth of data accumulated by enterprises, obtaining value out of it is extremely challenging. In particular, the data access bottleneck prevents domain experts from getting the right piece of data within a constrained time frame. The Optique Platform unlocks the access to Big Data by providing end users support for directly formulating their information needs through an intuitive visual query interface. The submitted query is then transformed into highly optimized queries over the data sources, which may include streaming data, and exploiting massive parallelism in the backend whenever possible. The Optique Platform thus responds to one major challenge posed by Big Data in data-intensive industrial settings.**

Keywords*—*8.II.VIII.VI Knowledge management applications, 8.II.IV.IV Distributed databases, 8.V.II.VI Graphical user interfaces, 8.II.IV.VIII Query processing, 8.II.IV.XIII Temporal databases

I. INTRODUCTION

This paper presents the Optique Platform, a new and novel solution to one of the great challenges posed by Big Data: End users' difficulty of getting hold of just the data they need within the time frame they have for accessing it.

In order for end users to create value out of the vast resources of data that is now to a rapidly increasing degree available, they need to be able to explore the data in new and unforseen ways. A key aspect of this is the ability to flexibly access the data, and in particular the ability to pose *ad hoc* queries that join information from different sources. Any system that, in the context of Big Data, pretends to put such capabilities in the hands of the end users must provide a wide range of advanced functionalities. The designers of such solutions must at least seek to touch a balance on the following, partially contradictory, requirements:

(i) *Prerequisites:* The system must have an acceptable installation overhead. In the context of Big Data this will in most cases rule out solutions that require, e.g., a massive reorganization of data sources or a massive manual re-engineering of meta-data, like tagging.

(ii) *Usability:* End users must be able to access the data on their own, without the need for a specialized IT support staff. In the context of Big Data, this requires that most of the complexity of the various data sources must by default be hidden from the end users. End users must then be exposed to details on demand, in a gradual and controlled way.

(iii) *Scalability:* The solution must scale, not only with respect to the volumes of Big Data, but also with respect to schema complexity and data velocity.

(iv) *Scope:* Big Data scenarios typically involve a wide variety of data types and sources, and a combination of static and streaming data arriving with a high velocity. The system must be able to accommodate the variety and velocity of the underlying data sources.

In this article, we concentrate on the information needs of end users from engineering disciplines that need to access large, *complex* and *structured* corporate data sources, i.e. not only large volumes of data, but also complex schemata, with varied types, and often (in Sect. VI) streaming data. These engineers require the ability to pose *ad hoc* complex, structured queries over such data, typically combining information from a variety of tables.

If we disregard for a moment the extreme size and complexity and the streaming aspect, the shape of the data and the targeted queries is therefore similar to the situation in conventional relational data stores. In order to provide an automated, end-to-end connection between complex information needs and relational data stores, a technology known as *Ontology Based Data Access (OBDA)* has recently emerged. The central idea of OBDA is to capture information needs in the form of (relational) queries, expressed in terms familiar and comprehensible to end users, and to translate these information needs into queries over the data sources.

More specifically, OBDA uses an *ontology* to describe the end users' domain vocabulary in a way that allows a rigorous mathematical interpretation. End user queries are formulated using that ontology. On the other hand a set of *mappings* is provided that describes for each concept, relationship, and attribute in the ontology, how it is represented in the data sources. The ontology and mappings together form a declarative description of the OBDA setup that can be used to drive the query translation process.

A number of prototypes implementing the OBDA idea have been developed, but they concentrate mostly on the aspect of rewriting formal relational queries from an end user schema to a source schema. This is important and challenging, but we argue that it is only one aspect of what is required to provide an end-to-end solution for accessing Big Data sources. In the remainder of this article we will point out some ways in which the conventional OBDA approach breaks down under the stress of Big Data, and present how the Optique Platform remedies the situation.

II. The Optique Platform

The Optique Platform [4] is based on OBDA, but goes beyond the OBDA paradigm by addressing the four requirements for Big Data Access from Sect. I: (i) *Prerequisites:* OBDA requires an ontology and mappings. In a Big Data setting, these can become enormous artifacts that need to be authored and maintained. (ii) *Usability:* most end users cannot be expected to know a formal query language to formulate their information needs. They need the support of a user interface for query formulation, and one that can help them to find their way in an ontology with thousands of concepts. (iii) *Scalability:* existing OBDA approaches do not deal well with very large ontologies and mappings. (iv) *Scope:* much of the relevant data is often more than 'relational,' it may be temporal, streaming, geospatial, etc., and a system needs to be equipped to handle such data types appropriately.

The Optique project[1] is addressing these issues, driven by real-world case studies provided by Siemens AG and Statoil ASA. The project is developing an integrated end-to-end platform for end user access to Big Data, according to the architecture in Fig. 1. Note in particular the Query Formulation and Ontology & Mapping Management components that extend the classical OBDA setup of Query transformation, planning and execution. Also note that query execution includes the combination (federation) of information from distributed data sources, and includes the processing of streaming data and results.

The Optique platform is designed to provide end user access to data without moving it out of the storage in which it currently resides. Typically, this will be a RDBMS. It is known however that RDBMS are not sufficiently scalable for many Big Data applications. Therefore the Optique platform includes the ADP (Athena Distributed Processing) system [5] as a massively parallel database back-end, that will improve scalability whenever the application allows relocating the data to an elastic cloud. Moreover, the capabilities of ADP are used for federation between multiple data sources, and to combine streaming and static data, see Sect. VI.

In the remainder of this article we present the different components of the Optique platform, stressing how the Big Data challenges are addressed. To illustrate our points we employ the following running example:

> *Which hydrocarbon fields containing a wellbore with gas content.*

We assume that the data is stored according to the following relational schema:

- ***Well***(<u>*id*</u>, *name, type*)
- ***Field***(<u>*id*</u>, *name*)
- ***Wellbore***(<u>*id*</u>, *name, content, well_fk, field_fk*)
- ***DevelopmentWellbore***(<u>*wellbore_fk*</u>, *production_facility*)

where primary keys are underlined, and foreign keys' names end in *_fk*.

This example is based on a case study we performed on the Norwegian Petroleum Directorate's FactPages data set [10]. This is one of the information sources routinely used by domain experts in the Optique project's Statoil case study. The query is similar to real information needs in the Statoil case study, but kept simple for the purpose of illustration. Note that there is a dedicated table in the schema for development wellbores, which are a special kind of wellbore. In Sect. VI we will extend this example with a temporal dimension.

III. Reducing Installation Overhead

Building an ontology and connecting it to the data sources via mappings is a costly process, especially for large and complex databases. To aid this process, tools that can extract a preliminary ontology and mappings from the source schema are useful. To improve the quality of the ontology and mappings, but also to maintain them, e.g. when source schemas change, tool support for editing ontologies and mappings is crucial. While ontology editing is well covered by existing tools, mapping editors are not as readily available, which is why such an editor [9] is developed as part of the Optique platform.

In order to ease the production of initial versions of the ontology and mappings, the Optique platform includes a *bootstrapping* component that takes a set of database schemata as the input, and returns an ontology and a set of mappings that connect the terms occurring in the ontology to the schema elements. "Guessing" an ontology from a database schema is no easy task, since the database modelling step that produces a database schema from (explicit or implicit) knowledge of a domain is typically lossy. Automatic bootstrapping means to make a best effort at reverse engineering this step and is necessarily imperfect. However, depending on the quality of the data source schemata, the results often provide a very good starting point for later manual optimisations that can be applied as required.

Our current implementation is based on the approach called 'direct mapping' by the W3C:[2] every table in the database (except for those representing many-to-many relationships) is mapped to one class in the ontology; every data attribute is mapped to one data property; and every foreign key to one object property.

Additionally, explicit and implicit database constraints from the schema can be used to enrich the ontology with axioms about the classes and properties from these direct mappings. For instance, in the example schema, the DevelopmentWellbore table's primary key is also a foreign key pointing to the Wellbore table. It turns out to be an effective heuristic in such cases to add class inclusion axioms between the directly mapped classes, in this case $\mathsf{DevelopmentWellbore} \sqsubseteq \mathsf{Wellbore}$.

Even after applying such heuristics to enrich the ontology, it will still usually be too close to the source schema. It is however increasingly often the case that a high quality ontology of (parts of) the domain already exists, that captures the domain experts' vocabulary better than the directly mapped ontology.

[1] http://www.optique-project.eu/

[2] http://www.w3.org/TR/rdb-direct-mapping/

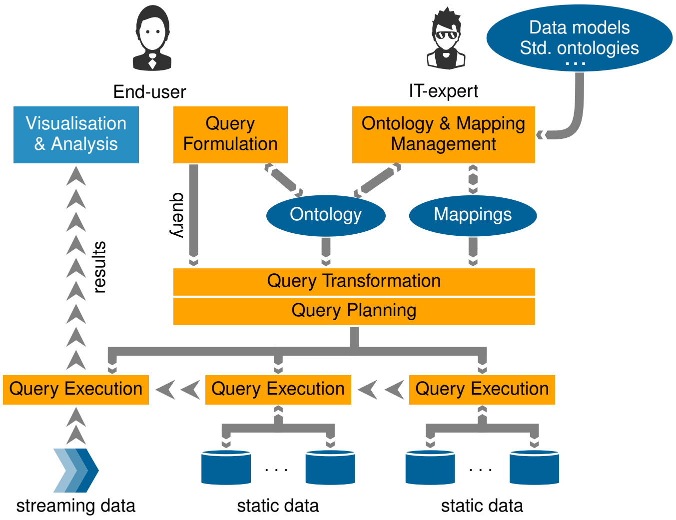


Fig. 1. Architecture of the Optique Platform

When such a high quality ontology is available, the bootstrapping component allows importing it and using it in the bootstrapping process. This is achieved through a heuristic alignment of the directly mapped and the imported ontologies, using the LogMap system [3] LogMap is a highly scalable ontology matching system that discovers ontology-to-ontology mappings, e.g. class equivalence axioms, between the vocabularies of the input ontologies.

Special care needs to be taken to avoid introducing unwanted consequences: for instance the bootstrapper will avoid adding alignment axioms that would lead to inconsistencies, or faulty consequences like Well $\sqsubseteq$ WellBore that are not supported by the domain ontology. This is based on novel techniques to avoid violations of the so called consistency and conservativity principles [11].

We evaluated our bootstrapper on Statoil's corporate exploration and production data store, an SQL database comprising over 1500 tables with a total of over 19000 columns. Parts of the directly mapped ontology were automatically aligned with an independently hand-crafted ontology based on the NPD FactPages data set. The resulting ontology was sufficient to cover 30–50% of a previously collected catalogue of representative information needs of petroleum exploration experts.

This is a very good coverage for a fully automated process, but to achieve a really high coverage of end-user information needs, the ontology and mappings will have to be improved further; and both have to be maintained when end user requirements or data sources change. The Optique platform includes a dedicated mapping editor for this purpose [9]. An important insight is that the data sources, ontology and mappings together are a complex combination of artifacts where it is easy to introduce mistakes. To support the maintenance and evolution of mappings, advanced analysis techniques are required that are based on the equivalence between the SQL queries in mappings and formulae of first-order logic [6].

IV. Eliciting Information Needs

As discussed above, end users should be able to pose expressive queries for an arbitrary domain in order to exploit available Big Data sources. Ideally, the query specification mechanism should be usable even for users without specialised IT skills. Accordingly, the following approaches can be found in the literature:

- Query language editors allow highly expressive queries through query languages such as SPARQL[3]. However, knowledge of a query language is a too exigent requirement for end users in most cases.
- Information retrieval approaches mimic traditional keyword queries. Although simple to use, expressiveness is rather limited.
- Natural language interfaces aim to interpret a query as a whole in order to give more relevant responses than with keywords. However, ambiguities and linguistic variability limit their effectiveness.
- Visual approaches try to achieve good usability and sufficient expressiveness, although finding a good tradeoff is difficult. For example, visual SPARQL query builders are still too complicated for mainstream users, while typical facet-based systems limit queries to one class.

To address this challenge we have proposed a novel visual query interface named OptiqueVQS [12]. It allows the formulation of rather expressive queries through a visual user interface that hides the syntax of the query language – see Figure 2. The bottom-left widget can be used to browse the concept taxonomy and include new concepts to the query. This is shown as a tree in the top widget and can be manipulated, e.g. to select or delete a concept variable, reacting after each query change. A selected concept can be further refined using the controls of the bottom-right widget, which are built from the concept properties in the ontology. Figure 2 represents the query built with OptiqueVQS for the information need presented in Sect. II.

[3] http://www.w3.org/TR/sparql11-query/

Fig. 2. Snapshot of the query interface of OptiqueVQS.

Overall, OptiqueVQS integrates concept browsing, facet-based search and visual query manipulation in the query interface. Behind the scenes, a SPARQL query is built and sent to the query transforming component of the Optique platform – for our running example we obtain:

```
SELECT ?field WHERE {
  ?field a npdfp:Field .
  ?wb a npdfp:Wellbore ;
      npdfp:wellboreForField ?field ;
      npdfp:wellboreContent "GAS"^^xsd:string .
}
```

Since ontologies can be very large in a Big Data context, information about classes is loaded gradually on demand and there are input fields to find classes and properties without having to navigate endless lists. Moreover, the widget-based architecture allows the to accommodate different needs in a modular way, e.g. we have developed widgets to select geospatial objects on a map, or to specify a time window for streaming data.

We have tested OptiqueVQS in two preliminary user studies. Results show that participants were able to solve non trivial query tasks, even without previous training. This confirms our thesis that interaction with a query, instead of directly with the data, is suitable for end users if implemented carefully.

V. Scaling Up Query Transformation

Query transformation, the task of rewriting end-user queries to queries over the data sources, for OBDA is quite well researched, with several implementation available. But previous efforts have been directed mostly at data volume: the important issue was making the translation process independent of the size of the data store.

When large, complex schemas come into the equation in addition to large volumes, query transformation needs to be rethought: rewriting can be expensive, and even prohibitive, as the size of the ontology and the number of mappings increase. The *Ontop* query rewriting system,[4] which is used in Optique, solves this problem by embedding the consequences of the ontology into the mappings, generating so called $\mathcal{T}$-mappings. This is done off-line, i.e. independently of any concrete query. Actual queries then need to be rewritten only w.r.t. a small set of necessary mappings.

For instance, in the running example, we have the axiom

$$\mathsf{DevelopmentWellbore} \sqsubseteq \mathsf{Wellbore}$$

and the following mappings to the relational schema

$\mathsf{Wellbore}(id) \leftarrow$ `SELECT` id `FROM Wellbore`
$\mathsf{DevelopmentWellbore}(wellbore\_fk) \leftarrow$
  `SELECT` $wellbore\_fk$ `FROM DevelopmentWellbore`

[4] http://ontop.inf.unibz.it/

Ontop will embed the axioms above in the mappings by generating the following new $\mathcal{T}$-mapping:

Wellbore(id) ←
  SELECT id FROM Wellbore
  UNION
  SELECT $wellbore\_fk$ AS id FROM DevelopmentWellbore

Then to answer the triple pattern

```
?wb a :Wellbore
```

in our running example, we only need to consider the mappings and the data, while we can disregard the ontology axioms. Unlike other systems (triple stores for instance) there is no need to chase GBs of data to generate all the facts derived by the ontology.

The architecture of *Ontop* is depicted in Figure 3. The input consists of an ontology; a database, possibly including integrity constraints; and the mappings connecting the ontology and the database.

The *off-line* stage of Ontop pre-processes the ontology, mappings and database schema as follows:

1) The ontology is classified using an ontology reasoner.
2) $\mathcal{T}$-mappings are constructed to drastically simplify query rewritings. Intuitively, $\mathcal{T}$-mappings embed the hierarchy that can be inferred by the ontology into a new, extended set of mappings.
3) The $\mathcal{T}$-mappings are then optimised in order to eliminate redundancies.

The *on-line* stage takes a SPARQL query and unfolds it with respect to the $\mathcal{T}$-mappings. Then, it optimizes this query into an SQL query. The optimized query is executed by the data source.

We have shown using a range of standard benchmarks that *Ontop* outperforms many commercial triples stores with and without reasoning enabled, see e.g. [8].

In order to test the efficiency of *Ontop* in a challenging OBDA setting, we have created a scalable benchmark based on the previously mentioned NPD FactPages data set, ontology, and mappings [10], as well as a set of real-world queries created by users of the FactPages, but equiped with a data generator able to increase the initial dataset in order to allow a scalability analysis [2].

The results for *Ontop* show that most of the queries are translated into a single join query that can be evaluated efficiently—11000 queries per hour on the original 4**M** triples dataset, and that the execution times scale linearly w.r.t. the growth of the dataset—14 queries per hour on the 6**B** triples dataset.

In the general case, a user query translates to multiple SQL queries, some of which may be redundant. Therefore, being able to detect containment between queries is essential to efficient query answering in OBDA. The general problem of query containment between SQL queries is undecidable, as SQL can express arbitrary first-order queries. However, as theorem provers for first-order logic are quite well developed, we expect to be able to use them for detecting query containment in many cases. Once redundant queries are removed, we achieve further efficiency gains by executing queries in parallel using ADP.

VI. Streaming Queries

Most Big Data processing involves a temporal dimension—be it as time attributes within a static DB, or as timestamps on data items arriving in one of possibly many stream sources such as sensor and event data. Hence, an information system intended to cover a wide range of Big Data processing scenarios has to provide an expressive and intuitive means for querying and interpreting temporal and streaming data.

Adding the temporal and stream dimension to the OBDA paradigm leads to adaptations and changes in almost all components of the system. In the Optique platform, this includes the following:

- extensions to the visual query interface;
- the novel query language STARQL, providing operators needed for temporal and stream reasoning;
- extensions to the ontology and mapping languages, to specify time attributes in the source data, which are mapped to timestamps on assertions at the ontology level;
- extensions of the highly distributable SQL backend system ADP [5] with a means to handle streams as (virtual) tables.

The query-language framework STARQL (Streaming and Temporal ontology Access with a Reasoning-Based Query Language) [7] provides an expressive declarative interface to both types of temporal data (historical and streaming) within the paradigm of OBDA and as such adds a new approach to the recent venture of temporalizing and streamifying OBDA, see e.g. [1].

To explain the handling of STARQL queries, we extend the running example from the beginning of the paper: For all fields containing a well with gas content, we want to compute a moving 3 month average over the field's gas production. This can be done using the following STARQL query:

```
CREATE STREAM Sout AS
SELECT ?field, AVG(?p), NOW
FROM proddata[NOW-3month, NOW]->1 month, npdfp
WHERE {
       ?field a :Field .
       ?wb a :Wellbore ;
       :wellboreForField ?field ;
       :wellboreContent "GAS"^^xsd:string .}
SEQUENCE BY StdSeq AS SEQ
HAVING EXISTS i in SEQ:
       GRAPH i : { ?wb :producedGas ?p }
```

In STARQL, querying historical data and querying streaming data proceeds in an analogous way. In both cases, the query may refer to static data, i.e, data that are not equipped with a timestamp as they are assumed to be non-temporal or to hold at every time point (as in [1]). In general, the static part of a STARQL query is formulated in the WHERE clause. For instance, the example query asks for wellbores containing gas, which is assumed to be a static property for this query.

The answers produced by the static sub-query, are used as a pre-filter for the stream processing in the remainder of the query. This separation between the static and dynamic aspects is an important feature of STARQL that makes it possible to

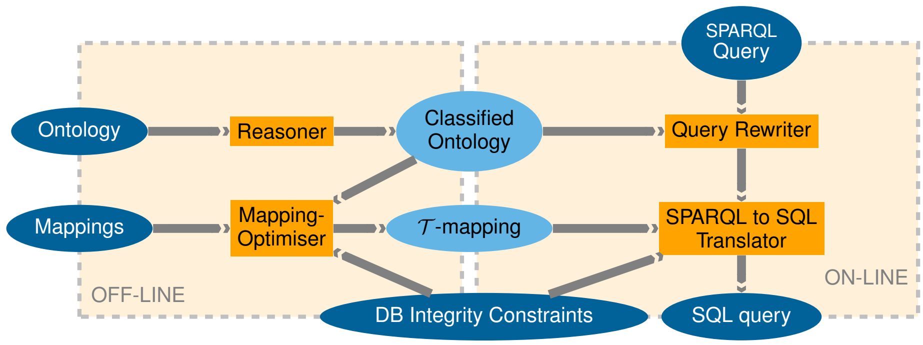


Fig. 3. *Ontop* Architecture

process high volume static data and high velocity streaming data within one query.

Relevant slices of the temporal data are specified with a window; in the case of historical data, this is a window with fixed endpoints; in case of streaming data, it is a moving window that contains a reference to the developing time `NOW` and a sliding parameter that determines the rate at which snapshots of the data are taken. In the example query, we have a three month sliding window over gas production data stored in `proddata`. The contents of the temporal data are grouped according to a sequencing strategy into a sequence of small graphs that represent different states. In the example, the sequencing strategy uses standard sequencing according to which assertions with the same timestamps come into the same graph. But the general state oriented view allows to implement other more advanced sequencing strategies such as various clustering strategies known from machine learning and time-series analysis.

On top of the sequence, relevant patterns and aggregations are formulated in the `HAVING`-clause, using a highly expressive template language. The example query asks for every three-month snapshot whether there is a state `i` such that field `?f` produced gas volume `?p` at that state. The safety conditions [7] guarantee that the STARQL query can indeed be unfolded into an ADP query.

VII. Conclusion

Giving end users with limited IT expertise flexible access to Big Data is a major bottleneck in data-intensive industries. We have argued how ontology-based data access (OBDA) can provide a solution: By capturing the end users' vocabulary in a formal model (ontology), and maintaining a set of mappings from this vocabulary to the data sources, we can automate the translation work previously done by the IT-experts. Yet current OBDA systems lack a holistic approach to the problem of establishing an integrated, end-to-end connection from the end user to large scale and distributed data sources. Specific problems pertain to usability, prerequisites, scope, and scalability. We have shown how these problems can be overcome by a novel combination of techniques, encompassing an end user oriented query interface for eliciting information needs, semi-automated methods for managing ontologies and mappings, new techniques for scalable query rewriting, temporal and streaming data processing.

These techniques are implemented in an integrated platform, in which all components ranging from low-level distributed query execution to visual end user interfaces interact seamlessly. To tackle the issue of cross-component communication and optimization, components communicate through a unified semantic layer of abstraction. Knowledge artifacts such as ontologies, mappings, metadata and queries are stored in the central semantic repository based on open standards such as OWL 2 QL for ontologies, SPARQL as a query language, and R2MRL for the mappings. As a result, the Optique platform provides a single point of entry for administrative tasks (e.g. management of mappings and ontologies) as well as visual components through which end users can satisfy their information needs in interacting with Big Data.

Acknowledgments: The authors would like to thank all other members of the Optique project for their contributions to the project; and in particular Diego Calvanese, Ian Horrocks, Evgeny Kharlamov, Ralf Möller, Domenico Fabio Savo, Evgenij Thorstensen, and Dmitriy Zheleznyakov for their contributions to and comments on drafts of this manuscript.

This work has been partially funded by the European Commission under FP7 Grant Agreement 318338, "Optique."

References

[1] Stefan Borgwardt, Marcel Lippmann, and Veronika Thost. Temporal query answering in the description logic DL-Lite. In Pascal Fontaine et al., editors, *Proc. FroCoS 2013*, volume 8152 of *LNCS*, pages 165–180. Springer, 2013.

[2] Diego Calvanese, Davide Lanti, Martin Rezk, Mindaugas Slusnys, and Guohui Xiao. A scalable benchmark for OBDA systems: Preliminary report. In *Proc. OWL Reasoner Evaluation (ORE 2014)*, volume 1207 of *CEUR*, 2014.

[3] Ernesto Jiménez-Ruiz, Bernardo Cuenca Grau, Yujiao Zhou, and Ian Horrocks. Large-scale interactive ontology matching: Algorithms and implementation. In *Proc. ECAI 2012*, pages 444–449, 2012.

[4] Evgeny Kharlamov, Ernesto Jiménez-Ruiz, Dmitriy Zheleznyakov, et al. Optique: Towards OBDA systems for industry. In *Eur. Sem. Web Conf. (ESWC) Satellite Events*, pages 125–140, 2013.

[5] Herald Kllapi, Eva Sitaridi, Manolis M. Tsangaris, and Yannis E. Ioannidis. Schedule optimization for data processing flows on the cloud. In *SIGMOD Conference'11*, pages 289–300, 2011.

[6] Domenico Lembo, Jose Mora, Riccardo Rosati, Domenico Fabio Savo, and Evgenij Thorstensen. Towards mapping analysis in ontology-based data access. In *Proceedings of RR 2014*, 2014. To appear.

[7] Özgür L. Özçep, Ralf Möller, and Christian Neuenstadt. A stream-temporal query language for ontology based data access. In *KI 2014*, volume 8736 of *LNCS*, pages 183–194. Springer, 2014.

[8] Mariano Rodriguez-Muro et al. Evaluating SPARQL-to-SQL translation in ontop. In *Proc. OWL Reasoner Evaluation (ORE 2013)*, volume 1015 of *CEUR*, pages 94–100, 2013.

[9] Kunal Sengupta, Peter Haase, Michael Schmidt, and Pascal Hitzler. Editing R2RML mappings made easy. In *ISWC 2013 Posters & Demos*, pages 101–104, 2013.

[10] Martin G. Skjæveland, Espen H. Lian, and Ian Horrocks. Publishing the Norwegian Petroleum Directorate's FactPages as semantic web data. In Harith Alani et al., editors, *Proc. ISWC 2013*, volume 8219, pages 162–177. Springer, 2013.

[11] A. Solimando, Ernesto Jiménez-Ruiz, and G. Guerrini. Detecting and correcting conservativity principle violations in ontology-to-ontology mappings. In *Proc. ISWC 2014*, 2014.

[12] Ahmet Soylu et al. OptiqueVQS – towards an ontology-based visual query system for Big Data. In *Proc. MEDES 2013*. ACM, 2013.

BIOGRAPHIES

Martin Giese, Peter Haase, Ernesto Jiménez-Ruiz, Davide Lanti, Özgür Özçep, Martin Rezk, Riccardo Rosati, Ahmet Soylu, Guillermo Vega-Gorgojo, Arild Waaler, and **Guohui Xiao** are researchers in the Optique project.

Appendix B

Article in PERSPECTIVES: "Tackling the Big Data Bottleneck"

PERSPECTIVES is a publication produced by DNV GL. It offers oil and gas industry professionals valuable insights and analysis of key challenges and opportunities facing the sector, and explores technical and macro-economic issues clouding the future. PERSPECTIVES is produced in print bi-annually to an audience of more than 6,000 senior industry players. The publication is also distributed digitally to a global audience of more than 30,000 oil and gas professionals. Dedicated electronic editions are sent to DNV GL pipelines, gas and subsea and floating production-focused contacts across the world. In the next edition of the magazine an article about Optique is included as an example where DNV GL explore the importance of working smarter within the industry.

TACKLING THE BIG DATA BOTTLENECK

A USD17.5 million joint industry-academia project is speeding up information retrieval from complex databases

Oil and gas companies face a long-standing industry problem in accessing data. It is one of the challenges of larger, deeper and more remote operations, but now comes with the added complexity of collecting and interpreting a huge surge of real-time, digital information generated by multiple fields and plants.

Over recent years, a range of advanced tools have come to the market to help operators make sense of this so-called 'Big Data', in order to understand how to bolster performance across thousands of wells and, in real-time, monitor the condition of advanced equipment. But the technical limitations of today's computing systems are already struggling to manage the amount of information that some operators are required to handle, sparking a search for smarter ways in which data could, and should, be analysed. Big Data solutions aim to effectively aid decision – making, allowing users to work more effectively by focusing on accurate information and how to use it when required.

Big Data is often characterised and quantified by reference to 'the three Vs' – volume, velocity, and variety – a description originally coined by Doug Laney, now research vice president of technology analysts Gartner Research.[1]

In explorative drilling, for instance, a company will evaluate an area, drill a well, gather real-time data and input this into its system to inform planning for the next well before drilling it. Companies may re-evaluate fields every week and in many places, driving the volume of data ever upwards.

"Our ambition is to allow engineers to independently navigate, retrieve and simplify complicated data, and reduce the time it takes to access what they need ... to just a matter of minutes"

Professor Arild Waaler, coordinator, Optique JIP

A collaborative response

As companies seek smarter ways to handle the influx of complex data, joint industry projects (JIPs) have begun to explore ways of saving time, money and energy through shared goals.

One such initiative is Optique, a four-year joint industry project between several world-leading academic institutions and industry partners. It exploits recent advances in semantic technologies, in which the meaning of data is explicitly represented as part of the data model. The aim is to develop a software platform to provide end-users with flexible, comprehensive, and timely access to large and complex industrial data sets – in processing petabytes of well data, for example – by making computers use the language users understand and are used to.

University of Oslo (UiO) professor Arild Waaler, who coordinates Optique, initiated the project in 2010 and has received backing from Norwegian oil company Statoil, DNV GL, German engineering group Siemens, and fluid Operations, a German provider of innovative cloud and data management solutions. The EUR13.8 million (USD17.5 m) programme launched in December 2012 with EUR9.7m European Union funding.

The Optique team expects its approach to reduce turnaround time for >

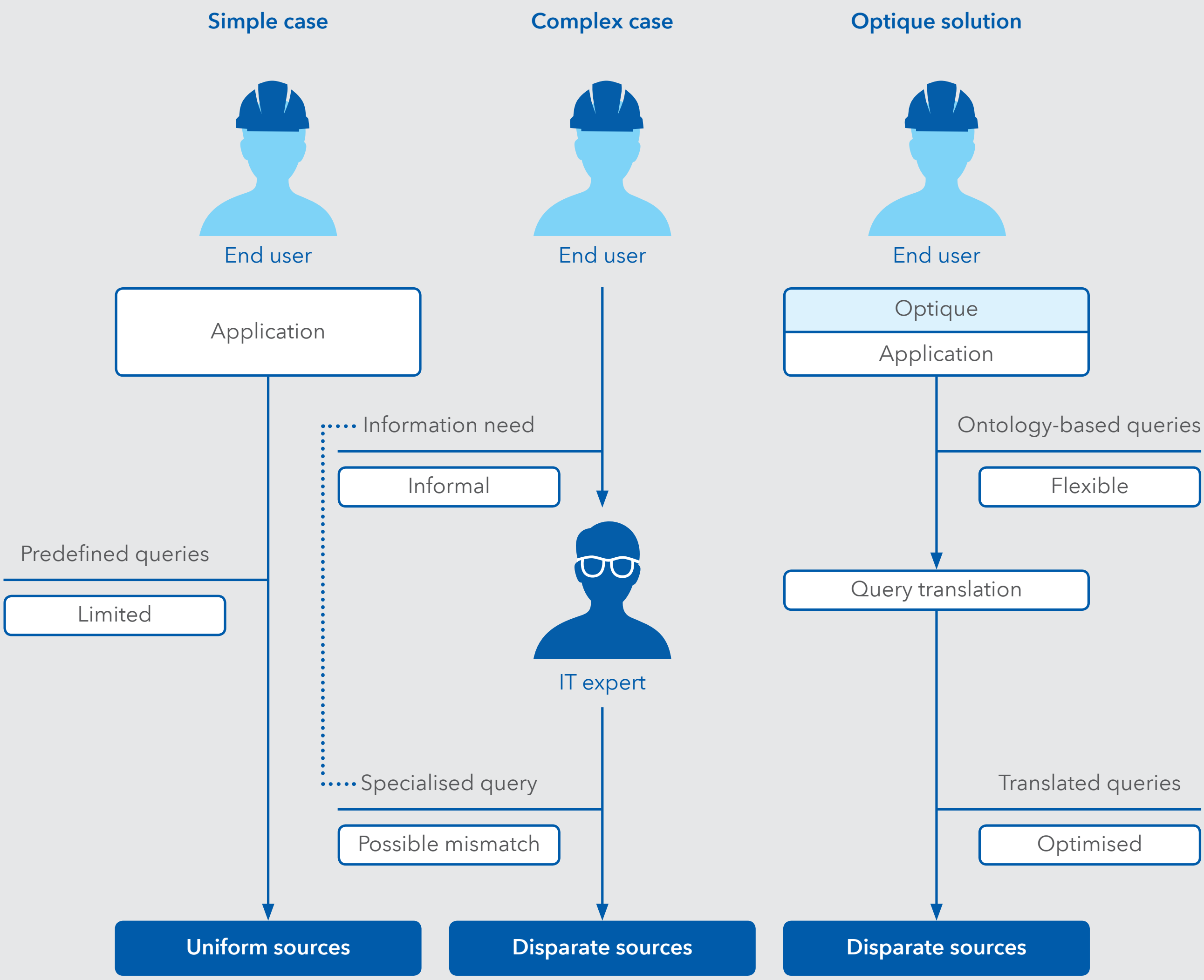


Figure 1: Common data access scenarios in enterprise context and the Optique approach

Simple case

Most users cannot write questions (queries) in special ways (structured languages) to access relevant data quickly and in required formats. A 'simple data access' model offers a limited range of set questions and information types to pull data from various databases that are equally easy or hard (uniform) to connect to and control.

Complex case

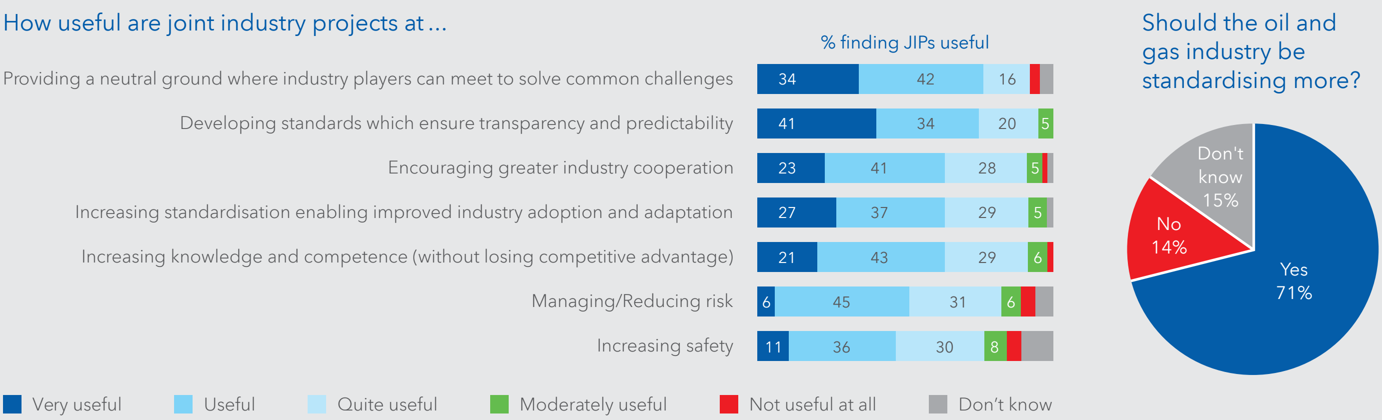
The complex case is a 'man-in-the-middle' approach where users send information needs to IT experts who in turn write more sophisticated queries. This finds the right information and presents it in ways that are useful for the purposes involved, but limited IT staff numbers mean it can take days to weeks for users to get it back.

Optique solution

The Optique JIP exploits advances in semantic technologies that can explore meaning and context behind words and sentences. The goal is software that allows people to use computer language that they can understand so they can get flexible, comprehensive, and timely access to large, complex industrial data sets.

Source: Optique JIP

Joint industry projects: What do participants get out of them?



Source: 65 responses to TNS Gallup AS survey (2014) of 67 participants in DNV GL-led JIPs

information requests from days to minutes, while also advancing to data sets whose size and complexity is beyond the reach of existing technologies.

The big picture
Waaler believes that the majority of current solutions for Big Data focus solely on volume and processing large amounts quickly. The Optique project adds another dimension to the three Vs: complexity.

"Traditional technologies are extremely good at volume, but compromise a lot on variety, velocity and complexity," he said. "Optique is unique in focusing on all these dimensions simultaneously. It also addresses trustworthiness by showing where data came from and how it has changed, providing transparency for the end user."

Take the variety aspect for instance: "Statoil has hundreds of terabytes of stratigraphy and seismic interpretation data that needs analysis in large and very complex databases. You cannot do this with only the methods developed for big volumes of data, but it is a main goal for Optique. We focus on variety, velocity and complexity, then consume as much data as we can without compromising too much of the other dimensions."

Optique aims to test and implement a long-term solution for data access that creates a tool for end users to find data on their own, which they cannot do now.

Waaler explained: "Geologists and engineers know what they need, but the problem is posting a complex query to multiple databases. This is impossible without sending a request to IT experts, a scarce resource. End users must wait for these experts to create complex queries. This may take up to several weeks and considerably delays decision-making."

Optique plans to provide tools to allow a user to query data without assistance from IT experts, and get the result in minutes, he said. "This will open up new exploratory and interactive ways of working as users get more relevant data sets in shorter time. We see Optique as the central tool for exploring information and returning timely, complete, and accurate results. Users can then focus fully on what they are trained in."

A challenge to industry
The Optique solution has been tried and tested in the laboratory. The next step is to implement it within the industry, and DNV GL has taken on the role of bridge builder between the theoretical and practical worlds. Waaler said remaining challenges include speeding up the performance of the back end by applying massively parallelised solutions and also tools to ease establishing and maintaining installations of the Optique platform.

In early 2015, the Optique team plans to present current results at a conference in Høvik, Norway. The aim is to recruit interested companies as partners to the project. The vision is that by 2020, Optique methods and technology will be incorporated into mainstream information management products delivered by trusted vendors.

"We will deliver a good concept, but this will not be something that can be delivered to the industry two years from now. I hope that by then we have something so impressive that the industry will want to continue to fund this project. I am optimistic," Waaler said. ■

For further details about Optique contact: Tore Hartvigsen, DNV GL project manager torehartvigsen@dnvgl.com

1 Laney, D: '3D Data Management: Controlling data volume, velocity, and variety'; Meta Group (now Gartner) (2001)

Appendix C

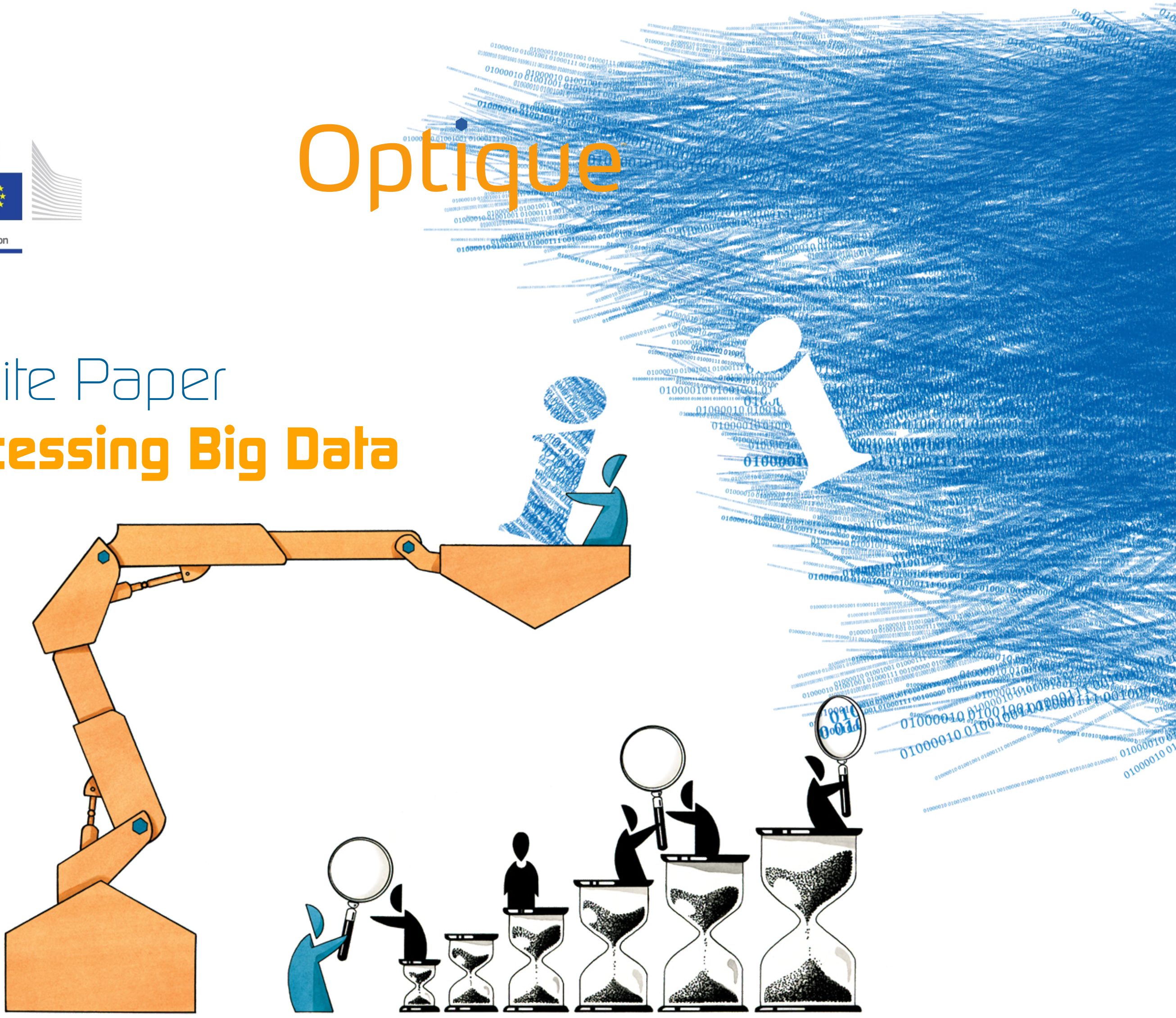
Executive Whitepaper

To provide information about Optique to enterprise executives and decision makers an Executive Whitepaper has been prepared. The paper introduce the Optique idea, products and project to readers without necessarily a profound IT technical background but with the interest to investigate how accessing Big data can be a challenge to an enterprise and that Optique can be a remedy to the challenge of Big Data Access. A series of White papers is planned produced in Optique, this being the first.

Optique

White Paper

Accessing Big Data



We are surrounded by vast and steadily increasing amounts of data, be it in industrial or personal contexts. In order to maintain a leading position in industry, one soon finds that it is paramount to not only handle this explosion of information effectively, but also to be able to access important data at a moment's notice.

Optique exploits recent advances in semantic technologies in order to develop a software platform that can provide end-users with flexible, comprehensive, and timely access to large and complex industrial data sets. It lets end-users access the data they need with the language they use. The platform reduces turnaround time for information requests from days to minutes while at the same time scaling up to datasets whose size and complexity is beyond reach of existing technologies.

This paper informs decision makers on the Big Data challenges faced by industry today. It introduces Optique, a joint project consisting of world leading academic institutions and industry partners that will address and propose a solution for these challenges.

Currently, access to Big Data is severely hampered; the data is stored in ways that are unintuitive and not transparent to the end-user. Optique addresses this by implementing and expanding on ground-breaking European research on semantic technologies. In this document we discuss the project objectives and the manner in which industry partners can join and influence the project.

This is the first paper in a series of white papers, introductory courses, and workshops organisd by the Optique project on current Big Data challenges.

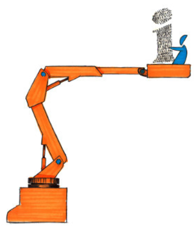
This paper was financed by the Optique Project with the grant agreement FP7-318338.

Improving Decision Making

Large amounts of data have been accumulated by industries over the last decades from a large variety of sources and in many different formats. With the help of this data, end-users are theoretically capable of making well-informed decisions that have a strong impact on the business, e.g., from a fiscal or safety point of view. For example, Statoil, the leading Norwegian Oil & Gas company, continuously searches for new oil and gas reservoirs through explorative drilling. This is, however, very expensive, and copious amounts of data need to be considered and analyzed: stratigraphy, seismic interpretation, production logs, maps, models, infrastructure, business information, etc.

In a perfect world, such decisions are made in a timely fashion, taking all relevant information into account. Unfortunately, this is not always that simple. The amount of data is increasing, leading to an *information overload*. In the case of Statoil, the required database queries can include processing TBs of data. While having as much information as possible can seem beneficial at first glance, it has some clear disadvantages: increased data access time and reduced actual relevancy of query results.

Data is stored in various heterogeneous formats over many differently structured databases. As a result, the gathering of *only relevant data* spread over disparate sources becomes a very time-consuming task. Often requiring the assistance of designated IT personnel, it can take weeks to retrieve the desired information. An estimated 30–70% of end-users' time is being allocated solely to retrieving data (Crompton, 2008). However, decisions must often be made quickly in response to current developments or as a cost saving measure. This leads to a bottleneck in the decision making process, where decisions do not take all relevant information into account or expensive expert time is wasted on waiting for necessary data.



Optique aims at *removing the bottleneck* in the decision making process by enabling the end-users to get the information they need without requiring any knowledge on how the data is stored, thus removing the need for an IT middle man. This is done by using and expanding on existing tools from semantic technologies, an approach that has the potential to reduce the data access time by several orders of magnitude (e.g., minutes instead of days).

Big Data

A common definition of *Big Data* focuses on the technical limitations of currently used computer systems and software. A *Big Data challenge* occurs when the current (standard) toolset limits what you can do with the data. Use of new hardware, software, or algorithms is required to solve the problem. Managing and using data have always been challenging; but when does data become Big Data? Big Data is often characterised by Volume, Velocity, and Variety. These "3 Vs" where originally coined by Dough Laney (Laney, 2001). Optique extends this definition with another dimension, Complexity.

"Big Data will deliver transformational benefits to enterprises within 2 to 5 years, and by 2015 will enable enterprises adopting this technology to outperform competitors by 20% in every available financial metric."–Gartner, 2012

Volume: For many enterprises massive amounts of data have been accumulated over decades. Disparate data systems have been used to solve specific needs and integration between systems has received little attention. For many enterprises this poses a technical challenge in terms of storing, finding, combining and analyzing the data.

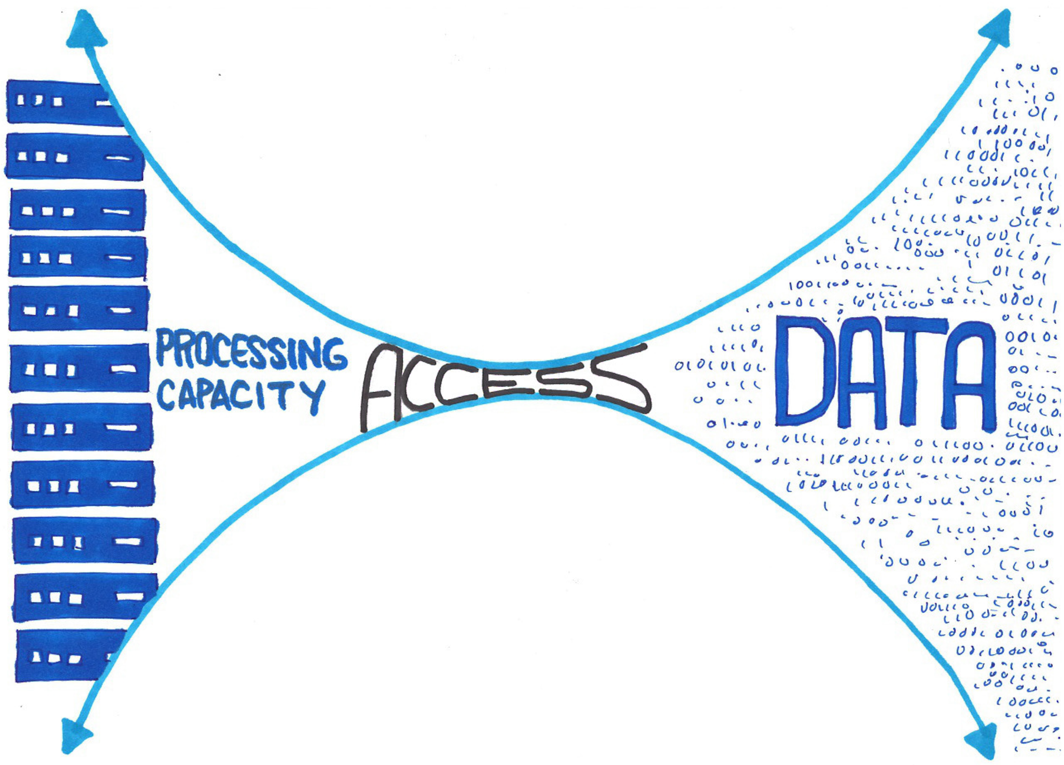
Velocity: For some problems, making the right decision in time is most important. Being informed about and being able to react to unpredictable behavior can reduce the risk for accidents. Tackling such problems in real-time poses additional requirements.

Variety: Data can be of any type; it can be structured or unstructured; it can be text, time series from sensors, click streams and log files from web sites, multimedia, position data from GPS loggers, etc. The challenge is to combine the various data, to set up the system, and to integrate new data as it becomes available.

Complexity: As more information is gathered and stored, the structure of existing data stores becomes higly complex. This results in databases that require in-depth knowledge of their schemata in order to pose queries. As a result, considerable time must be invested in understanding the structure before the data can be accessed.

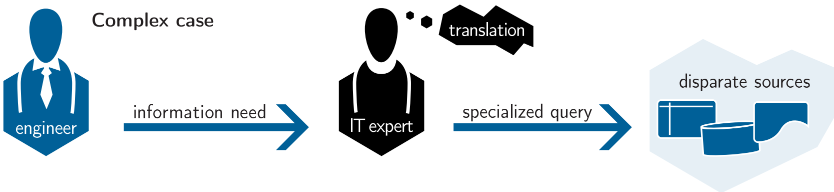
In order to be of substantial value for business, a Big Data solution must take all of these into account. This is however a difficult task, as optimizing in one dimension will invariably have an effect on the others. It is therefore important to find the right balance.

In addition to the above mentioned dimensions, other notions have been used to characterize Big Data as well. It is tempting to believe that a large amount of data will deliver the desired results. However this is grossly overly optimistic. It is therefore important to determine the *trustworthiness* of the data. Optique addresses this in particular by tracking the data's *provenance*, i.e., informing the end-user where the data came from and what transformations it has undergone. This ensures transparency between the data and the end-user.



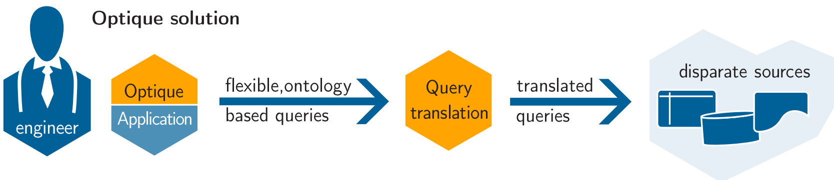
Optique: Intelligent Databases

Currently, accessing information in legacy data stores requires considerable background knowledge on the manner in which the data is stored. Hence, end-users are heavily reliant on predefined queries or IT experts to formulate and optimise special purpose queries. However, IT experts who have both in-depth knowledge of the databases and enough domain knowledge to understand the end-users are a scarce resource. As a consequence, end-users must wait for IT experts to create complex queries, possibly involving a number of complex joins between tables. This may take up to several weeks and considerably delays the decision making processes in the enterprise. End-users have limited access to tools for *exploring* the data themselves.



Optique's objective is to achieve a combination of intelligence, flexibility, and scalability in order to meet the end-users' requirements. From a front-end perspective, this is achieved through new and intuitive query formulation tools. These help end-users formulate queries without involving IT experts. Through Optique's support of streaming data and geospatial and temporal reasoning, end-users are able to formulate queries not supported by current systems. In particular, these changes give experts the opportunity to simply explore the data in ways that were previously not possible.

The ultimate objective is to implement an end-to-end solution that provides comprehensive and timely access to large scale data sets.



By not requiring in-depth knowledge of the databases' structure, the end-users can formulate their own queries using only familiar vocabulary and concepts. This provides end-users the ability to:

- **Find data** by identifying data sets that are relevant to a question at hand through direct access to the semantics of the data, and by supporting exploration of the data through *ad hoc* queries.
- **Reuse data** through a powerful and flexible query language along with support for coherent query formulation; as well as linking disparate sources to a common vocabulary;
- **Explore data** by connecting end-users directly to the data through Optique's end-to-end solution; the dramatic increase in data retrieval efficiency yields higher quality data exploration, and thus better data exploitation as well.
- **Exploit data in different, distant and unforeseen contexts** through support for building different conceptualizations into ontologies and mappings, thus allowing end-users to explore data from their own perspective, independently of the original purpose and organisation of the data.

Optique and Current Solutions

What is the difference between the Optique platform and the technology currently in use by industry, and what is the advantage of moving from legacy systems to the technology provided by Optique?

Data warehouses are a good example for current solutions: they have been in use for decades and most large corporations have significant IT investments in such systems.

Typically a data warehouse integrates data from disparate data sources with the purpose of creating a central data repository optimized for senior management reporting and data analysis. The data stored in the warehouse is uploaded from operational systems. During upload, the data might be cleansed or transformed according to defined business rules.

The benefits of a data warehouse are many: it enables maintenance of data history, improves data quality, and enables an organization to present its information consistently across several data sources.

But there are also limitations. In a data warehouse setup, each data source needs to be mapped to the data warehouse. Some of the mappings can be highly complex, especially when the nature of the sources differs a lot. When the number of data sources increases, a significant effort has to be made to keep the mappings consistent with the underlying data models. This is a bottleneck when bringing data warehouse architecture into the world of Big Data, and is a bottleneck that the Optique project addresses and has the ambition to loosen. In addition, data warehouses are based on traditional relational structures, while much of the data generated today is unstructured and does not fit neatly into traditional, structured relational models. Furthermore, data volume and velocity become a problem since new data must be extracted frequently from the source databases and loaded into the data warehouse. Optique addresses these issues by accessing the source database directly and by supporting the processing of data streams.

Optique Partner Programme

In order to further the collaboration between academia and industry, Optique has launched the European Partner Programme. This enables industry to not only stay informed, but also influence new technologies and scientific research, while ensuring the continued collaboration between academia and industry. Optique therefore offers three different levels of industry partnership:

Network Partner

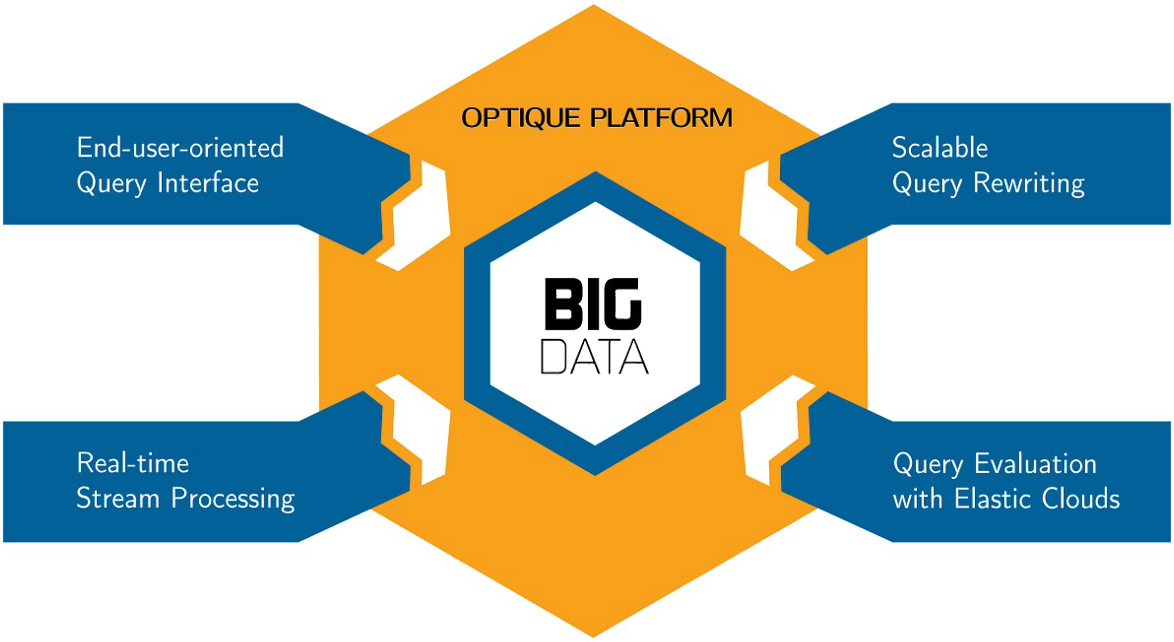
The **Network Partner** level is suitable for companies that primarily want information and networking opportunities with industry and academia.

Discussion Partner

The **Discussion Partner** level also includes an analysis of potential usage areas for the company and a dedicated contact person within the Optique project.

Pilot Partner

The **Pilot Partner** level additionally includes unlimited access to the Optique system for the duration of the membership, as well as access to Optique experts to implement and customize the Optique System based on the company's usage areas.



About Optique

Optique is a large-scale integrating EU project (IP) under FP7-ICT Objective 4.4 Intelligent Information Management. The program runs over 4 years (Start December 2012) with a total budget of EUR 13,800,000 (whereof the EU contribution is EUR 9,760,000). Optique will deliver a comprehensive program of research and technological development bringing together leading experts from European academia and industry. Optique will exploit recent ground-breaking European research on ontology based data access and scalable query execution to develop an extensible platform. Built on open standards and protocols, it will integrate effortlessly with existing systems to provide a complete and generic solution to the data access challenges posed by Big Data. The program will be based on use cases and data from energy sector (Siemens) and stratigraphy data from the Oil and Gas sector (Statoil).

Deliverables

In addition to research publications and reports, the main deliverable of the project will be an extensible platform prototype. This will serve as a basis for a complete and generic solution to the data access challenges posed by Big Data. The individual components will be available as open source, while an optimized and integrated solution will be available as a commercial product. The platform will include tools to support, e.g., query formulation and ontology and mapping management.

Development of the platform will be continuously evaluated against the requirements of the two use cases from Siemens and Statoil. Experience from the use case deployments will be used to develop high quality tutoring and training resources.

Optique Partners

Industry

Siemens AG is a globally operating technology company with core activities in the fields of energy, healthcare, industry, and infrastructure. They will contribute with use cases and big data from remote monitoring and diagnostics of gas/steam turbines, generators and compressors used, e.g., in power plants.

Statoil ASA is a leading Norwegian Oil & Gas company and will contribute with use cases and stratigraphy data for its exploration activities.

Fluid Operations AG is a leading provider of innovative cloud and data management solutions based on semantic technologies. They will be responsible for the development of the core Optique platform and will take lead in commercializing the software results.

DNV GL provides classification, technical assurance, software, and advisory services to the maritime, Oil & Gas, and energy sectors, in addition to certification services across a wide range of industries. They are responsible for information dissemination and will work together with the project partners on exploitation of project results.

Academia

University of Oslo coordinates the Optique program. They will in particular focus on developing the use cases and demonstrators from the Oil and Gas sector on stratigraphic data (Statoil use cases).

Oxford University will provide overall scientific coordination for the project and lead the work with the query formulation component.

Hamburg University of Technology will develop the use case demonstratrs for the energy sector (Siemens use cases) and the methodology and tools to deal with scalable query answering on real-time processing of stream-based data.

Sapienza University of Rome will deliver the ontology and mapping management component of the Optique system.

Free University of Bozen-Bolzano will deliver the query transformation subsystem.

National and Kapodistrian University of Athens will lead the work on distributed query execution. It will also lead the data mining and query log analysis activities for improving the performance of temporal query answering.

Project Advisory Board

The Project Advisory Board consists of representatives from the following companies:

Contact

Prof. Arild Waaler (UiO)
Project Coordinator
arild@ifi.uio.no
+47 22840055

Prof. Ian Horrocks (UOXF)
Scientific Director
ian.horrocks@ics.ox.ac.uk
+44 1865 273939

Paper prepared by:

Daniel P. Lupp (UiO), Tore Hartvigsen (DNV GL), and Pål Rylandsholm (DNV GL)

Bibliography

Crompton, J. 2008. *Putting the Focus on Data*. Keynote talk, W3C Workshop on Semantic Web in Oil & Gas Industry: Houston, TX, USA, available at http://www.w3.org/2008/12/ogws-slides/Crompton.pdf

Laney, D. 2001. *3D Data Management: Controlling Data Volume, Velocity, and Variety*. Gartner.