DEPARTMENT OF
COMPUTER
SCIENCE

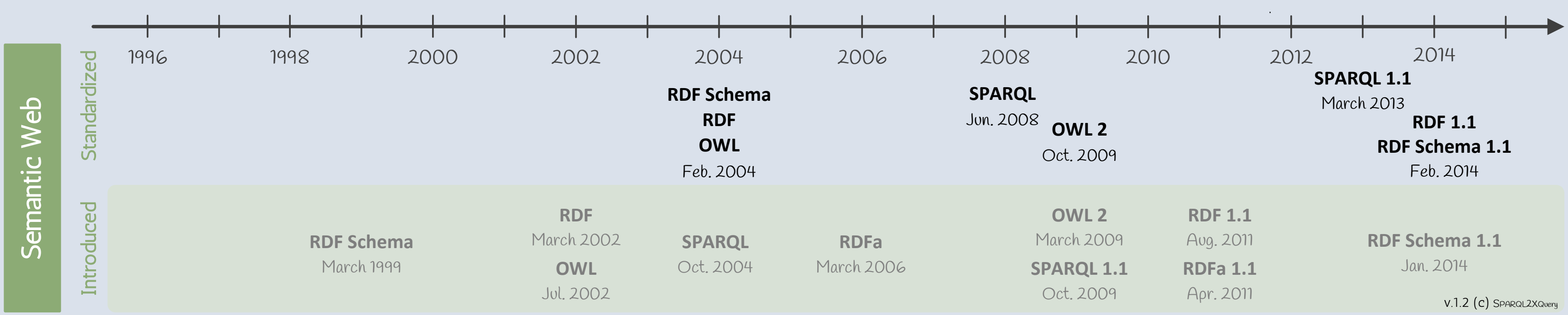
UNIVERSITY OF
OXFORD

Build it, and they will come: Applications of semantic technology

Ian Horrocks
Information Systems Group
Department of Computer Science
University of Oxford

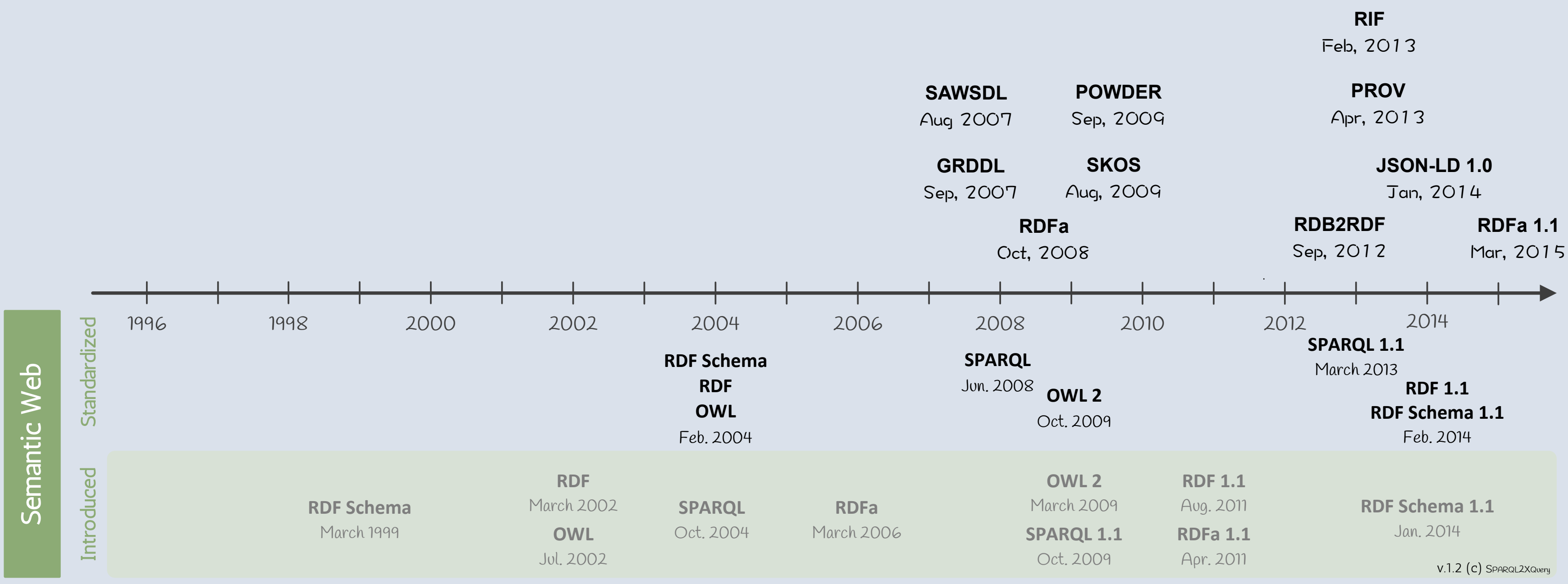
A Brief History of the Semantic Web

A Brief History of the Semantic Web



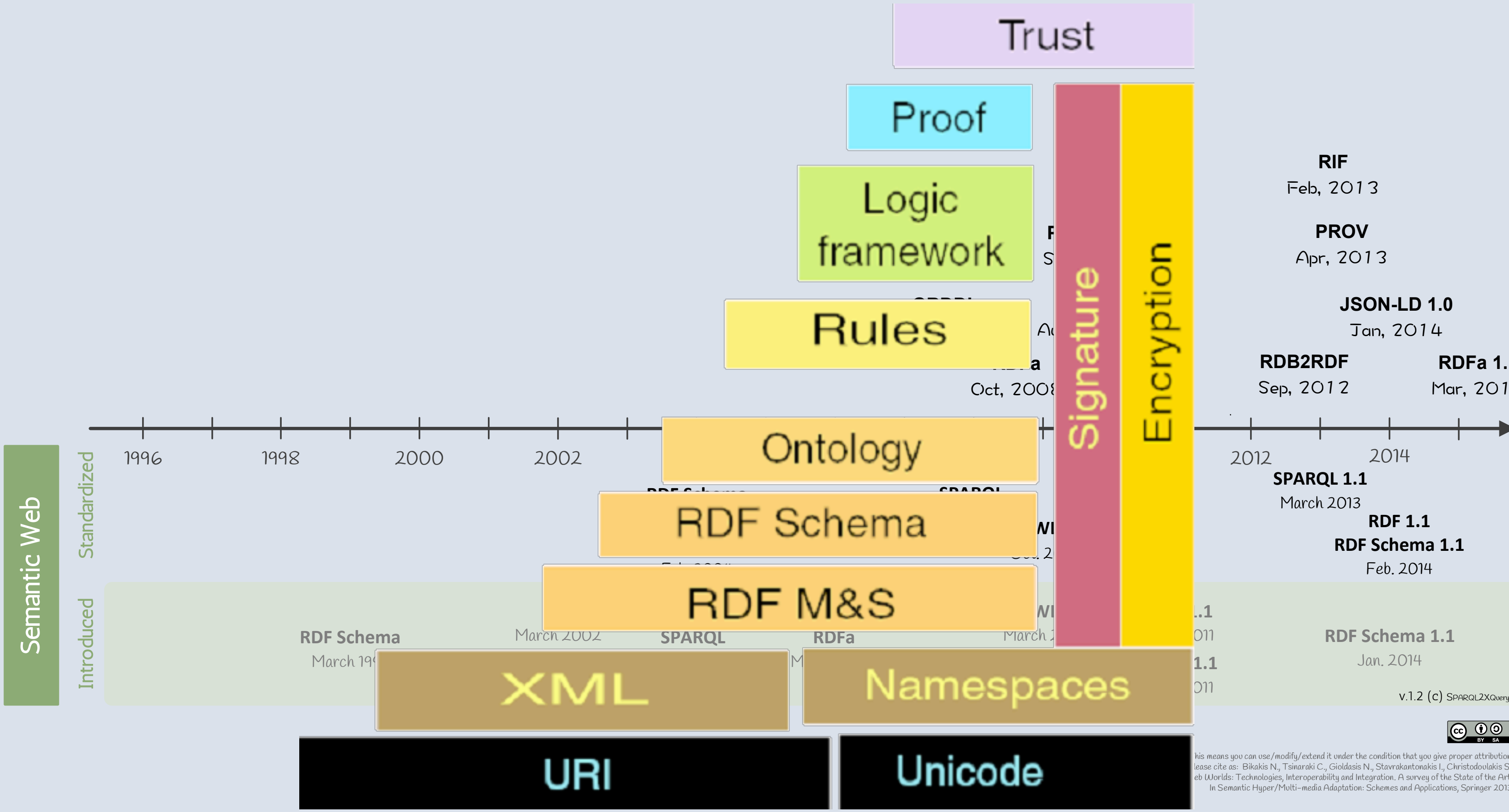
UNIVERSITY OF OXFORD DEPARTMENT OF COMPUTER SCIENCE
Information Systems Group
EPSRC Engineering and Physical Sciences Research Council
Optique

A Brief History of the Semantic Web



UNIVERSITY OF OXFORD DEPARTMENT OF COMPUTER SCIENCE
Information Systems Group
EPSRC Engineering and Physical Sciences Research Council
Optique

A Brief History of the Semantic Web



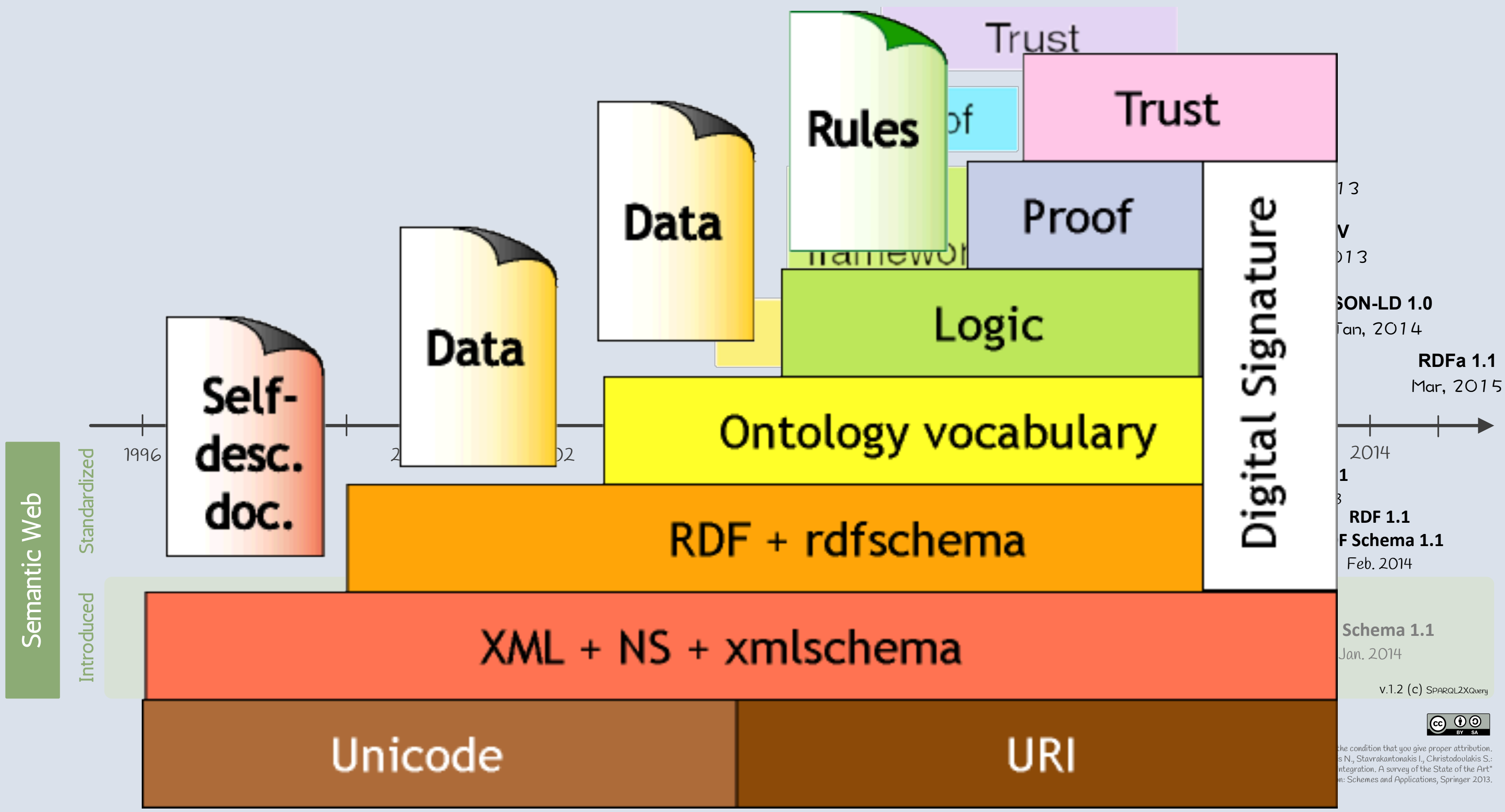
University of Oxford Department of Computer Science

Information Systems Group

EPSRC Engineering and Physical Sciences Research Council

Optique

A Brief History of the Semantic Web



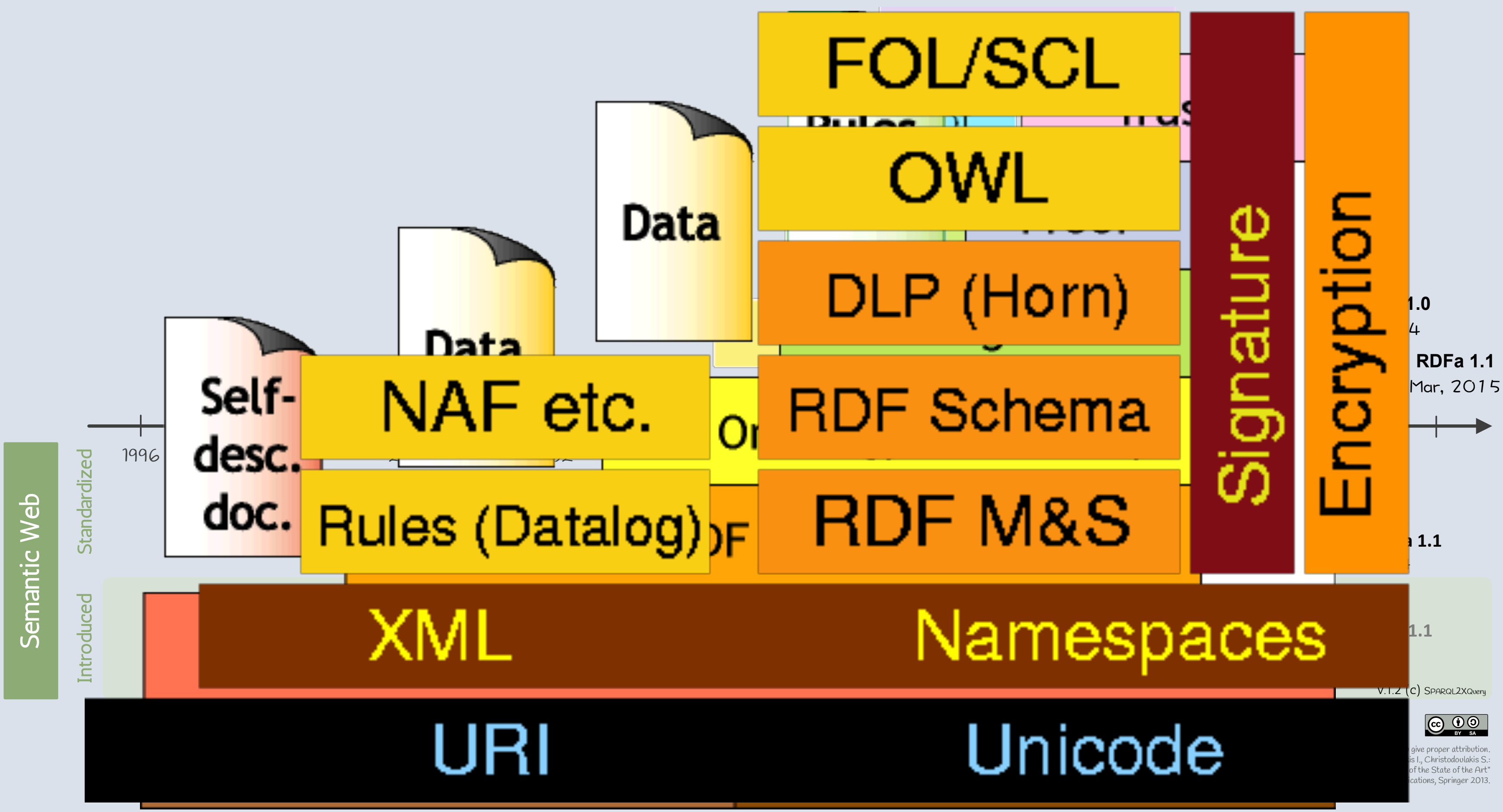
UNIVERSITY OF OXFORD DEPARTMENT OF COMPUTER SCIENCE

Information Systems Group

EPSRC Engineering and Physical Sciences Research Council

Optique

A Brief History of the Semantic Web



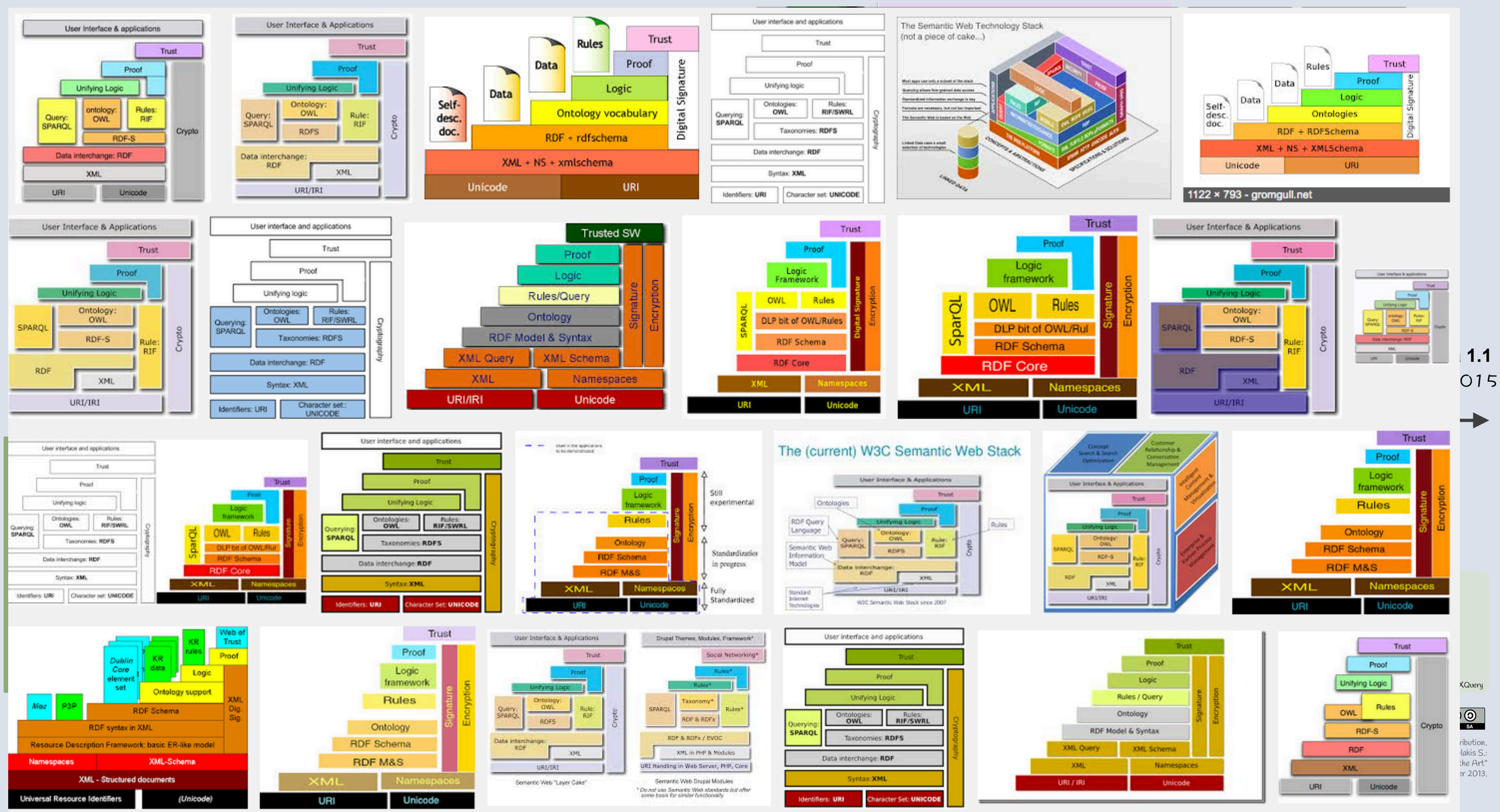
DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF OXFORD

Information Systems Group

EPSRC
Engineering and Physical Sciences Research Council

Optique

A Brief History of the Semantic Web



DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF OXFORD
Information Systems Group
EPSRC
Engineering and Physical Sciences Research Council
Optique

A Brief History of the Semantic Web

UNIVERSITY OF OXFORD DEPARTMENT OF COMPUTER SCIENCE

Information Systems Group

EPSRC Engineering and Physical Sciences Research Council

Optique

A Brief History of the Semantic Web

A Brief History of the Semantic Web

UNIVERSITY OF OXFORD DEPARTMENT OF COMPUTER SCIENCE

Information Systems Group

EPSRC Engineering and Physical Sciences Research Council

Optique

Why the Angst?

DEPARTMENT OF COMPUTER SCIENCE UNIVERSITY OF OXFORD

Information Systems Group

EPSRC Engineering and Physical Sciences Research Council

Optique

Why the Angst?

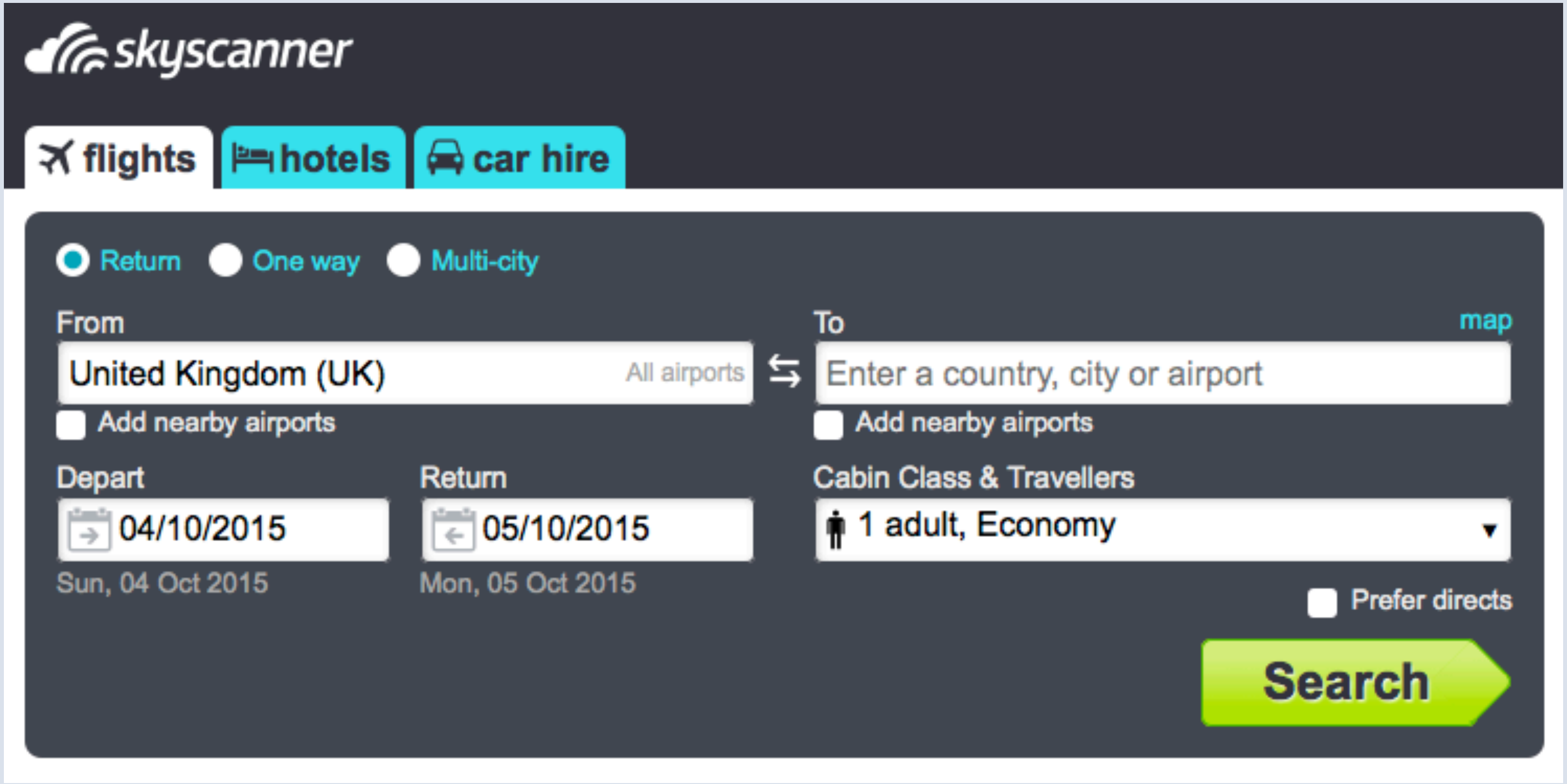
Why the Angst?

“A new form of Web content that is meaningful to computers will **unleash a revolution of new possibilities**”

Why the Angst?

Semantic **Web**

Why the Angst?



Why the Angst?

Google

Why the Angst?

Google

- Explicit KR sometimes needed, e.g., **Knowledge Graph**
  - Less rigorous treatment of semantics
  - Not using Semantic Web standards

Why the Angst?

Google

- Explicit KR sometimes needed, e.g., **Knowledge Graph**
  - Less rigorous treatment of semantics
  - Not using Semantic Web standards
- **Hiring** Semantic Web people

UNIVERSITY OF OXFORD
DEPARTMENT OF COMPUTER SCIENCE
Information Systems Group
EPSRC
Engineering and Physical Sciences Research Council
Optique

Why the Angst?

Semantic Web

UNIVERSITY OF OXFORD
DEPARTMENT OF COMPUTER SCIENCE
Information Systems Group
EPSRC
Engineering and Physical Sciences Research Council
Optique

Data Access:

Data Access: Statoil Exploration

- Geologists & geophysicists use data from previous operations in nearby locations to develop stratigraphic models of unexplored areas
  - TBs of relational data
  - using diverse schemata
  - spread over 1,000s of tables
  - and multiple data bases

Data Access: Statoil Exploration

- Geologists & geophysicists use data from previous operations in nearby locations to develop stratigraphic models of unexplored areas
  - TBs of relational data
  - using diverse schemata
  - spread over 1,000s of tables
  - and multiple data bases

Data Access

- 900 geologists & geophysicists
- 30-70% of time on data gathering
- 4 day turnaround for new queries

Data Access: Statoil Exploration

- Geologists & geophysicists use data from previous operations in nearby locations to develop stratigraphic models of unexplored areas
  - TBs of relational data
  - using diverse schemata
  - spread over 1,000s of tables
  - and multiple data bases

Data Access

- 900 geologists & geophysicists
- 30-70% of time on data gathering
- 4 day turnaround for new queries

Data Exploitation

- Better use of experts time
- Data analysis "most important factor" for drilling success

Data Access: SIEMENS Energy Services

- Service centres responsible for remote monitoring and diagnostics of 1,000s of gas/steam turbines
- Engineers use a variety of data for visualization, diagnostics and trend detection:
    - several TB of time-stamped sensor data
    - several GB of event data
    - data grows at 30GB per day

Service Requests

- 1,000 requests per center per year
- 80% of time used on data gathering

Diagnostic Functionality

- 2–6 p/m to add new function
- New diagnostics → better exploitation of data


Pipelines from
oil facilities?

$$
\begin{aligned}
Q(?x) \leftarrow\ & (?x, \textsf{rdf:type}, \textsf{:Pipeline}) \wedge \\
& (?x, \textsf{:fromFacility}, ?y) \wedge \\
& (?y, \textsf{rdf:type}, \textsf{:OilFacility})
\end{aligned}
$$

$$Q(?x) \leftarrow (?x, \text{rdf:type}, \text{:Pipeline}) \wedge (?x, \text{:fromFacility}, ?y) \wedge (?y, \text{rdf:type}, \text{:OilFacility})$$

(:p1, rdf:type, :Pipeline)
(:p1, :fromFacility, :f1)
(:f1, rdf:type, :OilFacility)
(:p2, rdf:type, :OilPipeline)
(:p2, :fromFacility, :f2)
(:f2, rdf:type, :OilFacility)
(:p3, rdf:type, :OilPipeline)

$$Q(?x) \leftarrow (?x, \text{rdf:type}, \text{:Pipeline}) \wedge (?x, \text{:fromFacility}, ?y) \wedge (?y, \text{rdf:type}, \text{:OilFacility})$$

(:p1, rdf:type, :Pipeline)
(:p1, :fromFacility, :f1)
(:f1, rdf:type, :OilFacility)
(:p2, rdf:type, :OilPipeline)
(:p2, :fromFacility, :f2)
(:f2, rdf:type, :OilFacility)
(:p3, rdf:type, :OilPipeline)

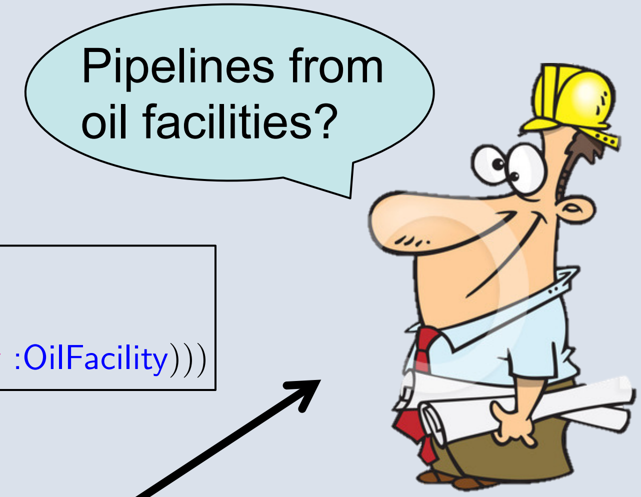
$$Q(?x) \leftarrow (?x, \text{rdf:type}, \text{:Pipeline}) \wedge (?x, \text{:fromFacility}, ?y) \wedge (?y, \text{rdf:type}, \text{:OilFacility})$$

(:p1, rdf:type, :Pipeline)
(:p1, :fromFacility, :f1)
(:f1, rdf:type, :OilFacility)
(:p2, rdf:type, :OilPipeline)
(:p2, :fromFacility, :f2)
(:f2, rdf:type, :OilFacility)
(:p3, rdf:type, :OilPipeline)

:p1

DEPARTMENT OF COMPUTER SCIENCE, UNIVERSITY OF OXFORD
Information Systems Group
EPSRC Engineering and Physical Sciences Research Council
Optique

$$
\begin{aligned}
Q(?x) \leftarrow\ & (?x, \text{rdf:type}, \text{:Pipeline}) \wedge \\
& (?x, \text{:fromFacility}, ?y) \wedge \\
& (?y, \text{rdf:type}, \text{:OilFacility})
\end{aligned}
$$



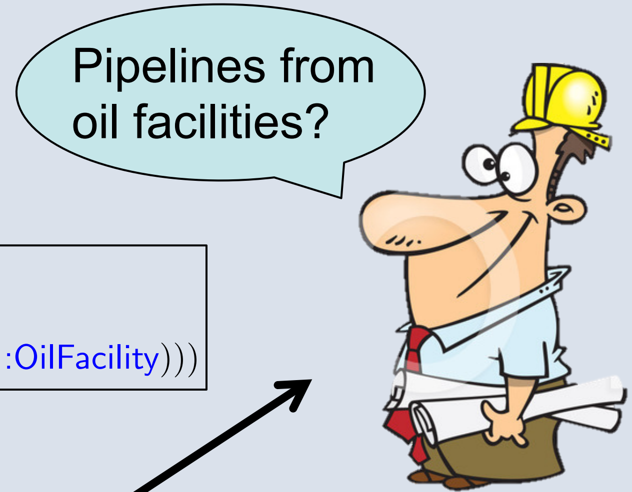

```
SubClassOf(:OilPipeline
  ObjectIntersectionOf(:Pipeline
    ObjectSomeValuesFrom(:fromFacility :OilFacility)))
```

(:p1, rdf:type, :Pipeline)
(:p1, :fromFacility, :f1)
(:f1, rdf:type, :OilFacility)
(:p2, rdf:type, :OilPipeline)
(:p2, :fromFacility, :f2)
(:f2, rdf:type, :OilFacility)
(:p3, rdf:type, :OilPipeline)

:p1

UNIVERSITY OF OXFORD DEPARTMENT OF COMPUTER SCIENCE
Information Systems Group
EPSRC Engineering and Physical Sciences Research Council
Optique

$$Q(?x) \leftarrow (?x, \text{rdf:type}, \text{:Pipeline}) \wedge (?x, \text{:fromFacility}, ?y) \wedge (?y, \text{rdf:type}, \text{:OilFacility})$$



SubClassOf(:OilPipeline
  ObjectIntersectionOf(:Pipeline
    ObjectSomeValuesFrom(:fromFacility :OilFacility)))

(:p1, rdf:type, :Pipeline)
(:p1, :fromFacility, :f1)
(:f1, rdf:type, :OilFacility)
(:p2, rdf:type, :OilPipeline)
(:p2, :fromFacility, :f2)
(:f2, rdf:type, :OilFacility)
(:p3, rdf:type, :OilPipeline)

:p1, :p2, :p3

UNIVERSITY OF OXFORD DEPARTMENT OF COMPUTER SCIENCE
Information Systems Group
EPSRC Engineering and Physical Sciences Research Council
Optique

$$Q(?x) \leftarrow (?x, \text{rdf:type}, \text{:Pipeline}) \land (?x, \text{:fromFacility}, ?y) \land (?y, \text{rdf:type}, \text{:OilFacility})$$

Pipelines from oil facilities?

```
SubClassOf(:OilPipeline
  ObjectIntersectionOf(:Pipeline
    ObjectSomeValuesFrom(:fromFacility :OilFacility)))
```

OWL 2 QL ontology

(:p1, rdf:type, :Pipeline)
(:p1, :fromFacility, :f1)
(:f1, rdf:type, :OilFacility)
(:p2, rdf:type, :OilPipeline)
(:p2, :fromFacility, :f2)
(:f2, rdf:type, :OilFacility)
(:p3, rdf:type, :OilPipeline)

:p1, :p2, :p3

$$Q(?x) \leftarrow (?x, \text{rdf:type}, \text{:Pipeline}) \wedge (?x, \text{:fromFacility}, ?y) \wedge (?y, \text{rdf:type}, \text{:OilFacility})$$

rewrite

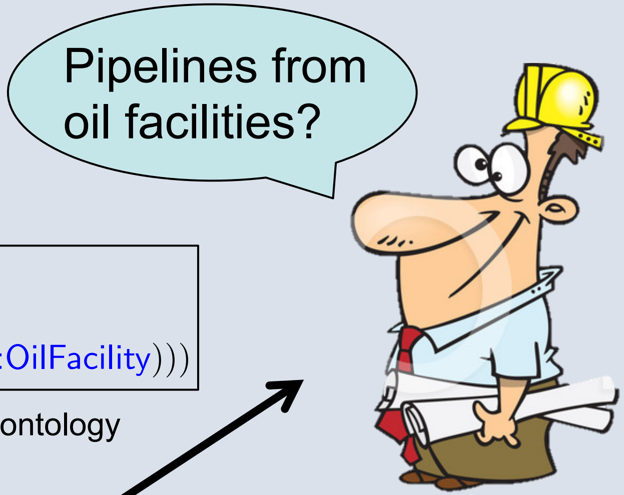
SubClassOf(:OilPipeline
  ObjectIntersectionOf(:Pipeline
    ObjectSomeValuesFrom(:fromFacility :OilFacility)))

OWL 2 QL ontology

$$Q'(?x) \leftarrow (?x, \text{rdf:type}, \text{:Pipeline}) \wedge (?x, \text{:fromFacility}, ?y) \wedge (?y, \text{rdf:type}, \text{:OilFacility}) \vee (?x, \text{rdf:type}, \text{:OilPipeline})$$

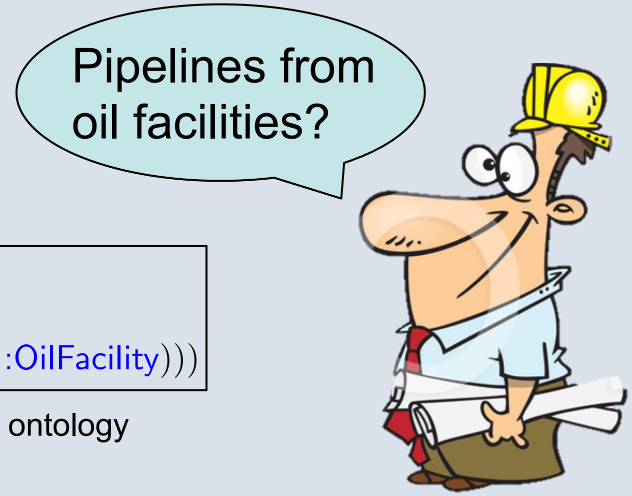
(:p1, rdf:type, :Pipeline)
(:p1, :fromFacility, :f1)
(:f1, rdf:type, :OilFacility)
(:p2, rdf:type, :OilPipeline)
(:p2, :fromFacility, :f2)
(:f2, rdf:type, :OilFacility)
(:p3, rdf:type, :OilPipeline)

:p1, :p2, :p3



University of Oxford Department of Computer Science
Information Systems Group
EPSRC Engineering and Physical Sciences Research Council
Optique

$$Q(?x) \leftarrow (?x, \text{rdf:type}, \text{:Pipeline}) \wedge (?x, \text{:fromFacility}, ?y) \wedge (?y, \text{rdf:type}, \text{:OilFacility})$$



rewrite

```
SubClassOf(:OilPipeline
  ObjectIntersectionOf(:Pipeline
    ObjectSomeValuesFrom(:fromFacility :OilFacility)))
```

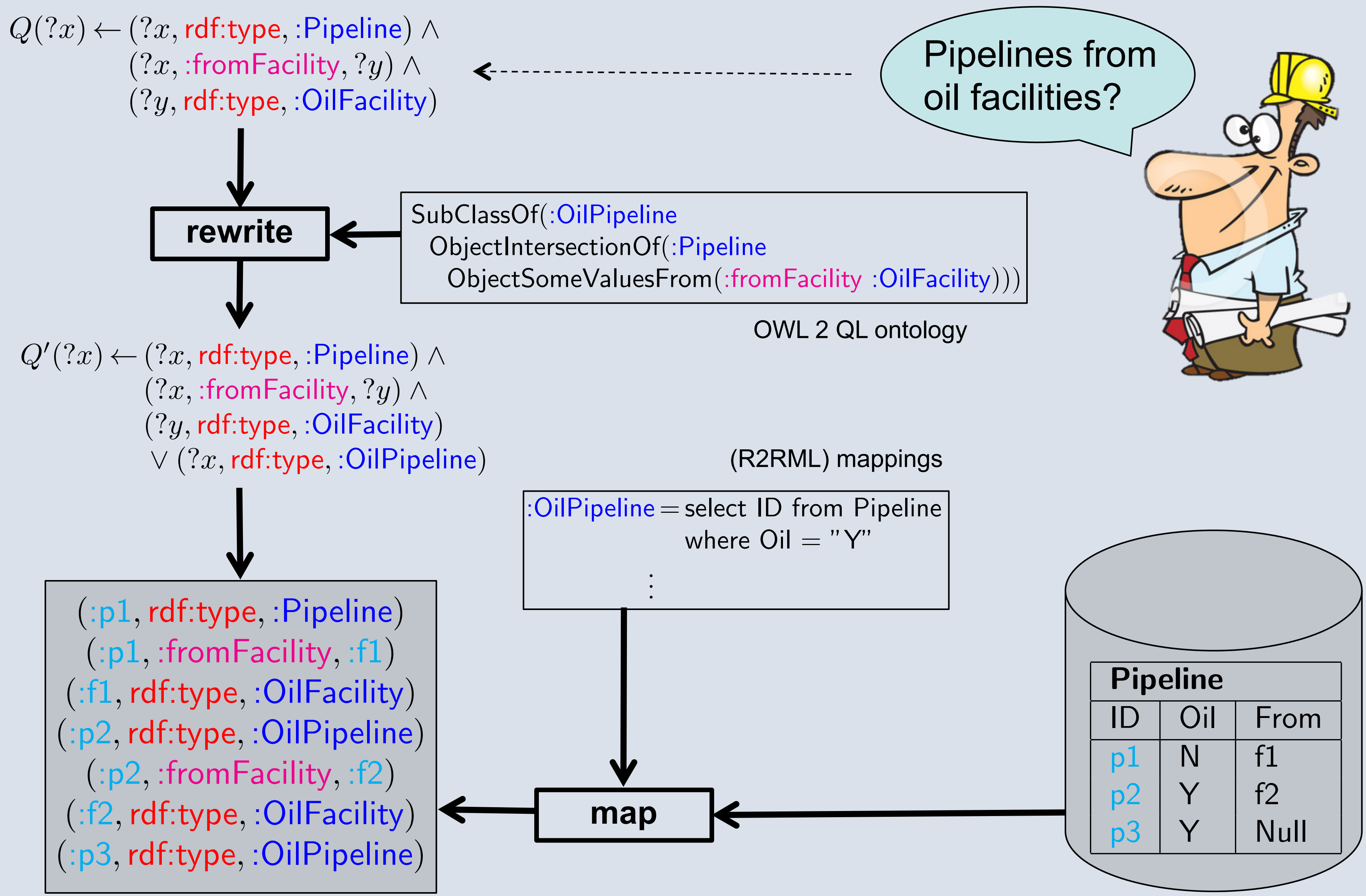
OWL 2 QL ontology

$$Q'(?x) \leftarrow (?x, \text{rdf:type}, \text{:Pipeline}) \wedge (?x, \text{:fromFacility}, ?y) \wedge (?y, \text{rdf:type}, \text{:OilFacility}) \vee (?x, \text{rdf:type}, \text{:OilPipeline})$$

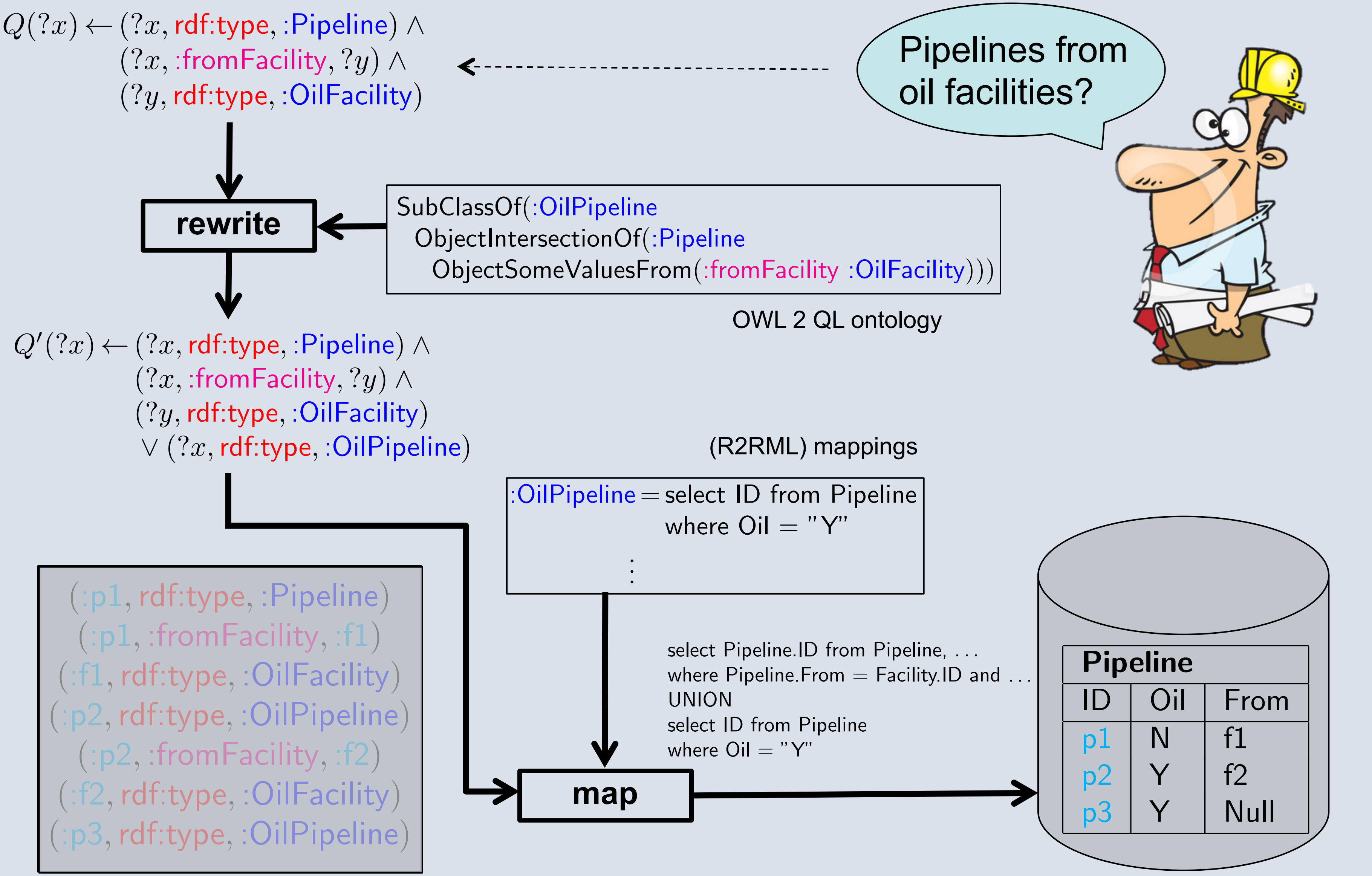
Pipeline

ID	Oil	From
p1	N	f1
p2	Y	f2
p3	Y	Null

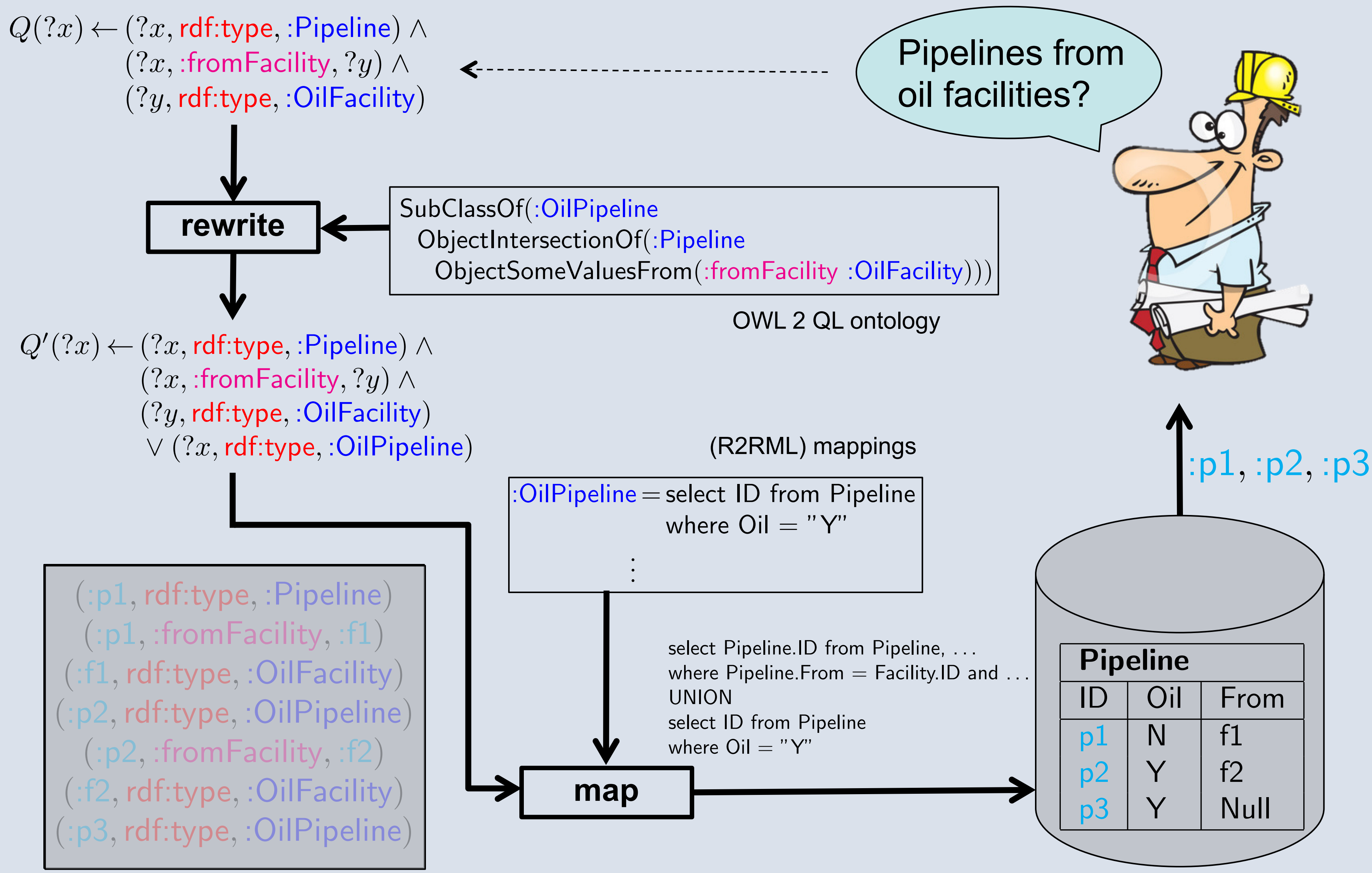
UNIVERSITY OF OXFORD DEPARTMENT OF COMPUTER SCIENCE
Information Systems Group
EPSRC Engineering and Physical Sciences Research Council
Optique



DEPARTMENT OF COMPUTER SCIENCE, UNIVERSITY OF OXFORD · Information Systems Group · EPSRC Engineering and Physical Sciences Research Council · Optique



$Q(?x) \leftarrow (?x, \text{rdf:type}, \text{:Pipeline}) \wedge (?x, \text{:fromFacility}, ?y) \wedge (?y, \text{rdf:type}, \text{:OilFacility})$
Pipelines from oil facilities?
rewrite
SubClassOf(:OilPipeline
ObjectIntersectionOf(:Pipeline
ObjectSomeValuesFrom(:fromFacility :OilFacility)))
OWL 2 QL ontology
$Q'(?x) \leftarrow (?x, \text{rdf:type}, \text{:Pipeline}) \wedge (?x, \text{:fromFacility}, ?y) \wedge (?y, \text{rdf:type}, \text{:OilFacility}) \vee (?x, \text{rdf:type}, \text{:OilPipeline})$
(R2RML) mappings
:OilPipeline = select ID from Pipeline where Oil = "Y"
⋮
select Pipeline.ID from Pipeline, . . .
where Pipeline.From = Facility.ID and . . .
UNION
select ID from Pipeline
where Oil = "Y"
map
(:p1, rdf:type, :Pipeline)
(:p1, :fromFacility, :f1)
(:f1, rdf:type, :OilFacility)
(:p2, rdf:type, :OilPipeline)
(:p2, :fromFacility, :f2)
(:f2, rdf:type, :OilFacility)
(:p3, rdf:type, :OilPipeline)
Pipeline
ID
Oil
From
p1
N
f1
p2
Y
f2
p3
Y
Null



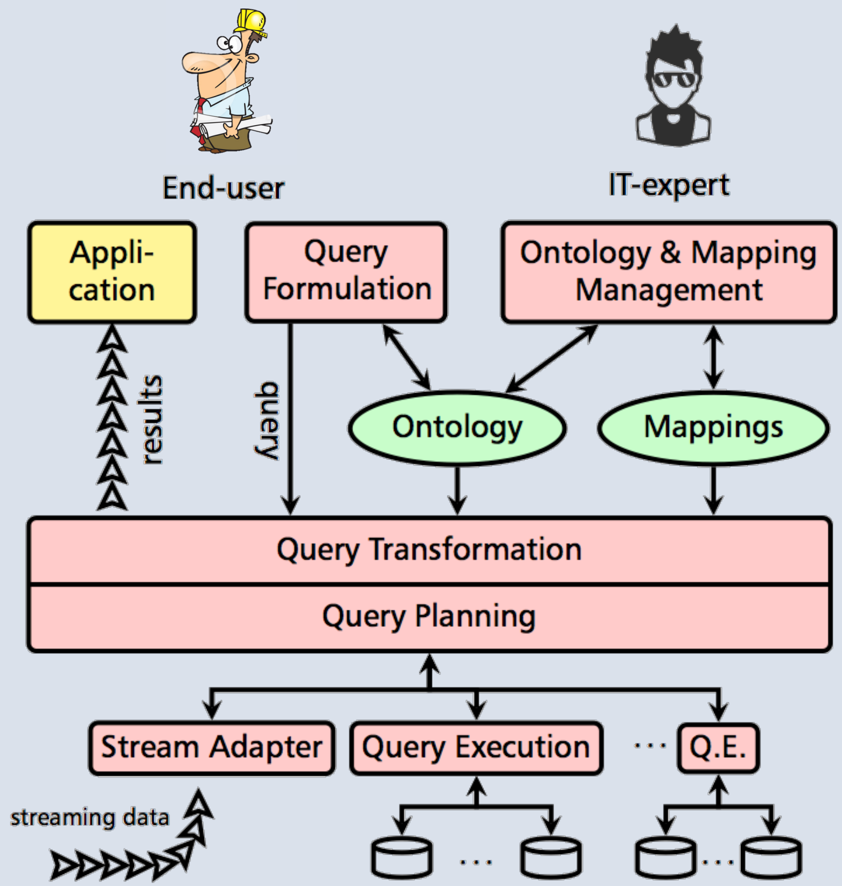
Information Systems Group

DEPARTMENT OF COMPUTER SCIENCE, UNIVERSITY OF OXFORD

EPSRC Engineering and Physical Sciences Research Council

Optique

Optique Architecture



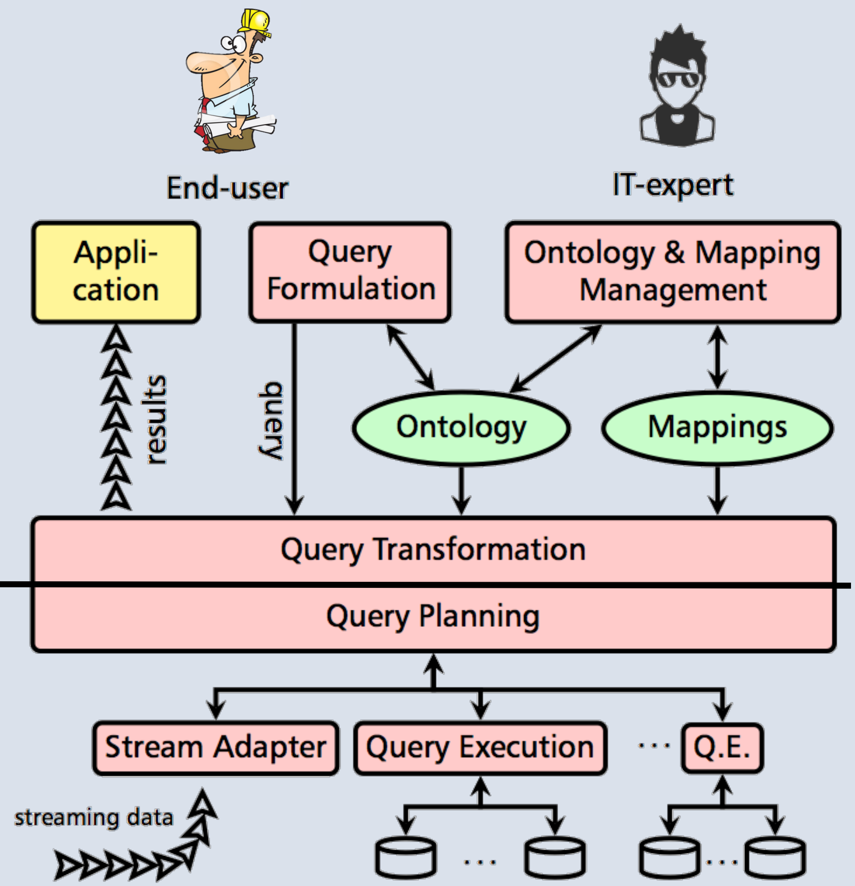
Optique Architecture

Query rewriting:

- uses ontology & mappings
- computationally hard
- ontology & mappings small

Query evaluation:

- ind. of ontology & mappings
- computationally tractable
- data sets very large



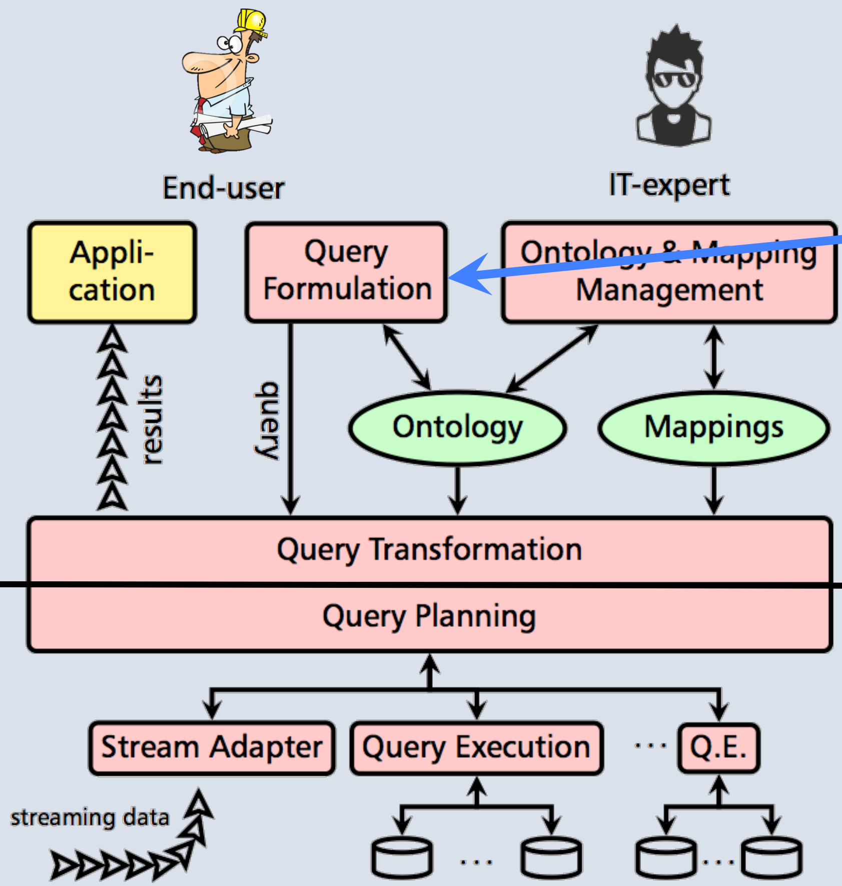
Optique Architecture

Query rewriting:

- uses ontology & mappings
- computationally hard
- ontology & mappings small

Query evaluation:

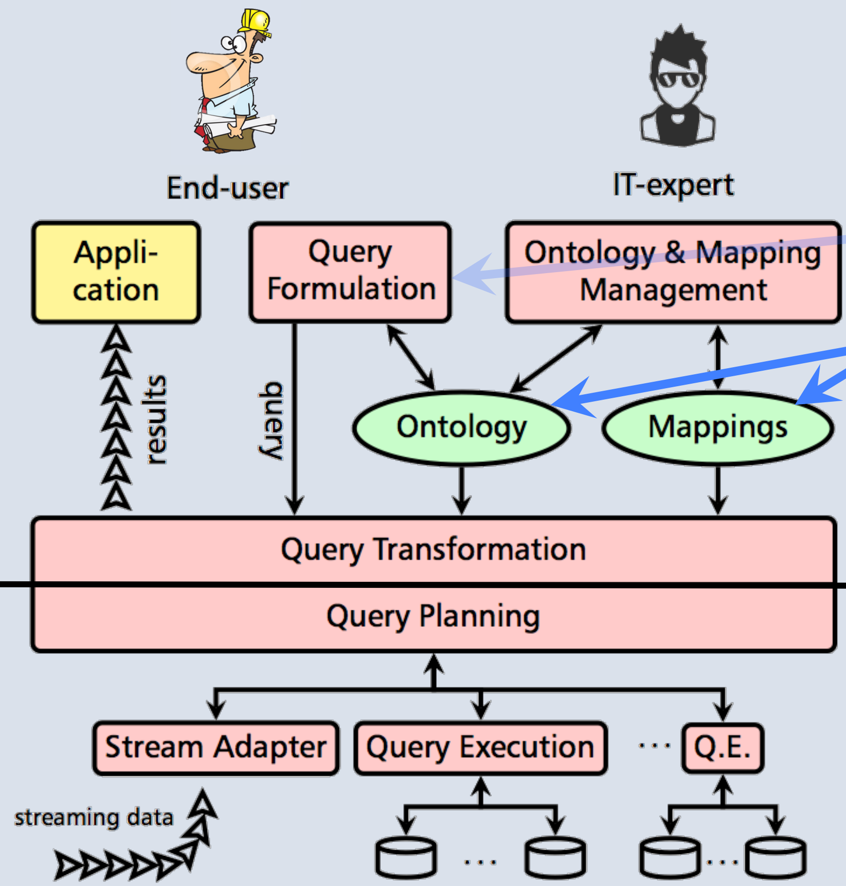
- ind. of ontology & mappings
- computationally tractable
- data sets very large



Other features:

support for query formulation

Optique Architecture



Query rewriting:
- uses ontology & mappings
- computationally hard
- ontology & mappings small

Query evaluation:
- ind. of ontology & mappings
- computationally tractable
- data sets very large

Other features:

support for query formulation

"Bootstrapping" Ontology & mappings

Research Issues

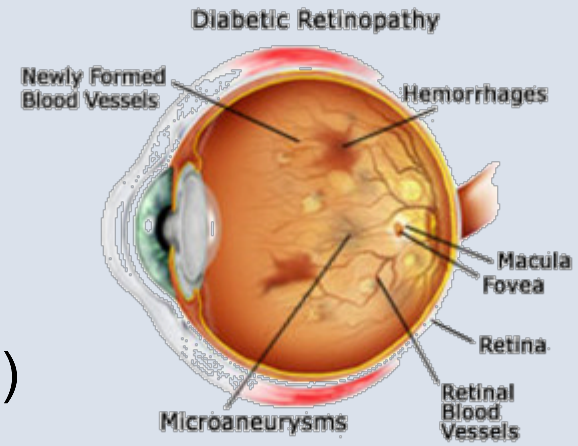
Research Issues

- **Expressive power**:
  - OWL QL (necessarily) has (very) restricted expressive power
  - Could use mappings to translate data into triples and use OWL RL
- **Scalability**:
  - Query size may increase exponentially in size of ontology
  - Rewritten queries may be hard for existing DBMSs
  - Extensive optimisation required [Bagosi et al]
- **Ontology and mapping engineering**:
  - Ontology engineering known to be hard
  - Less known about mappings, but likely to require similar tool support

Data Analysis: KAISER PERMANENTE

Data Analysis: KAISER PERMANENTE

- **HEDIS**[1] is a Performance Measure specification issued by **NCQA**[2]
  - E.g., all diabetic patients must have annual eye exams
- Meeting HEDIS standards is a requirement for government funded healthcare (**Medicare**)
- Checking/reporting is difficult and costly
  - Complex specifications & annual revisions
  - Disparate data sources
  - Ad hoc schemas including implicit information



[1] Healthcare Effectiveness Data and Information Set
[2] National Committee for Quality assurance

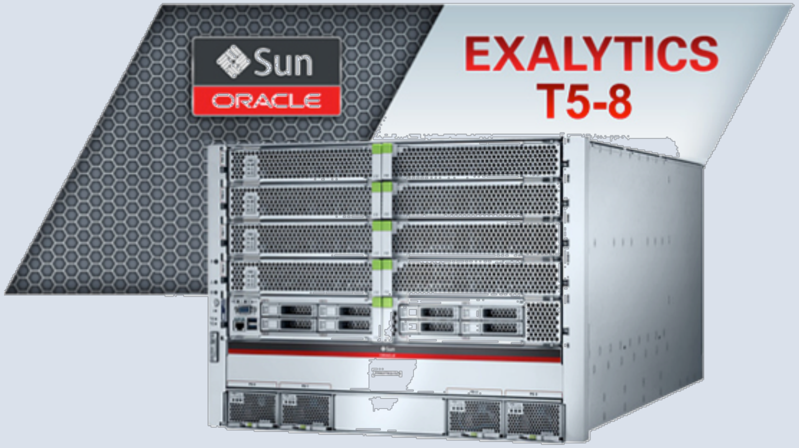
Semantic Technology Solution

Semantic Technology Solution

- Capture HEDIS diabetic care spec using **OWL RL & SWRL**
  - 174 axioms/rules
- Load data into (**RDFox**) triple store
  - Data from 466k patients (Georgia region); approx 100M records
  - Translated into 548M triples (32GB RAM)
- Use **materialisation** and **SPARQL queries** to identify relevant patients and check HEDIS conformance
  - Entire process takes ≈ 7,000s on a commodity server
  - Extends graph to 731M triples (43GB)

RDFox OWL RL Engine

- Targets SOTA **main-memory, mulit-core** architecture
  - Optimized in-memory storage with ‘mostly’ lock-free parallel inserts
  - Commodity server with 128 GB can store $>10^9$ triples
  - 10-20 x speedup with 32/16 threads/cores
  - LUBM 120K in 251s (20M triples/s) on T5-8 with 4TB/1024 threads



RDFox OWL RL Engine

- Targets SOTA **main-memory, mulit-core** architecture
  - Optimized in-memory storage with 'mostly' lock-free parallel inserts
  - Commodity server with 128 GB can store $>10^9$ triples
  - 10-20 x speedup with 32/16 threads/cores
  - LUBM 120K in 251s (20M triples/s) on T5-8 with 4TB/1024 threads

- Native equality reasoning (owl:sameAs) via rewriting
- Incremental reasoning (delete 5k triples from LUBM 50K in <1s)

Research Issues

Research Issues

- **Expressive power**:
  - Capturing HEDIS spec requires negation and aggregation
  - Current solution interleaves SPARQL queries & materialisation
  - Extending RDFox to support aggregation & stratified negation
- **Scalability**:
  - KP have approx 10M patients in total (20 times larger)
  - RDFox can store 10B triples on a 1TB machine [Nenov et al]
  - Working on (semantic) partitioning and distributed materialisation

Observations and **Opinions**

Language Design: Features/Bugs

Language Design: Features/Bugs

RDF blank nodes

- ✗ Existential semantics are not always intuitive or appropriate
- ✗ NP complexity for base layer (compared to AC^0 for DBs)
- ✗ Stack(s) is (are) broken
- ✓ Sometimes useful (of course)
- ✓ Endless entertainment/papers for theoreticians

Open world semantics

- ✗ Not always intuitive or appropriate (for DB people/applications)
- ✗ Difficult to combine with, e.g., defaults and NAF
- ✓ Appropriate for Web
- ✓ Appropriate (often) for data integration

Language Design: Features/Bugs

Only unary and binary predicates

- ✗ "Incompatibility" with DBs
- ✗ Reification costly and problematical
  - Mapping from DBs is non-trivial
  - Canonical URI creation is critical
- ✓ Often sufficient in practice
- ✓ Simple(r) data structures and algorithms

RDF über alles

- ✗ Verbose, and difficult to fully specify/constrain/parse syntax
- ✗ Nonsensical statements (rdf:type, rdf:type: rdf:type)
- ✓ Single storage and querying infrastructure
- ✓ Uniform/combined data and schema queries

Language Design: Features/Bugs

Lacks/includes necessary/useless features

- ✗ Specification should have included _________*
- ✗ Specification should not have included _________*

THERE IS NO ONE PERFECT LANGUAGE

- ✓ Standardisation critical for infrastructure development and uptake
- ✓ RDF+OWL+SPARQL provides a **huge advantage/opportunity**

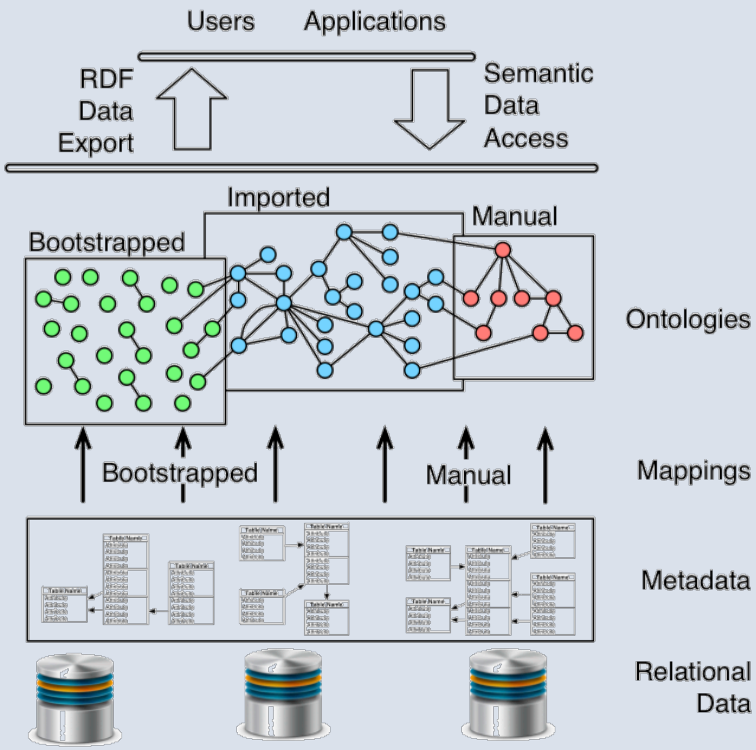
* Insert favourite/most-hated feature

Barriers to Uptake

Barriers to Uptake

Ontology (and mapping) engineering

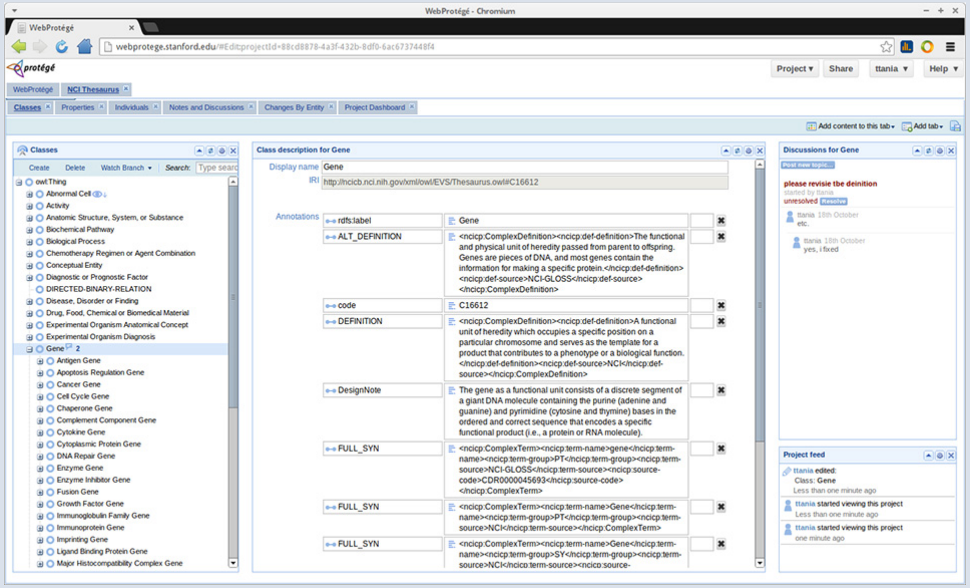
- Critically dependent on good quality ontologies (and mappings)
- Sophisticated tools and methodologies available
- But its still hard!



Barriers to Uptake

Ontology (and mapping) engineering

- Critically dependent on good quality ontologies (and mappings)
- Sophisticated tools and methodologies available
- But its still hard!



Kudos to Protégé

DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF OXFORD
Information Systems Group
EPSRC
Engineering and Physical Sciences Research Council
Optique

Barriers to Uptake

Scalability -v- expressive power

- Users expect semantic technology features in addition to DB features
- RDF entailment already NP-complete
- OWL 2 DL entailment NP-Hard w.r.t. size of data and N2ExpTime-complete w.r.t. size of ontology+data
- OWL 2 profiles "more simply and/or efficiently implemented"
  - OWL 2 QL – AC^0 data complexity (same as DBs)
  - OWL 2 EL – PTime-complete combined and data complexity
  - OWL 2 RL – PTime-complete combined and data complexity
- but at the cost of reduced expressive power

Scalability Beyond the Profiles?

LUBM ontology includes the axioms

```
    SubClassOf(:ResearchAssistant
                 ObjectIntersectionOf(:Person
                 ObjectSomeValuesFrom(:worksFor :ResearchGroup)))
EquivalentClasses(:Employee
                 ObjectIntersectionOf(:Person
                 ObjectSomeValuesFrom(:worksFor :Organization)))
    SubClassOf(:ResearchGroup :Organization)
```

and LUBM 1 data includes 547 instances of :ResearchAssistant

Scalability Beyond the Profiles?

LUBM ontology includes the axioms

SubClassOf(:ResearchAssistant
ObjectIntersectionOf(:Person
ObjectSomeValuesFrom(:worksFor :ResearchGroup)))
EquivalentClasses(:Employee
ObjectIntersectionOf(:Person
ObjectSomeValuesFrom(:worksFor :Organization)))
SubClassOf(:ResearchGroup :Organization)

and LUBM 1 data includes 547 instances of :ResearchAssistant

How many of these RAs are in the answer to the following query?

```
SELECT ?x WHERE { ?x rdf:type :Employee }
```

Scalability Beyond the Profiles?

LUBM ontology includes the axioms

```
    SubClassOf(:ResearchAssistant
                    ObjectIntersectionOf(:Person
                    ObjectSomeValuesFrom(:worksFor :ResearchGroup)))
EquivalentClasses(:Employee
                    ObjectIntersectionOf(:Person
                    ObjectSomeValuesFrom(:worksFor :Organization)))
    SubClassOf(:ResearchGroup :Organization)
```

and LUBM 1 data includes 547 instances of :ResearchAssistant

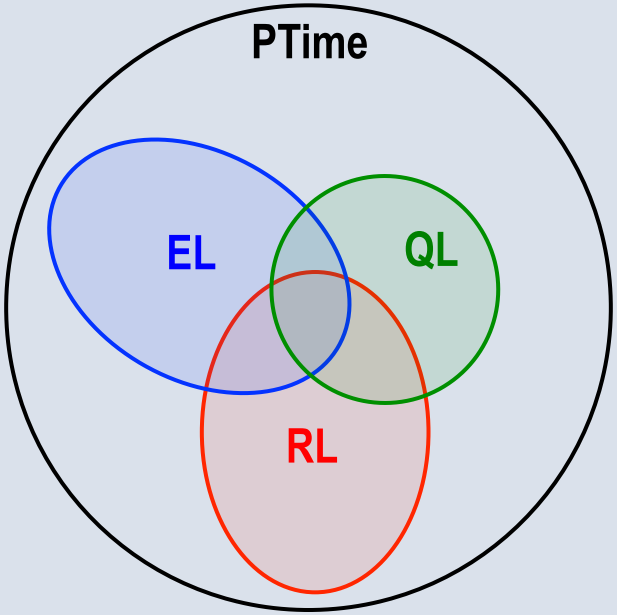
How many of these RAs are in the answer to the following query?

```
SELECT ?x WHERE { ?x rdf:type :Employee }
```

Answer computed by (most) RL reasoners: **none of them**

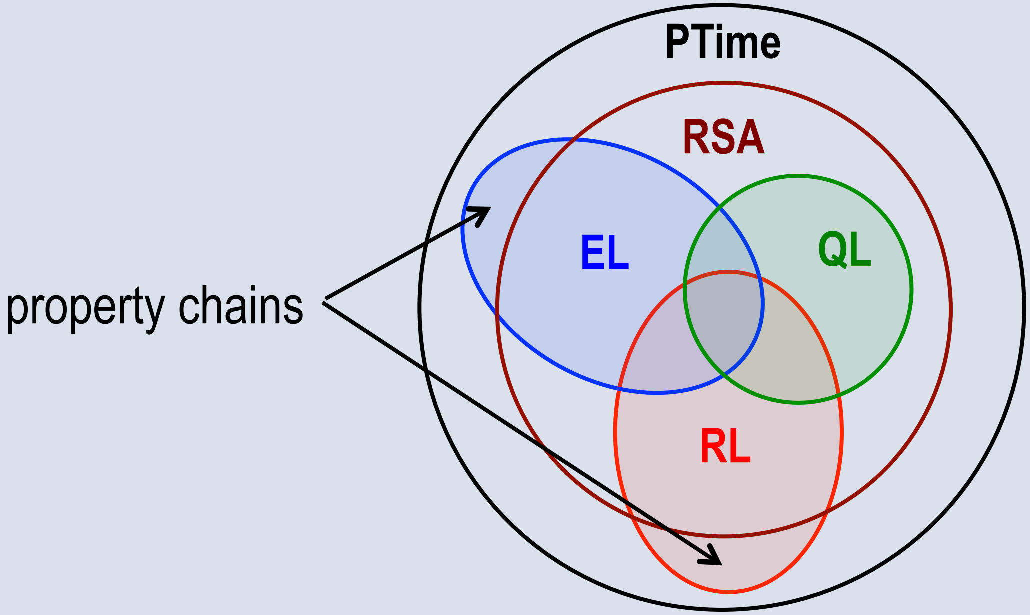
Combined Approach

- Combined approach allows RL reasoners to be used to support scalable query answering for all profiles



Combined Approach

- Combined approach allows RL reasoners to be used to support scalable query answering for all profiles
- And even for many non-profile ontologies in RSA class



Combined Approach

- Combined approach allows RL reasoners to be used to support scalable query answering for all profiles
- And even for many non-profile ontologies in RSA class
- How does it work?
  - Overapproximate $\mathcal{O}$ (e.g., "Skolemise" RHS exstentials) into $\mathcal{O}'$ i.e., $\mathcal{O}' \models \mathcal{O}$
  - Queries answered w.r.t. $\mathcal{O}'$ to give complete but unsound answers
  - Spurious answers are eliminated by a filtration step

Combined Approach

$\mathcal{O} \rightsquigarrow \mathcal{O}'$

$\mathcal{O}$:

SubClassOf(:ResearchAssistant
ObjectIntersectionOf(:Person
ObjectSomeValuesFrom(:worksFor :ResearchGroup)))

$\mathcal{O}'$:

SubClassOf(:ResearchAssistant
ObjectIntersectionOf(:Person
ObjectHasValue(:worksFor :RG1)))
(:RG1, rdf:type, :ResearchGroup)

Combined Approach

$\mathcal{O}$

SubClassOf(:ResearchAssistant
ObjectIntersectionOf(:Person
ObjectSomeValuesFrom(:worksFor :ResearchGroup)))

$\rightsquigarrow$

$\mathcal{O}'$

SubClassOf(:ResearchAssistant
ObjectIntersectionOf(:Person
ObjectHasValue(:worksFor :RG1)))
(:RG1, rdf:type, :ResearchGroup)

Combined Approach

$\mathcal{O} \rightsquigarrow \mathcal{O}'$

$\mathcal{O}$

SubClassOf(:ResearchAssistant
ObjectIntersectionOf(:Person
ObjectSomeValuesFrom(:worksFor :ResearchGroup)))

EquivalentClasses(:Employee
ObjectIntersectionOf(:Person
ObjectSomeValuesFrom(:worksFor :Organization)))

$\mathcal{O}'$

SubClassOf(:ResearchAssistant
ObjectIntersectionOf(:Person
ObjectHasValue(:worksFor :RG1)))
(:RG1, rdf:type, :ResearchGroup)

SubClassOf(:Employee
ObjectIntersectionOf(:Person
ObjectHasValue(:worksFor :ORG1)))
(:ORG1, rdf:type, :Organization)
SubClassOf(ObjectIntersectionOf(:Person
ObjectSomeValuesFrom(:worksFor :Organization))
:Employee)

Combined Approach

$\mathcal{O} \rightsquigarrow \mathcal{O}'$

$\mathcal{O}$

SubClassOf(:ResearchAssistant
ObjectIntersectionOf(:Person
ObjectSomeValuesFrom(:worksFor :ResearchGroup)))

EquivalentClasses(:Employee
ObjectIntersectionOf(:Person
ObjectSomeValuesFrom(:worksFor :Organization)))

SubClassOf(:ResearchGroup :Organization)

$\mathcal{O}'$

SubClassOf(:ResearchAssistant
ObjectIntersectionOf(:Person
ObjectHasValue(:worksFor :RG1)))
(:RG1, rdf:type, :ResearchGroup)

SubClassOf(:Employee
ObjectIntersectionOf(:Person
ObjectHasValue(:worksFor :ORG1)))
(:ORG1, rdf:type, :Organization)
SubClassOf(ObjectIntersectionOf(:Person
ObjectSomeValuesFrom(:worksFor :Organization))
:Employee)

SubClassOf(:ResearchGroup :Organization)

Combined Approach

$\mathcal{O} \rightsquigarrow \mathcal{O}'$

$\mathcal{O}$	$\mathcal{O}'$
SubClassOf(:ResearchAssistant ObjectIntersectionOf(:Person ObjectSomeValuesFrom(:worksFor :ResearchGroup)))	SubClassOf(:ResearchAssistant ObjectIntersectionOf(:Person ObjectHasValue(:worksFor :RG1))) (:RG1, rdf:type, :ResearchGroup)
EquivalentClasses(:Employee ObjectIntersectionOf(:Person ObjectSomeValuesFrom(:worksFor :Organization)))	SubClassOf(:Employee ObjectIntersectionOf(:Person ObjectHasValue(:worksFor :ORG1))) (:ORG1, rdf:type, :Organization) SubClassOf(ObjectIntersectionOf(:Person ObjectSomeValuesFrom(:worksFor :Organization)) :Employee)
SubClassOf(:ResearchGroup :Organization)	SubClassOf(:ResearchGroup :Organization)

$$\mathcal{O}' \cup \{(\text{:RA1}, \text{rdf:type}, \text{:ResearchAssistant})\} \models (\text{:RA1}, \text{rdf:type}, \text{:Employee})$$

UNIVERSITY OF OXFORD DEPARTMENT OF COMPUTER SCIENCE
Information Systems Group
EPSRC Engineering and Physical Sciences Research Council
Optique

Combined Approach

$\mathcal{O}$ $\rightsquigarrow$ $\mathcal{O}'$

$\mathcal{O}$	$\mathcal{O}'$
SubClassOf(:ResearchAssistant ObjectIntersectionOf(:Person ObjectSomeValuesFrom(:worksFor :ResearchGroup)))	SubClassOf(:ResearchAssistant ObjectIntersectionOf(:Person ObjectHasValue(:worksFor :RG1))) (:RG1, rdf:type, :ResearchGroup)
EquivalentClasses(:Employee ObjectIntersectionOf(:Person ObjectSomeValuesFrom(:worksFor :Organization)))	SubClassOf(:Employee ObjectIntersectionOf(:Person ObjectHasValue(:worksFor :ORG1))) (:ORG1, rdf:type, :Organization) SubClassOf(ObjectIntersectionOf(:Person ObjectSomeValuesFrom(:worksFor :Organization)) :Employee)
SubClassOf(:ResearchGroup :Organization)	SubClassOf(:ResearchGroup :Organization)

$\mathcal{O}' \cup \{(\text{:RA1}, \text{rdf:type}, \text{:ResearchAssistant})\} \models (\text{:RA1}, \text{rdf:type}, \text{:Employee})$

SELECT ?x, ?y WHERE { ?x :worksFor ?z . ?y :worksFor ?z . ?z rdf:type :Employee }

Scalable Query Answering for OWL DL?

Given an **OWL DL ontology** $\mathcal{O}$ dataset $\mathcal{D}$ and query q

- We can transform $\mathcal{O}$ into strictly stronger OWL RL ontology $\mathcal{O}_u$
  - Roughly speaking, Skolemise, and transform $\vee$ into $\wedge$

Scalable Query Answering for OWL DL?

Given an **OWL DL ontology** $\mathcal{O}$ dataset $\mathcal{D}$ and query q

- We can transform $\mathcal{O}$ into strictly stronger OWL RL ontology $\mathcal{O}_\mathsf{u}$
  - Roughly speaking, Skolemise, and transform $\vee$ into $\wedge$
- RL reasoning w.r.t. $\mathcal{O}_\mathsf{u}$ gives upper bound answer U

$$\mathsf{ans}(q, \langle \mathcal{O}, \mathcal{D} \rangle) \subseteq \mathsf{ans}(q, \langle \mathcal{O}_\mathsf{u}, \mathcal{D} \rangle) = \mathsf{U}$$

Scalable Query Answering for OWL DL?

Given an **OWL DL ontology** $\mathcal{O}$ dataset $\mathcal{D}$ and query q

- We can transform $\mathcal{O}$ into strictly stronger OWL RL ontology $\mathcal{O}_{\mathsf{u}}$
  - Roughly speaking, Skolemise, and transform $\vee$ into $\wedge$
- RL reasoning w.r.t. $\mathcal{O}_{\mathsf{u}}$ gives upper bound answer U

$$\mathsf{ans}(q, \langle \mathcal{O}, \mathcal{D} \rangle) \subseteq \mathsf{ans}(q, \langle \mathcal{O}_{\mathsf{u}}, \mathcal{D} \rangle) = \mathsf{U}$$

- If L = U, then both answers are sound and complete

$$\mathsf{L} \subseteq \mathsf{ans}(q, \langle \mathcal{O}, \mathcal{D} \rangle) \subseteq \mathsf{U}$$

UNIVERSITY OF OXFORD DEPARTMENT OF COMPUTER SCIENCE
Information Systems Group
EPSRC Engineering and Physical Sciences Research Council
Optique

Scalable Query Answering for OWL DL?

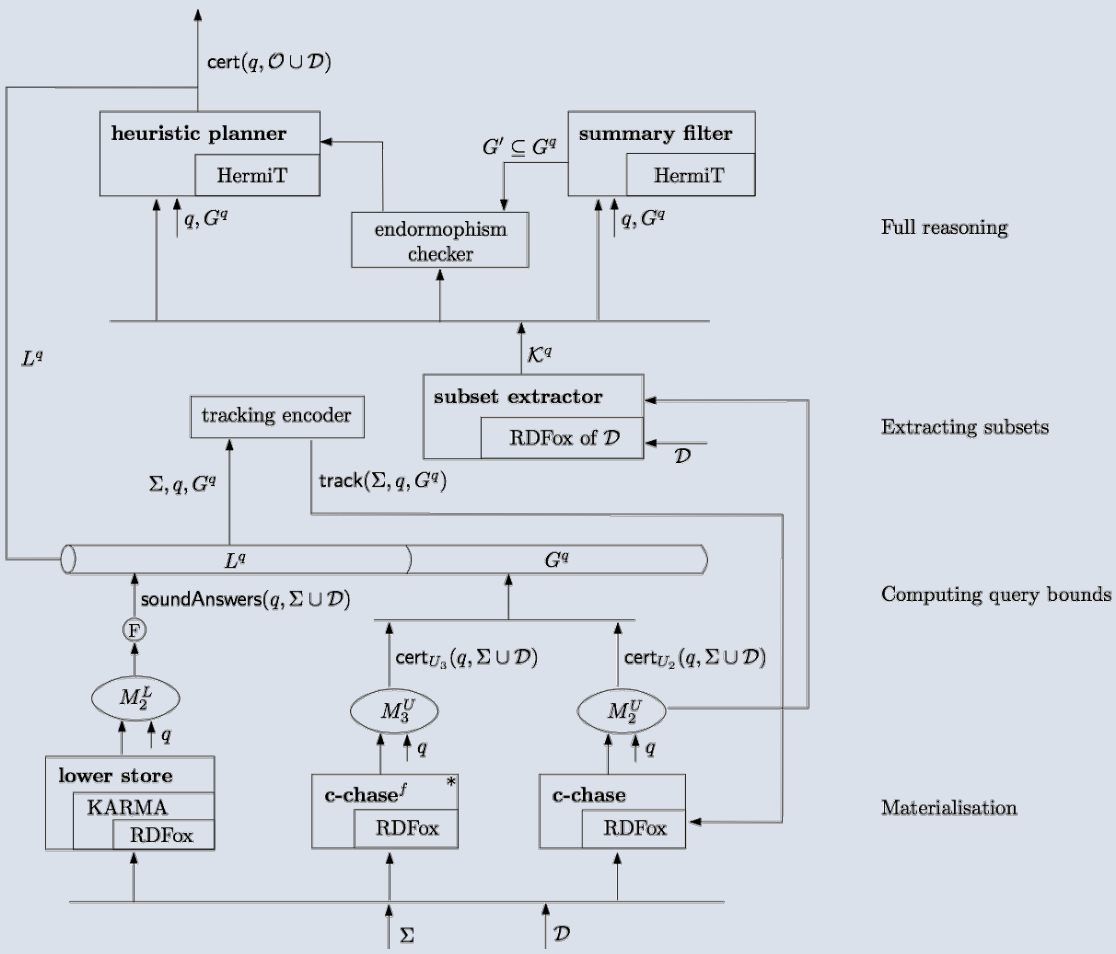
DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF OXFORD
Information Systems Group
EPSRC
Engineering and Physical Sciences Research Council
Optique

Scalable Query Answering for OWL DL?

- If L ≠ U, then U \ L identifies a (small) set of "possible" answers
  - Delineates range of uncertainty
  - Can more efficiently check possible answers using, e.g., HermiT (but still infeasible if dataset is large)
  - Can use U \ L to identify small(er) "relevant" subset of axioms/data sufficient to check possible answers (using proof tracing)

- Further optimisations
  - Use ELHO "combined" technique to tighten lower bound [Stefanoni et al]
  - Use "summarisation" technique to tighten upper bound [Dolby et al]
  - …

PAGOdA System



PAGOdA System

(b) LUBM query processing

A Perspective on the Semantic Web

A Perspective on the Semantic Web

UNIVERSITY OF OXFORD DEPARTMENT OF COMPUTER SCIENCE

Information Systems Group

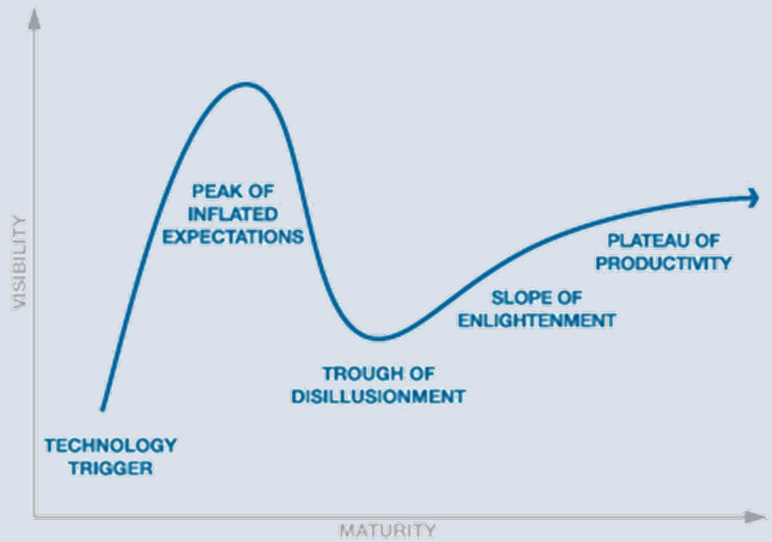
EPSRC Engineering and Physical Sciences Research Council

Optique

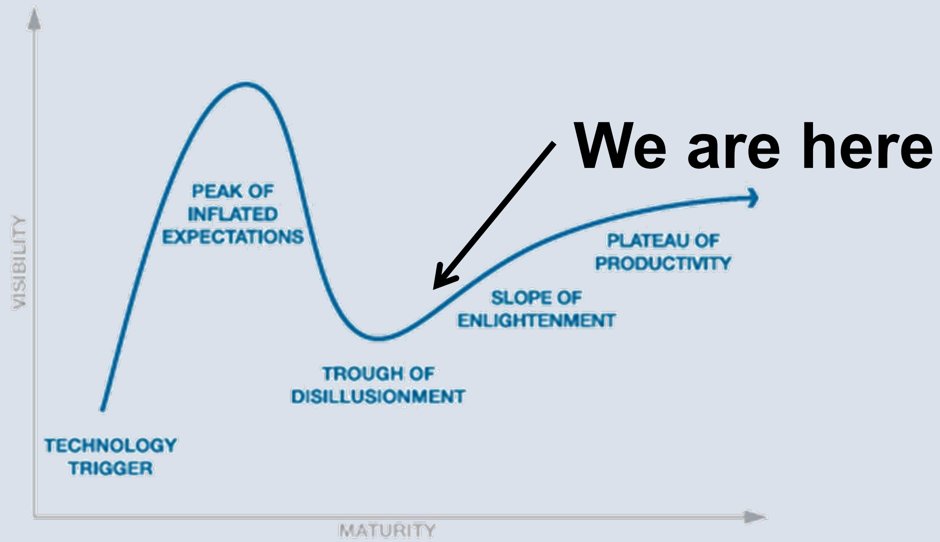
A Perspective on the Semantic Web

- **Graph DB** with rich and flexible schema
- Applications in **data integration and analysis** (as well as the Web)
- **Competing** with Graph/NoSQL DBs, Bigtable, HBase, …
- **RDF+OWL+SPARQL** standards and technologies offer important advantages

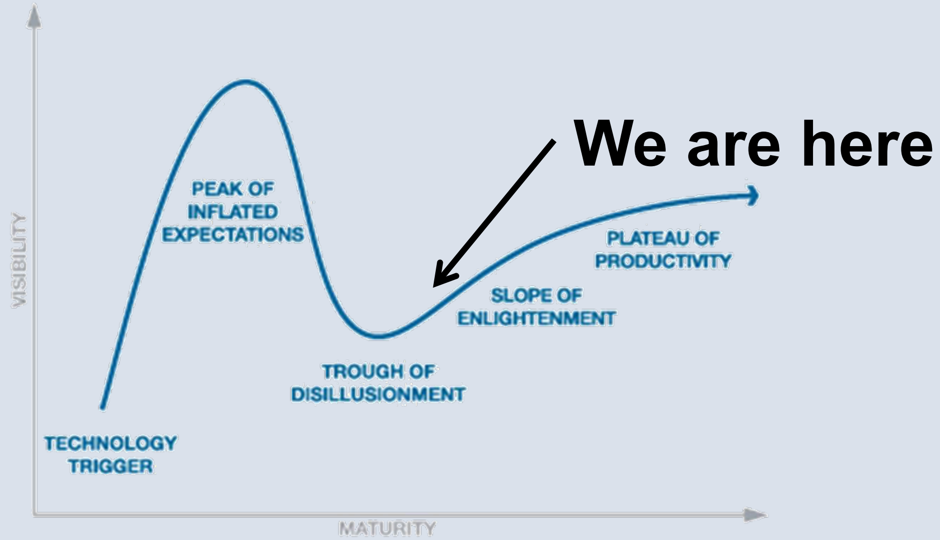
A Perspective on the Semantic Web



A Perspective on the Semantic Web



A Perspective on the Semantic Web



- We have the (right) **languages**
- We have the (right) **technology**
- We have interest and even enthusiasm from (potential) **users**
- All(!) we need to do is **engage** and (continue to) **deploy**

Acknowledgements

EPSRC Engineering and Physical Sciences Research Council

SIXTH FRAMEWORK PROGRAMME

SEVENTH FRAMEWORK PROGRAMME

DARPA

DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF OXFORD

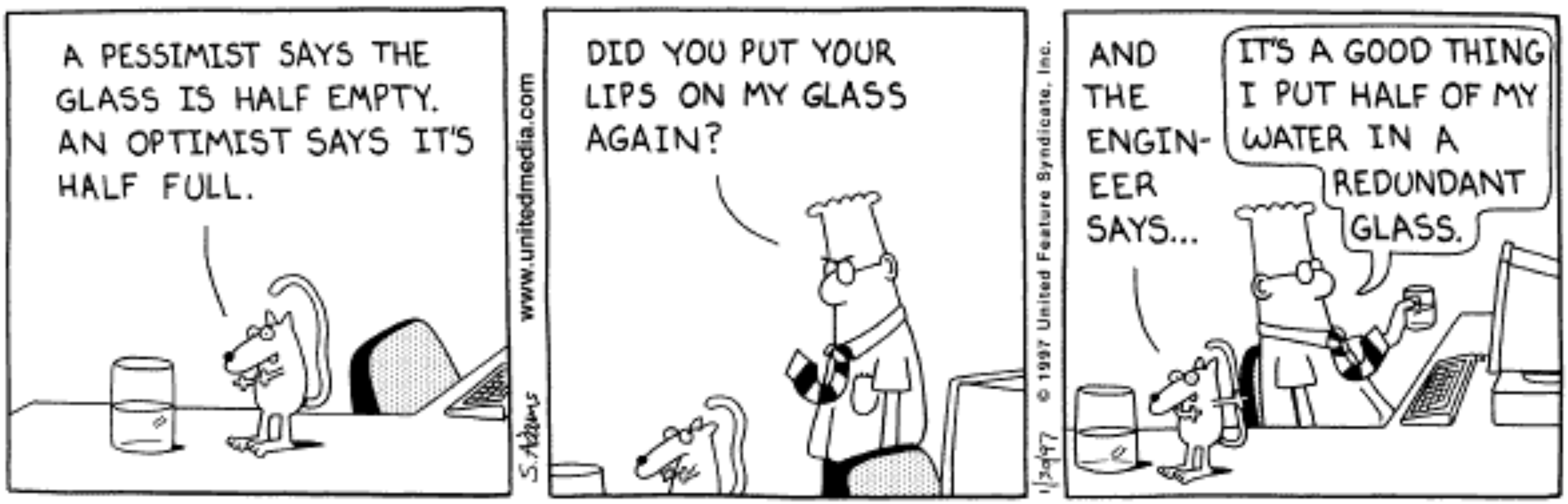
Information Systems Group

EPSRC
Engineering and Physical Sciences Research Council

Optique

Thank you for listening

Thank you for listening



Any questions?

UNIVERSITY OF OXFORD DEPARTMENT OF COMPUTER SCIENCE

Information Systems Group

EPSRC Engineering and Physical Sciences Research Council

Optique

References

[Kazakov et al] Yevgeny Kazakov, Markus Krötzsch, Frantisek Simancik: The Incredible ELK - From Polynomial Procedures to Efficient Reasoning with $\mathcal{EL}$ Ontologies. J. Autom. Reasoning 53(1): 1-61 (2014)

[Steigmiller et al] Andreas Steigmiller, Thorsten Liebig, Birte Glimm: Konclude: System description. J. Web Sem. 27: 78-85 (2014)

[Armas Romero et al] Ana Armas Romero, Bernardo Cuenca Grau, Ian Horrocks: MORe: Modular Combination of OWL Reasoners for Ontology Classification. In ISWC 2012: 1-16.

[Bagosi et al] Timea Bagosi, Diego Calvanese, Josef Hardi, Sarah Komla-Ebri, Davide Lanti, Martin Rezk, Mariano Rodriguez-Muro, Mindaugas Slusnys, Guohui Xiao: The Ontop Framework for Ontology Based Data Access. CSWS 2014: 67-77

References

[Nenov et al] Y. Nenov, R. Piro, B. Motik, I. Horrocks, Z. Wu, and J. Banerjee: RDFox: A Highly-Scalable RDF Store. In ISWC 2015.

[Staab & Studer] Steffen Staab, Rudi Studer: Handbook on Ontologies. Springer 2009.

[Dolby et al] J. Dolby, A. Fokoue, A. Kalyanpur, E. Schonberg, K. Srinivas: Scalable highly expressive reasoner (SHER). J. Web Sem. 7(4): 357-361 (2009)

[Stefanoni et al] Giorgio Stefanoni, Boris Motik, Ian Horrocks: Introducing Nominals to the Combined Query Answering Approaches for EL. AAAI 2013