# My research areas

**Michael Zakharyaschev** http://www.dcs.bbk.ac.uk/~michael/

(with all my publications and funded research projects)

- Knowledge representation and reasoning in AI
  - ontologies, description logics
  - query languages
  - ontology-mediated queries
  - temporal and spatial representation, reasoning and querying
- Semantic technologies
  - ontology-based data access
  - Statoil use case
  - Siemens use case
- Mathematical logic
  - modal logic, temporal logic, spatial logic, dynamic logic
  - complexity theory

# Querying resources on the Web

**Resource Description Framework (RDF)** http://www.w3.org/RDF/

a general method for conceptual description or modelling of information for web resources

- represents data and metadata as triples (subject, predicate, object)
  - Siemens phone data

    dbr:Siemens_C25 dbo:manufacturer dbr:Siemens .
    dbr:Siemens_C25 rdf:type yago:SiemensMobilePhones .
  - IBM phone data

    dbr:IBM_Simon dbo:manufacturer dbr:IBM .
    dbr:IBM_Simon rdf:type yago:IBMMobilePhones .

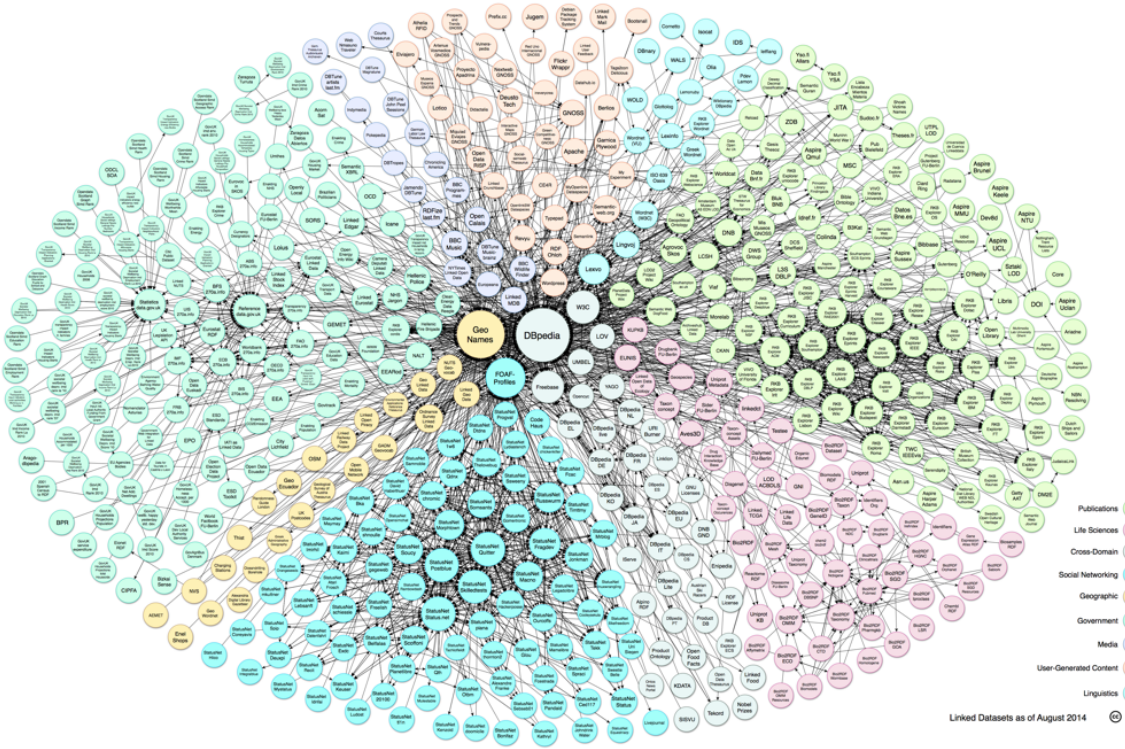
**SPARQL query:**

```
SELECT ?C
WHERE
{  ?C rdf:type dbo:Organisation .
   ?C foaf:made ?P .
   ?P rdf:type yago:Telephone . }
```

**Ontology**

$\forall x, y(manufacturer(x, y) \rightarrow maker(x, y))$
$\forall x, y(maker(x, y) \rightarrow made(y, x))$
$\forall x, y(manufacturer(x, y) \rightarrow organisation(y))$
$\forall x(SiemensMobilePhone(x) \rightarrow Telephone(x))$
$\forall x(IBMMobilePhone(x) \rightarrow Telephone(x))$

# Linked Data


Geo Names
DBpedia
W3C
LOV
FOAF-Profiles
Freebase
Publications
Life Sciences
Cross-Domain
Social Networking
Geographic
Government
Media
User-Generated Content
Linguistics
Linked Datasets as of August 2014

# Querying relational data: Statoil case

(from Norwegian Petroleum Directorate's FactPages)

show me the wellbores completed before 2008 where Statoil as a drilling operator sampled less than 10 meters of cores

5 days later:

```
SELECT DISTINCT cores.wlbName, cores.lenghtM, wellbore.wlbDrillingOperator, wellbore.wlbCompletionYear
FROM
    ( (SELECT wlbName, wlbNpdidWellbore, (wlbTotalCoreLength * 0.3048) AS lenghtM
       FROM wellbore_core
       WHERE wlbCoreIntervalUom = '(ft )' )
      UNION
      (SELECT wlbName, wlbNpdidWellbore, wlbTotalCoreLength AS lenghtM
       FROM wellbore_core
       WHERE wlbCoreIntervalUom = '(m )' )
    ) as cores,
    ( (SELECT wlbNpdidWellbore, wlbDrillingOperator, wlbCompletionYear
       FROM wellbore_development_all
      UNION
      (SELECT wlbNpdidWellbore, wlbDrillingOperator, wlbCompletionYear
       FROM wellbore_exploration_all )
      UNION
      (SELECT wlbNpdidWellbore, wlbDrillingOperator, wlbCompletionYear
       FROM wellbore_shallow_all )
    ) as wellbore
WHERE wellbore.wlbNpdidWellbore = cores.wlbNpdidWellbore
    ...
```

**In STATOIL:**

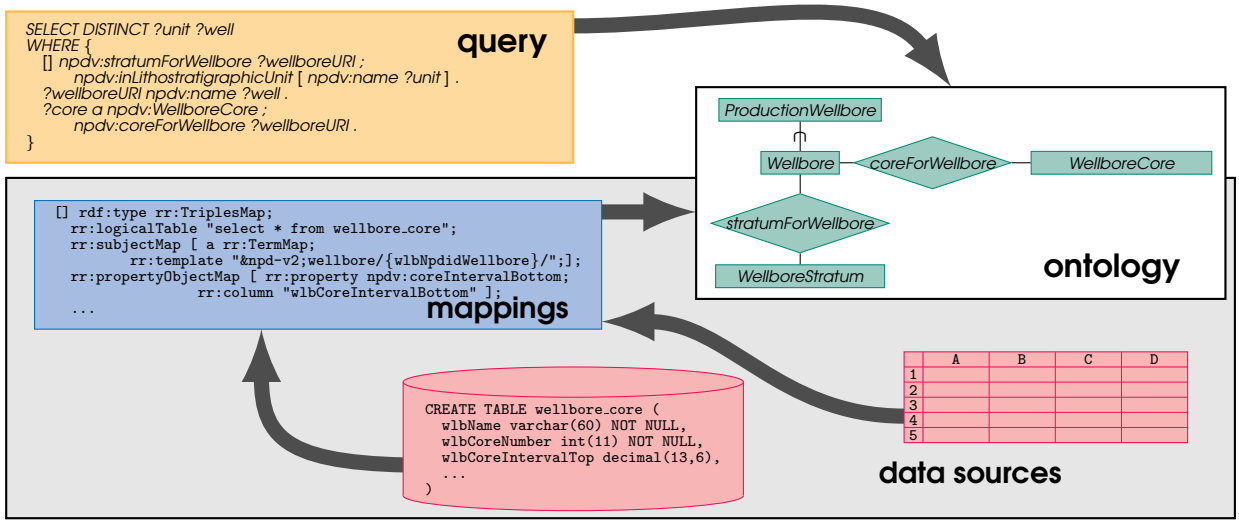
**1,000 TB of relational data**

**2,000 tables**

**different schemas**

**30–70% of time on data gathering**

# Ontology-based data access (OBDA)



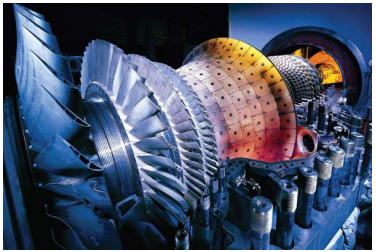

## Ontology

- gives a high-level conceptual view of the data
- provides a convenient & natural vocabulary for user queries
- enriches incomplete data with background knowledge

# Temporal OBDA: Siemens case

Siemens remote-diagnostic centres analyse data by querying aggregated sensor measurements

| turbine_Id | dateTime | activePower | rotorSpeed | mainFlame | ... |
|---|---|---|---|---|---|
| | | ... | | | |
| t03 | 2015-04-04 12:20:48 | 2 | 1550 | 0 | |
| t03 | 2015-04-04 12:20:49 | 1.8 | 1400 | null | |
| t01 | 2015-04-04 12:20:52 | 1.7 | 1350 | 1 | |
| | | ... | | | |



**Typical query:** when an **active power trip** occurred, i.e., the active power was above 1.5MW for a period of at least 10 seconds, 3 seconds after which there was a period of at least one minute where active power was below 0.15MW

**Metric temporal logic ontology, $\Pi$,** with axioms such as

$$\textbf{ActivePowerTrip}(v) \leftarrow \textbf{Turbine}(v) \wedge \boxminus_{[0,1m]} \textbf{ActivePowerBelow0.15}(v) \wedge \diamondminus_{[60s,63s]} \boxminus_{[0,10s]} \textbf{ActivePowerAbove1.5}(v)$$

**Ontology-mediated query** $(\Pi, \textbf{ActivePowerTrip}(t)@x)$

(find turbines $t$ and moments $x$ at which $t$ had an active power trip)

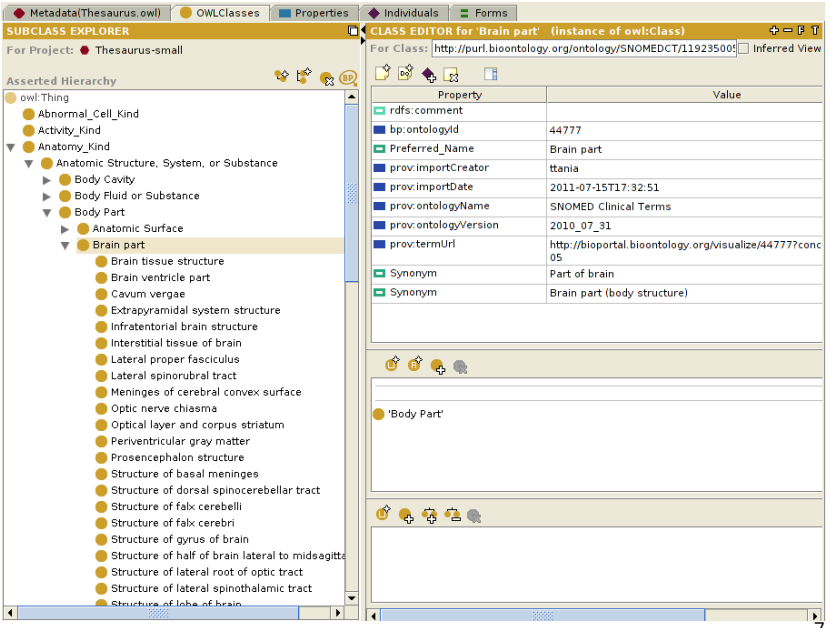
# Ontology Applications: Healthcare

**SNOMED CT** is a formal ontology of medical terms, synonyms and definitions used in clinical documentation and reporting

most comprehensive, multilingual clinical healthcare terminology in the world

- electronic record systems
- catalogues of services
- clinical decision support
- laboratory reporting
- genetic databases
- ...

used in most developed countries in the world



# Semantic Web Applications: Media

**BBC** website for the Football World Cup 2010: 32 teams, 8 groups, 776 players

too many pages to create, too few journalists

Semantic Web technologies used:

- ontology describes the interrelation between facts of the World Cup
- such metadata stored as RDF triples


- e-governments
- libraries
- news papers
- dbpedia
- . . .

# Research projects

## 'Practical'

- Extracting RDF data from unstructured texts
- Visualising SPARQL queries
- Extracting ontologies and mappings from databases
- Application for querying DBLP (computer science bibliography)
- Data integration via RDF
- Algorithms for answering temporal ontology-mediated queries
- Extracting 'modules' from ontologies
- Ontology-based information systems in your favourite area (movies, football, food, travels, etc.)

## 'Theoretical'

- Characterise ontology-mediated queries that can be reduced to SQL queries or SPARQL queries (with or without transitive closure)
- Given an ontology in an expressive language, determine whether it can be 'equivalently' rewritten in a simpler language