Semantic Web & related technologies

Carlos Buil Aranda

Ontology Engineering Group
Facultad de Informática
Universidad Politécnica de Madrid
cbuil@fi.upm.es
NeSC 25th September 2009

Table of Contents

What is the Semantic Web?

- An extension of the current Web...
  - ... where information and services are given well-defined and explicitly represented meaning, ...
  - ... so that it can be shared and used by humans and machines, ...
  - ... better enabling them to work in cooperation

- How?
  - Promoting information exchange by tagging web content with machine processable descriptions of its meaning.
  - And technologies and infrastructure to do this

Ontology Languages

- Work on Semantic Web has concentrated on the definition of a collection or “stack” of languages.
  - Used to support the representation and use of metadata
  - Basic machinery that we can use to represent the extra semantic information needed for the Semantic Web

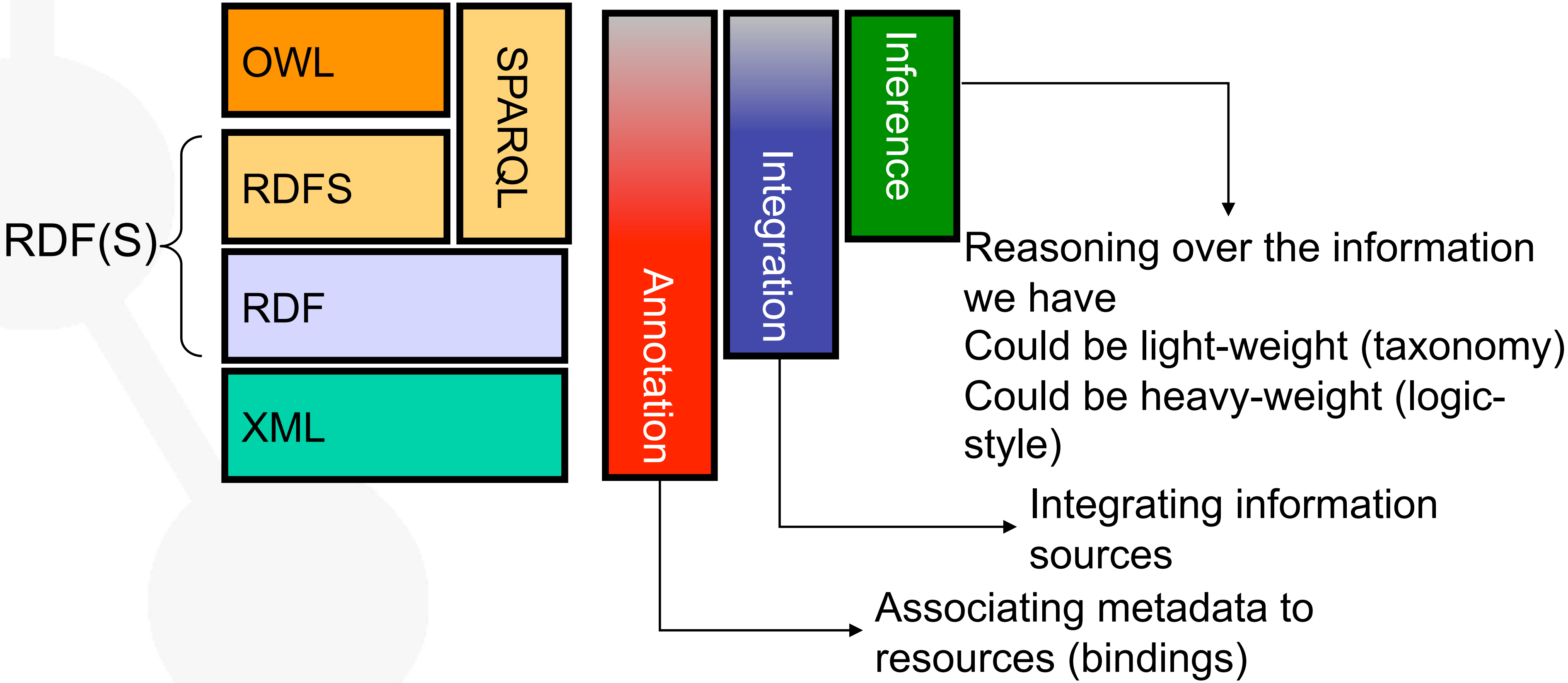
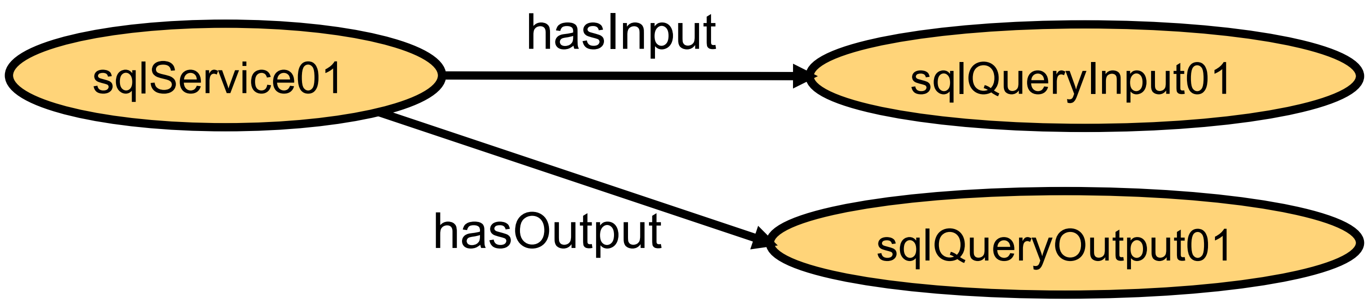


Table of Contents

- RDF stands for Resource Description Framework
  - It is a W3C Recommendation (http://www.w3.org/RDF)
- “RDF is to data what HTML is to documents”
- RDF is a graphical formalism (+ XML syntax + semantics)
  - for representing metadata
  - for describing the semantics of information in a machine-accessible way
- Provides a simple data model based on triples.

- Statements are <subject, predicate, object> triples:
  - <sqlService01,hasInput,sqlQueryInput01>, <sqlService01,hasOutput,sqlQueryOutput01>



- Statements describe properties of resources
- A resource is any object that can be pointed to by a URI
  - a document, a picture, a paragraph on the Web, http://www.epcc.ed.ac.uk/~ally, a book in the library, a real person, isbn://0141184280
- Properties themselves are also resources (URIs)

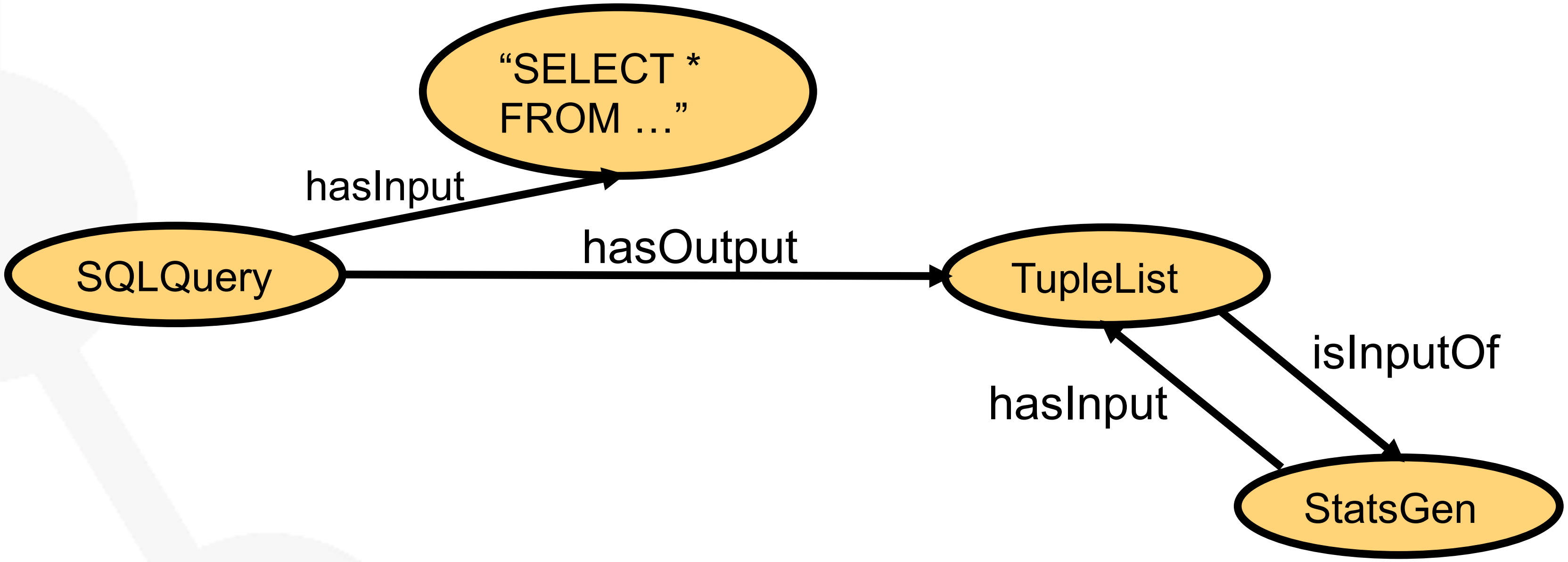
The RDF data model

```
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"     xmlns:CRISP-
DMIOntology="http://www.admire-project.eu/ontologies/CRISP-DMIOntology.owl#"
```

```
<SQLQueryService rdf:about="#sqlService01">
     <SQLQueryServiceName rdf:datatype="&rdf;XMLLiteral">
                              sqlQueryService01
                 </SQLQueryServiceName>
       <CRISP-DMIOntology:hasOutput rdf:resource="#sqlQueryOutput01"/>
      <CRISP-DMIOntology:hasInput rdf:resource="#sqlQueryInput01"/>
</SQLQueryService>
```

```
<CRISP-DMIOntology:TupleList rdf:about="#sqlQueryOutput01">
                 <outputSQLquery rdf:datatype="&rdf;XMLLiteral“></outputSQLquery>
</CRISP-DMIOntology:TupleList>
```

- The subject of one statement can be the object of another
- Such collections of statements form a directed, labeled graph



- The object of a triple can also be a "literal" (a string)

- Goal: to have a simple method that can express any fact
- RDF applications can put together RDF files and learn things from all of them
  - by linking documents together by the common vocabularies they use
  - by allowing any document to use any vocabulary
- Use cases
  - You want to integrate data from different sources without custom programming.
  - You want to offer your data for re-use by other parties

- SPARQL is the query language for RDF
- SPARQL is a graph-matching query language.
- A SPARQL query consists of three parts:
  - Pattern matching: optional, union, nesting, filtering.
  - Solution modifiers: projection, distinct, order, limit, offset.
  - Output part: construction of new triples, …

```
SELECT ?Name ?Email
WHERE
{
?X :hasName ?Name
?X :hasEmail ?Email
} LIMIT 10
```

```
{ P1 P2 }
{ P3 P4 }

{ P1
P2
OPTIONAL { P5 } }

{ P3
P4
OPTIONAL { P7
OPTIONAL { P8 } } }
```

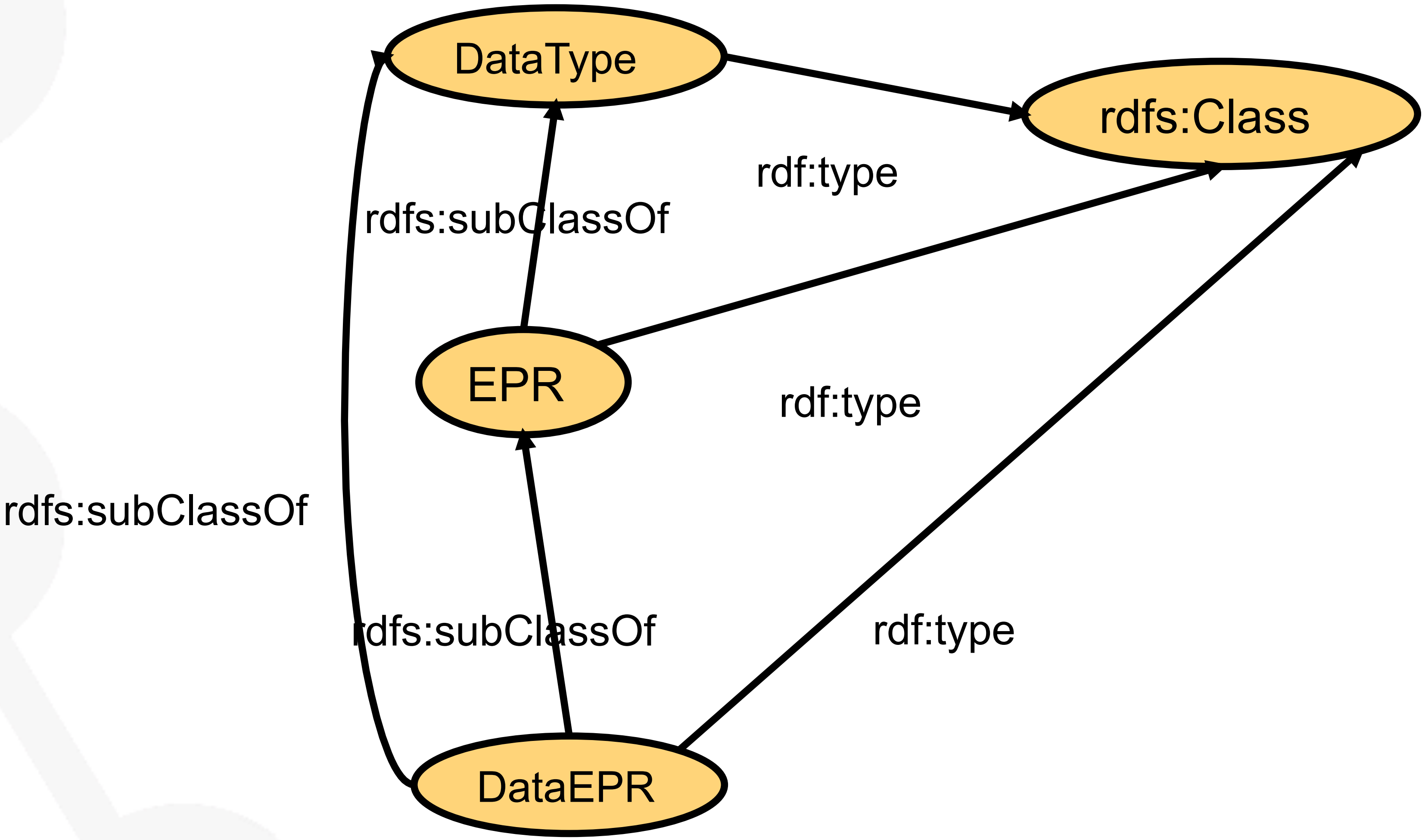
```
{P1} UNION
{ P9 }

{ P9
FILTER ( R ) }
```

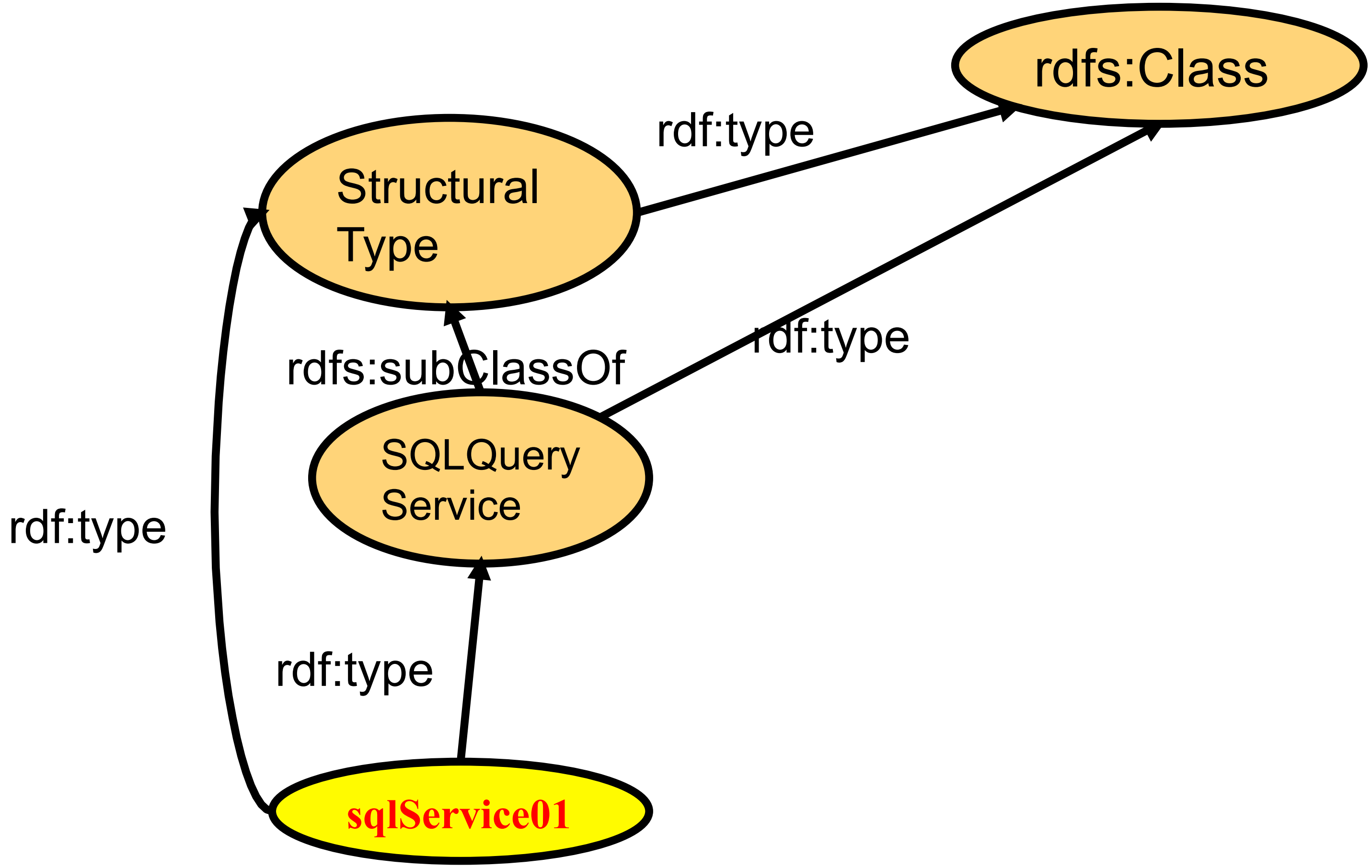
Table of Contents

RDFS: RDF Schema

- RDF Schema is another W3C Recommendation
  - http://www.w3.org/TR/rdf-schema/
- It extends RDF with a schema vocabulary that allows you to define basic vocabulary terms and the relations between those terms
  - Class, type, subClassOf,
  - Property, subPropertyOf, range, domain
  - it gives "extra meaning" to particular RDF predicates and resources
  - this "extra meaning", or semantics, specifies how a term should be interpreted
- The combination of RDF and RDF Schema is normally known as RDF(S)


DataType
rdfs:Class
rdf:type
rdfs:subClassOf
EPR
rdf:type
rdfs:subClassOf
rdfs:subClassOf
rdf:type
DataEPR

RDF(S) Inference



The RDF data model

```
<owl:Class rdf:about="#SQLQueryService">
    <rdfs:subClassOf rdf:resource="&CRISP-DMIOntology;StructuralType"/>   </
owl:Class>
```

Related Technologies (RDF(S))

- RDF(S) Gives
  - Ability to use simple schema/vocabularies to describe our resources
  - **Consistent** vocabulary use and **sharing**
  - Simple **inference**
  - **Query mechanisms**
- RDF(S) does not give
  - No localised range and domain constraints
  - No existence/cardinality constraints
  - No transitive, inverse or symmetrical properties

Table of Contents

OWL Basics (on top of RDF and RDFS)

- Set of constructors for concept expressions
  - Booleans: and/or/not
    - A DataResource is a RelationalResource or a RDFResource
  - Quantification: some/all
    - *RDFDataResources* can only have RDF*Data* that has data format *XML-RDF* or *N-Triples*
- Axioms for expressing constraints
  - Necessary and Sufficient conditions on classes
    - A *DataResource* that *hasDataFormat* XML-*RDF* is a *RDFRepository*.
  - Disjointness
    - *XML-RDF* is disjoint with *N-Triples*
  - Property characteristics: transitivity, inverse

Reasoning Tasks

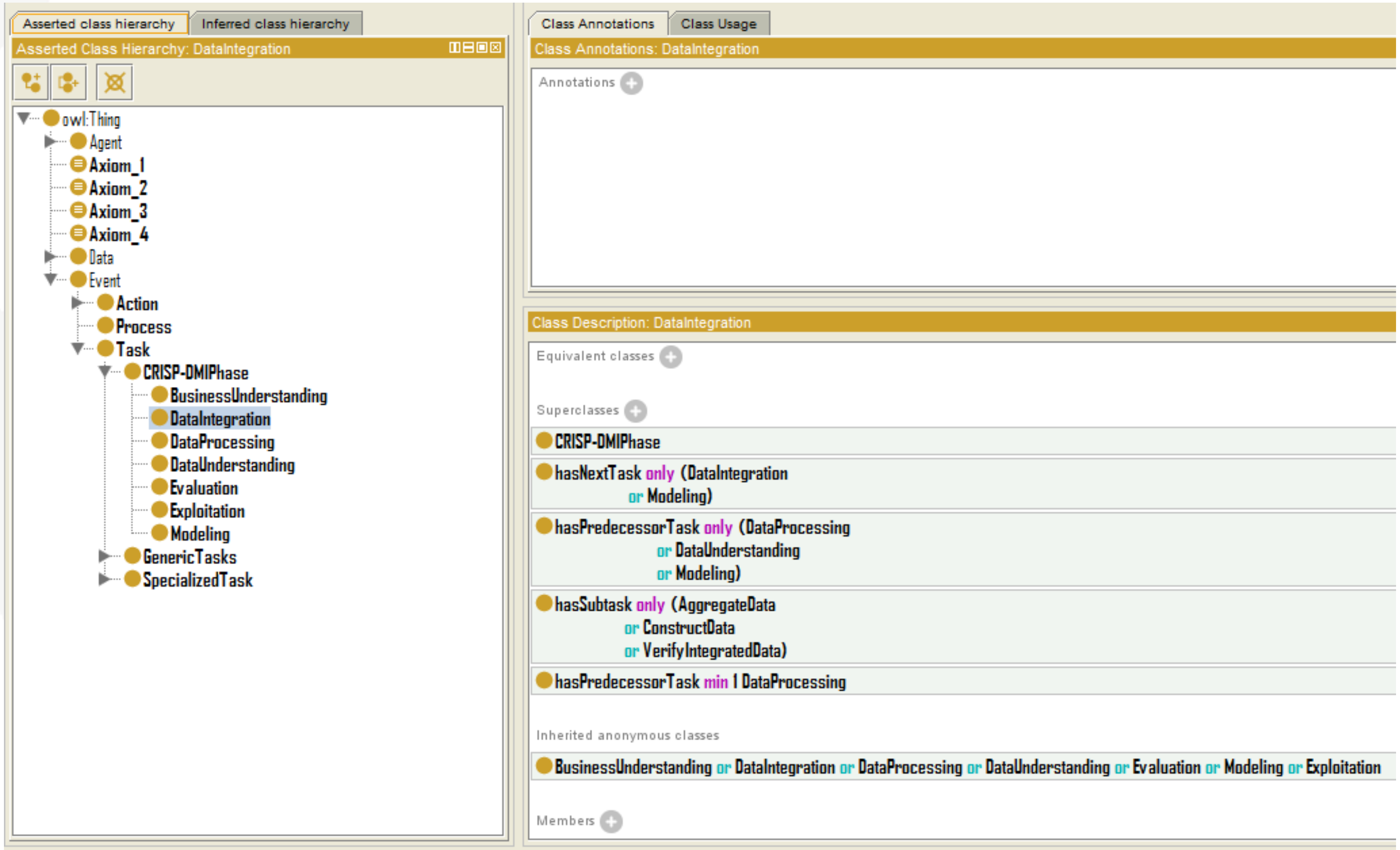
- OWL DL based on a well understood Description Logic ($SHOIN(D_n)$)
  - Formal properties well understood (complexity, decidability)
  - Known reasoning algorithms
  - Implemented systems (highly optimised)

- Because of this, we can reason about OWL ontologies
  - Subsumption reasoning
    - Allows us to infer when one class is a subclass of another
    - Can then build concept hierarchies representing the taxonomy.
  - Satisfiability reasoning
    - Tells us when a concept is unsatisfiable
      - i.e. when it is impossible to have instances of the class.
    - Allows us to check whether our model is consistent.
  - Instance Retrieval/Instantiation
    - What are the instances of a particular class C?
    - What are the classes that x is an instance of?

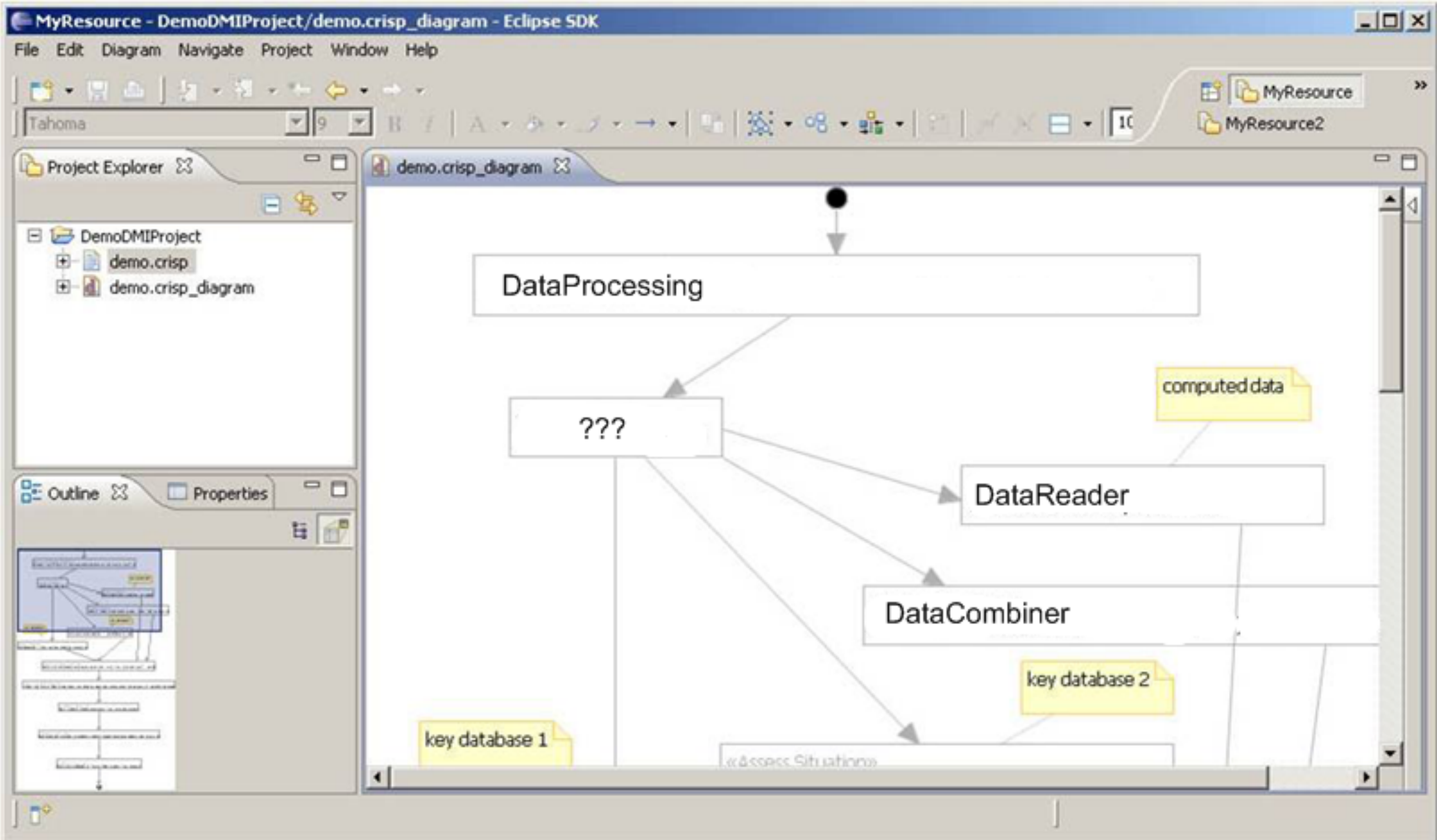
Table of Contents

- What is the Semantic Web?
- Related Technologies
  - RDF
  - RDF(S)
  - OWL
    - Inference Examples
    - Consistency Checking Examples
- Linked Data
- OGSA-DAI & DQP Extensions
- Conclusions

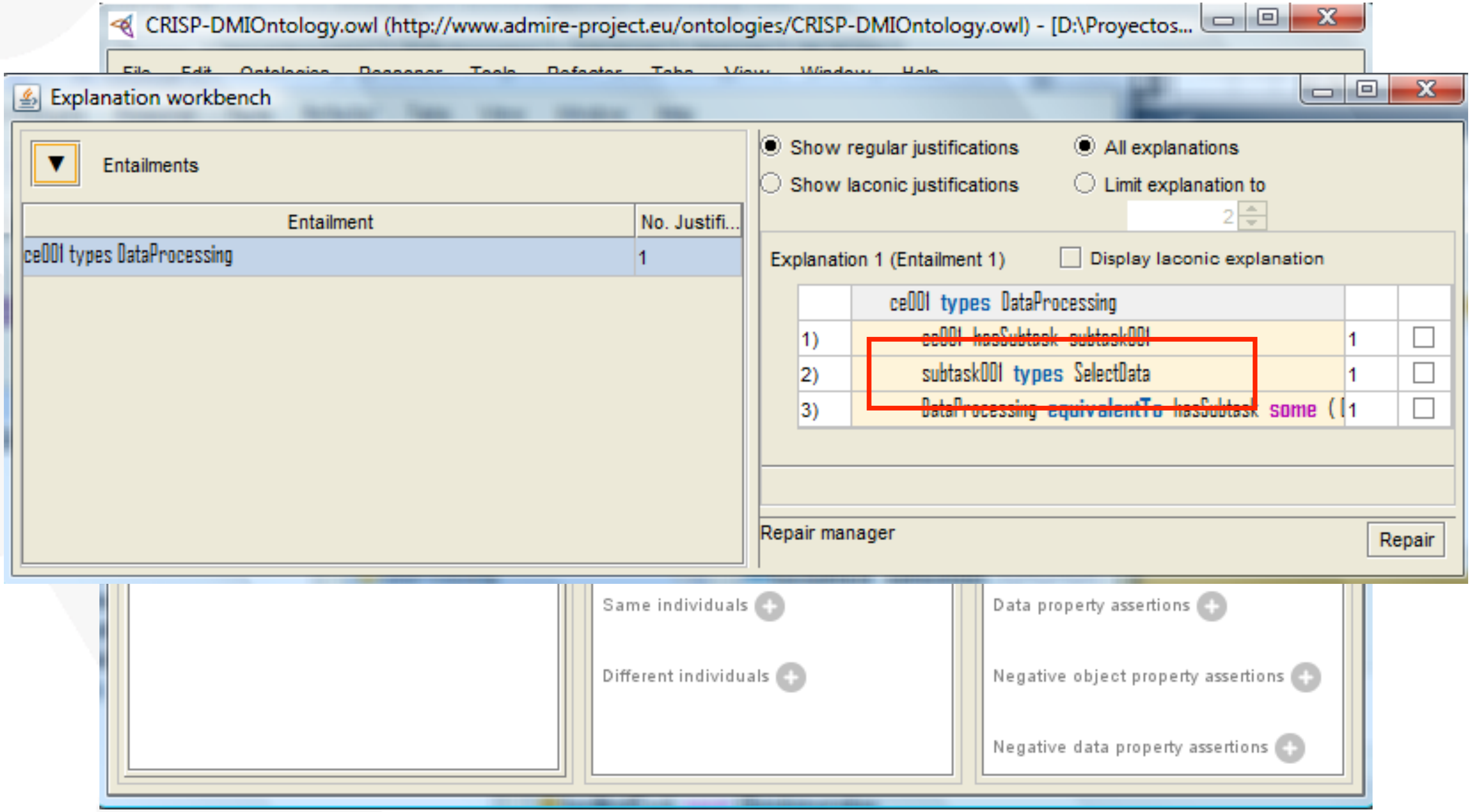
ADMIRE Ontology: CRISP-DMI Ontology



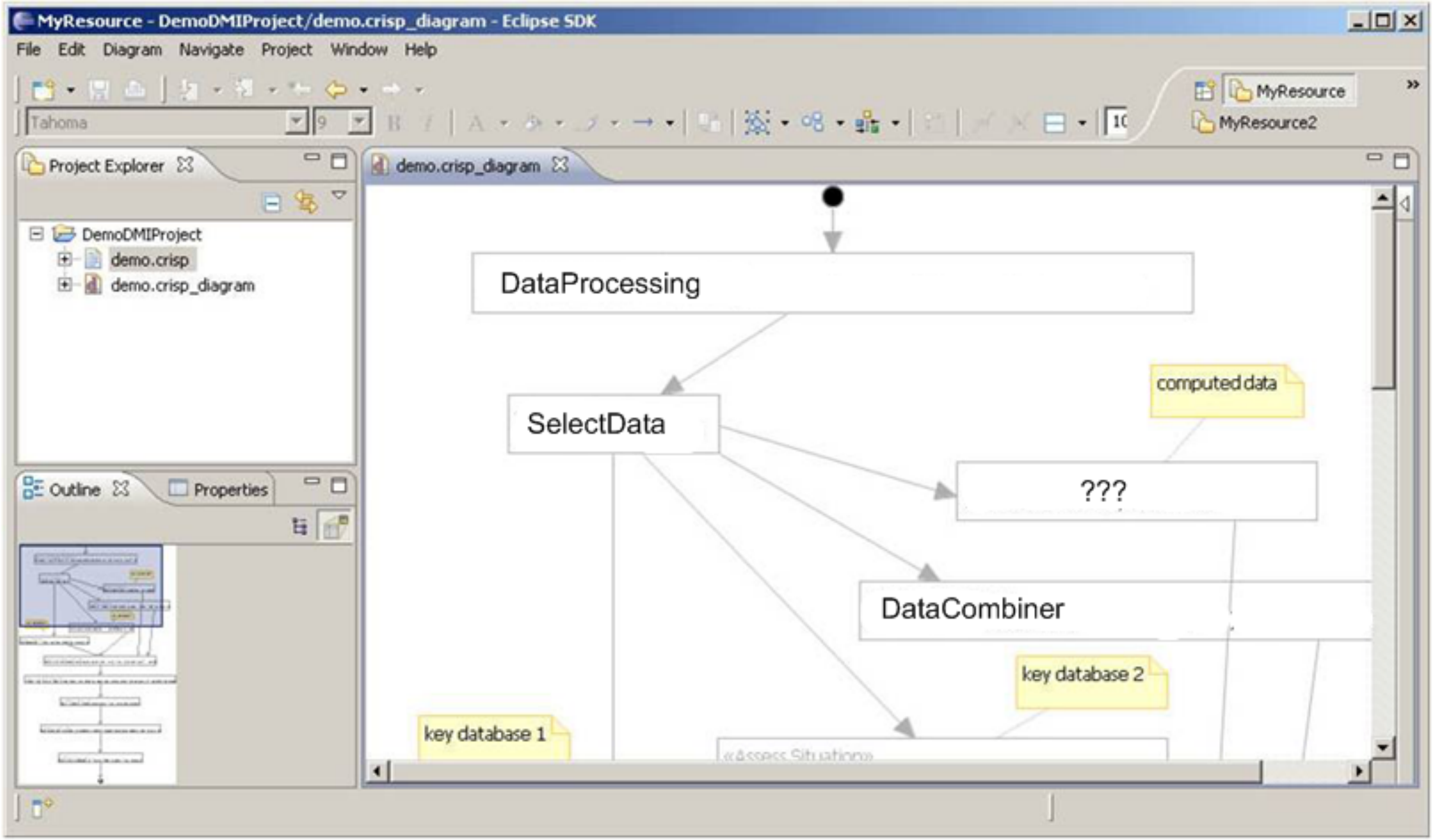
Inference


MyResource - DemoDMIProject/demo.crisp_diagram - Eclipse SDK
File Edit Diagram Navigate Project Window Help
Tahoma
MyResource
MyResource2
Project Explorer
DemoDMIProject
demo.crisp
demo.crisp_diagram
Outline
Properties
demo.crisp_diagram
DataProcessing
???
computed data
DataReader
DataCombiner
key database 2
key database 1

Inference

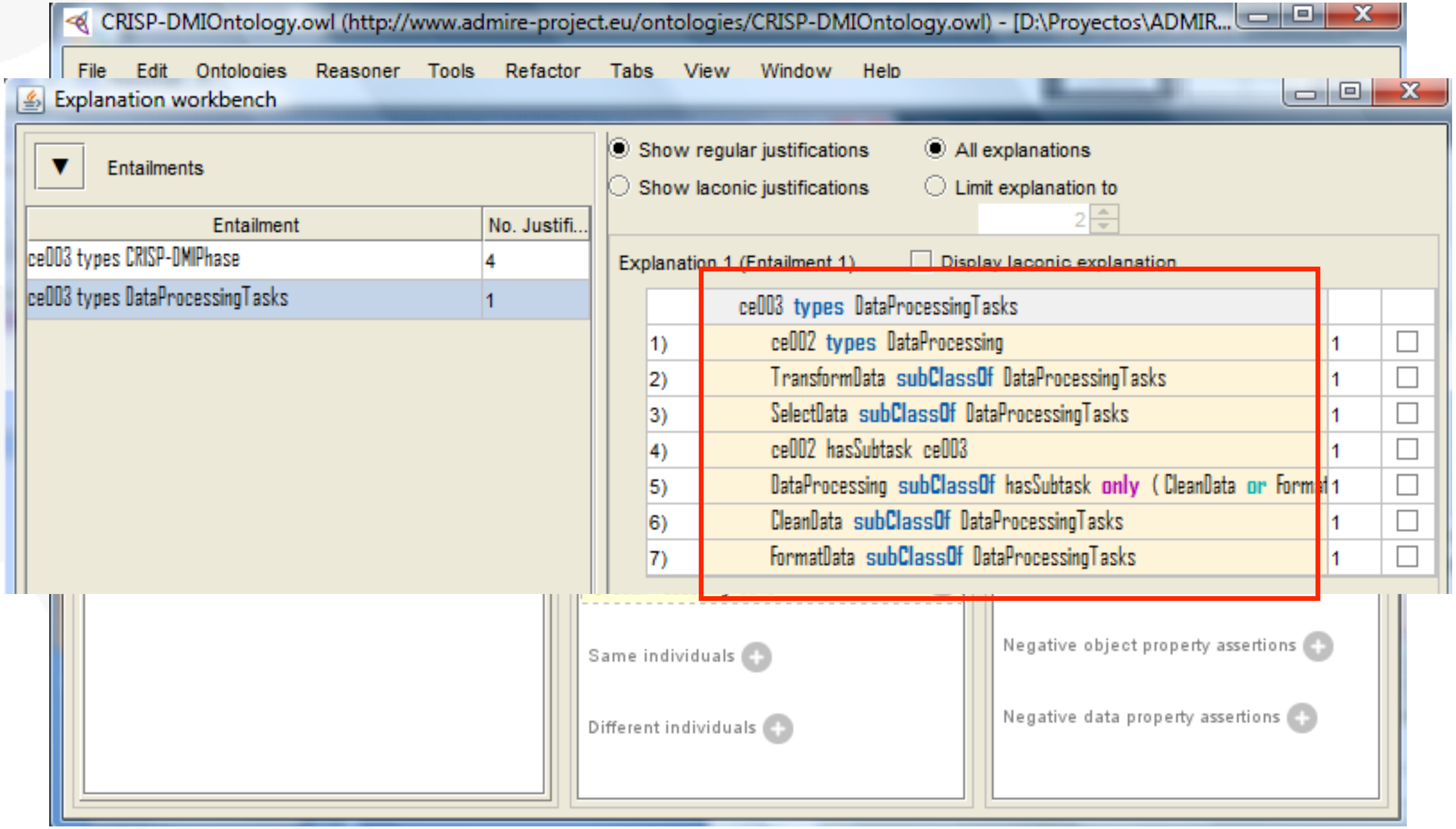

CRISP-DMIOntology.owl (http://www.admire-project.eu/ontologies/CRISP-DMIOntology.owl) - [D:\Proyectos...
Explanation workbench
Entailments
Entailment
No. Justifi...
ce001 types DataProcessing
1
Show regular justifications
Show laconic justifications
All explanations
Limit explanation to
2
Explanation 1 (Entailment 1)
Display laconic explanation
ce001 types DataProcessing
1) ce001 hasSubtask subtask001
2) subtask001 types SelectData
3) DataProcessing equivalentTo hasSubtask some (
Repair manager
Repair
Same individuals
Different individuals
Data property assertions
Negative object property assertions
Negative data property assertions

Inference



MyResource - DemoDMIProject/demo.crisp_diagram - Eclipse SDK
File Edit Diagram Navigate Project Window Help
Tahoma
MyResource
MyResource2
Project Explorer
DemoDMIProject
demo.crisp
demo.crisp_diagram
Outline
Properties
demo.crisp_diagram
DataProcessing
SelectData
computed data
???
DataCombiner
key database 2
key database 1

Inference


CRISP-DMIOntology.owl (http://www.admire-project.eu/ontologies/CRISP-DMIOntology.owl) - [D:\Proyectos\ADMIR...
File Edit Ontologies Reasoner Tools Refactor Tabs View Window Help
Explanation workbench
Entailments
Entailment
No. Justifi...
ce003 types CRISP-DMIPhase 4
ce003 types DataProcessingTasks 1
Show regular justifications
Show laconic justifications
All explanations
Limit explanation to
2
Explanation 1 (Entailment 1)
Display laconic explanation
ce003 types DataProcessingTasks
1) ce002 types DataProcessing 1
2) TransformData subClassOf DataProcessingTasks 1
3) SelectData subClassOf DataProcessingTasks 1
4) ce002 hasSubtask ce003 1
5) DataProcessing subClassOf hasSubtask only (CleanData or Format 1
6) CleanData subClassOf DataProcessingTasks 1
7) FormatData subClassOf DataProcessingTasks 1
Same individuals
Different individuals
Negative object property assertions
Negative data property assertions

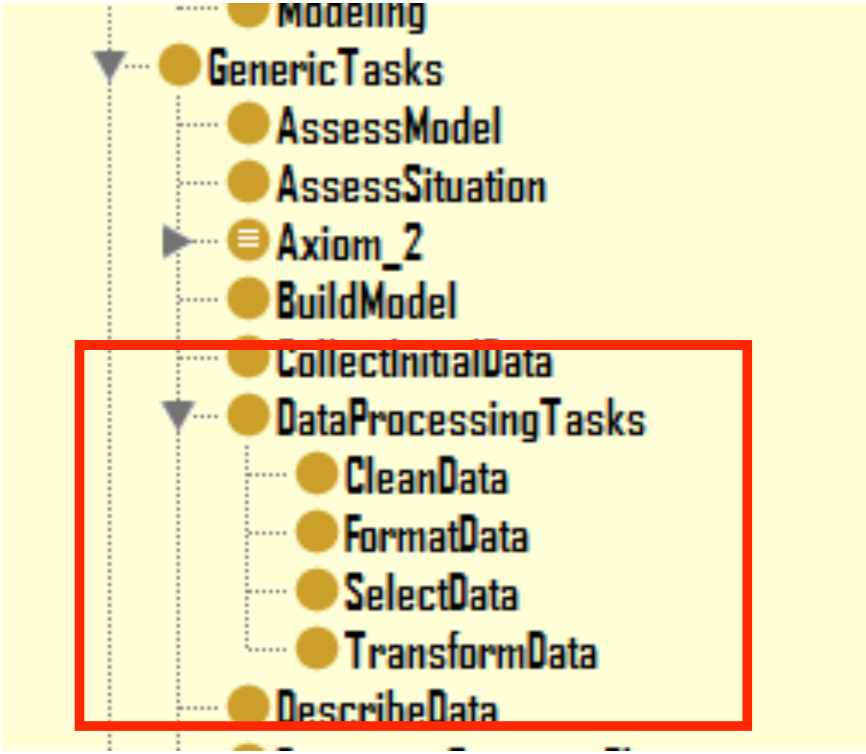
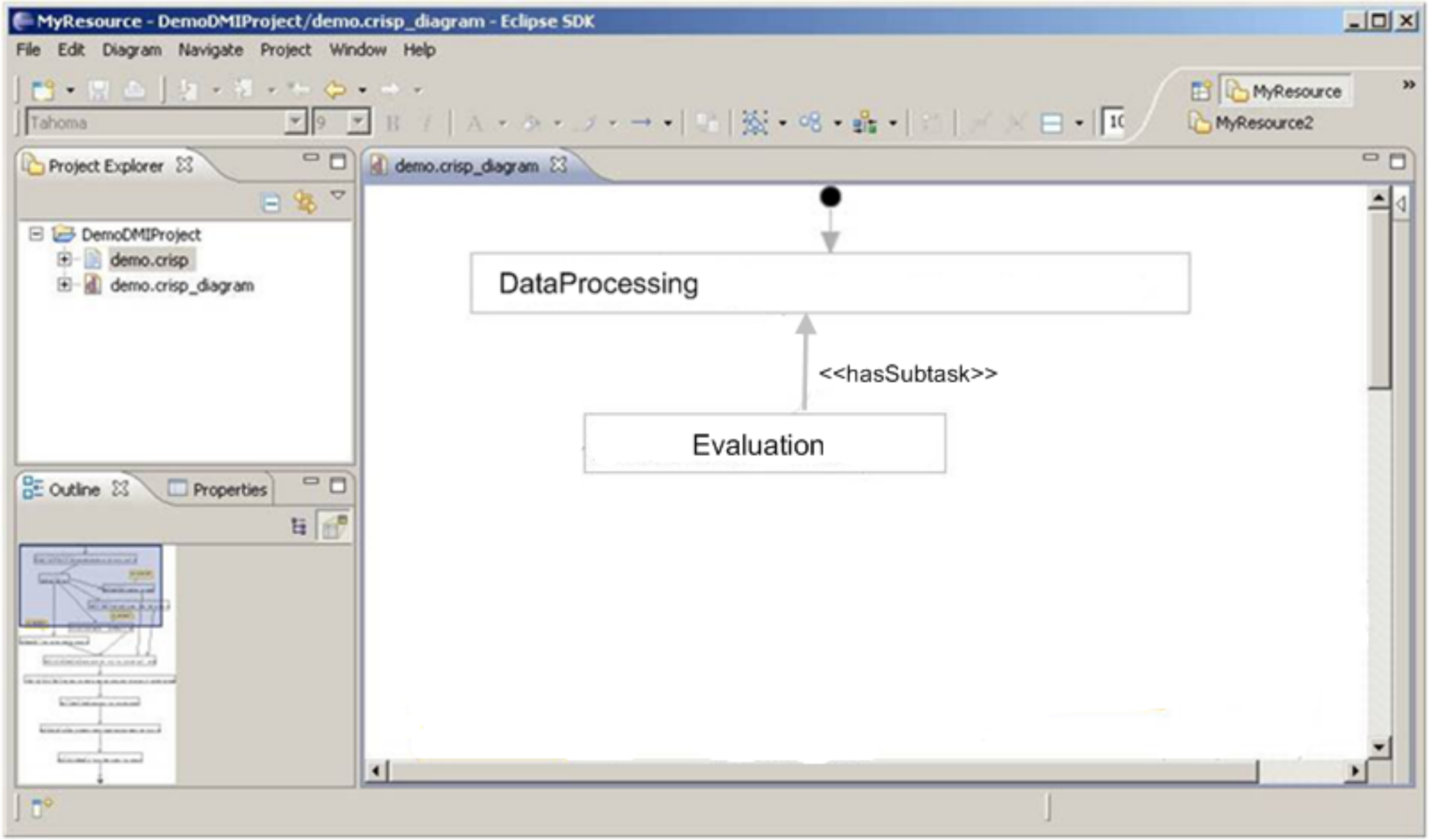

Modeling
GenericTasks
AssessModel
AssessSituation
Axiom_2
BuildModel
CollectInitialData
DataProcessingTasks
CleanData
FormatData
SelectData
TransformData
DescribeData

Table of Contents

- What is the Semantic Web?
- Related Technologies
  - RDF
  - RDF(S)
  - OWL
    - Inference Examples
    - Consistency Checking Examples
- Linked Data
- OGSA-DAI & DQP Extensions
- Conclusions

Consistency Checking



Consistency Checking

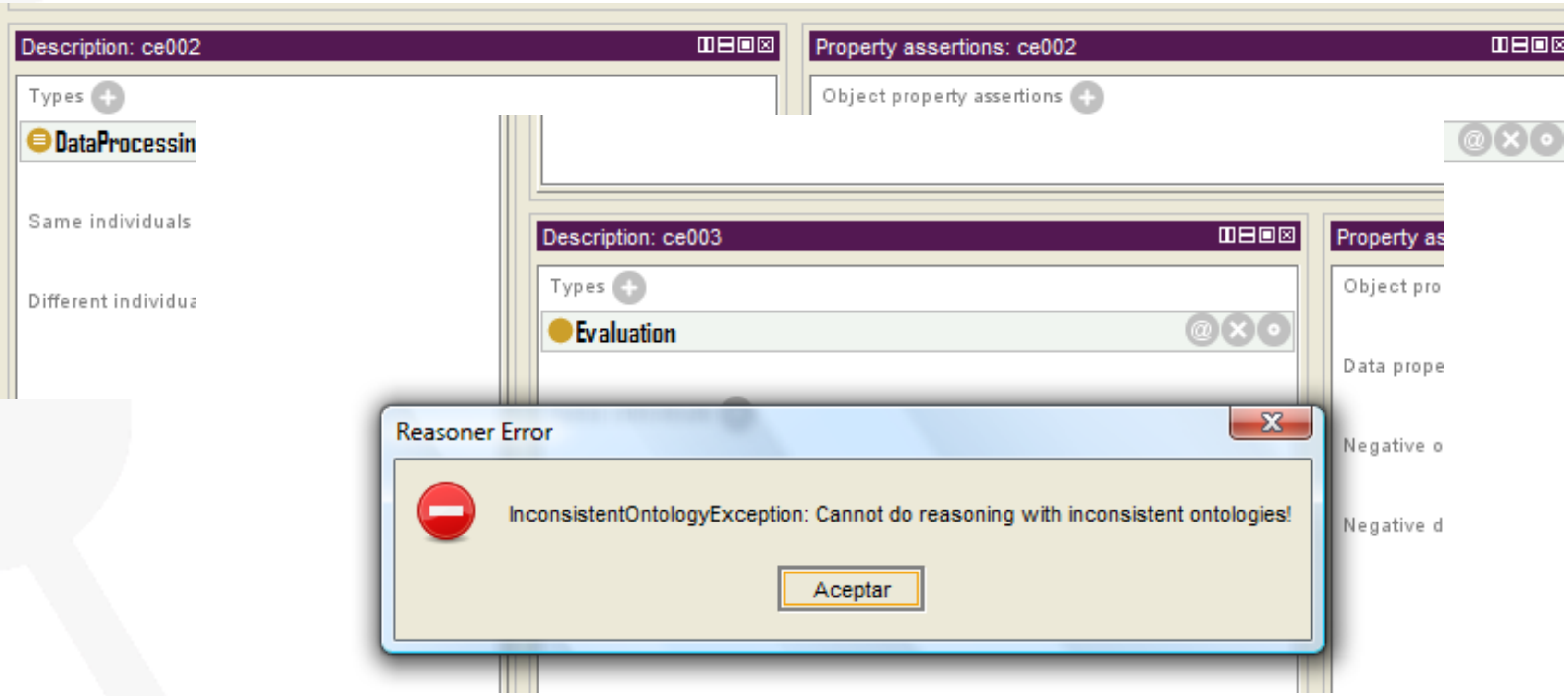
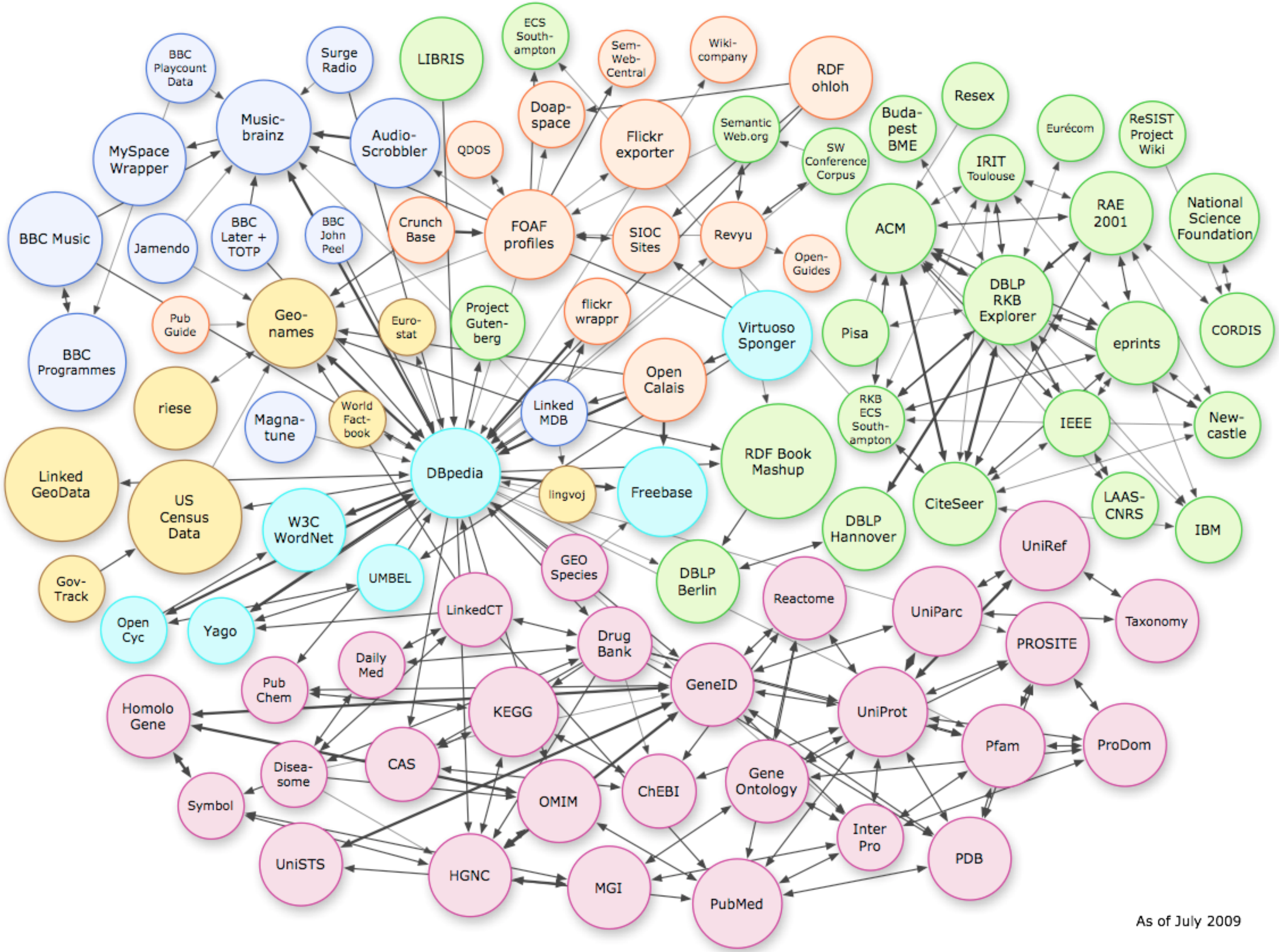


Table of Contents

- “Linked Data is about using the Web to connect related data that wasn't previously linked”
- Linked Data design principles:
  - Use URIs to identify things that you expose to the Web as resources.
  - Use HTTP URIs so that people can locate and look up (dereference) these things.
  - Provide useful information about the resource when its URI is dereferenced.
  - Include links to other, related URIs in the exposed data as a means of improving information discovery on the Web.

Linked Data



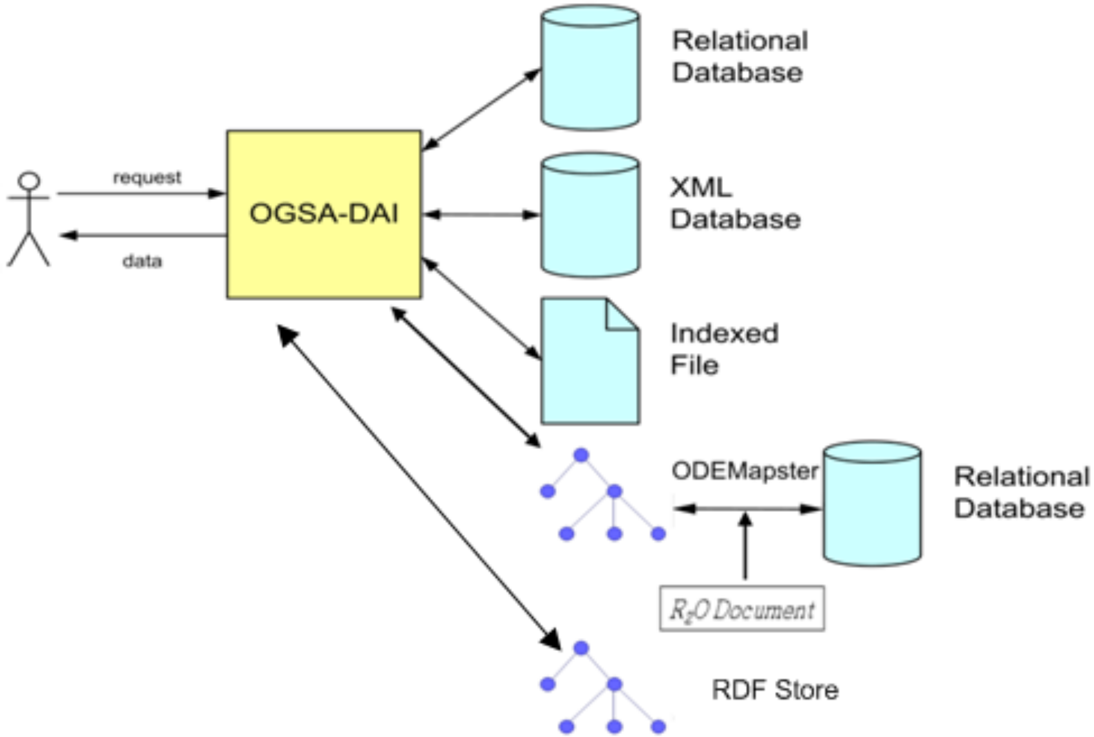
As of July 2009

Linked Data Example

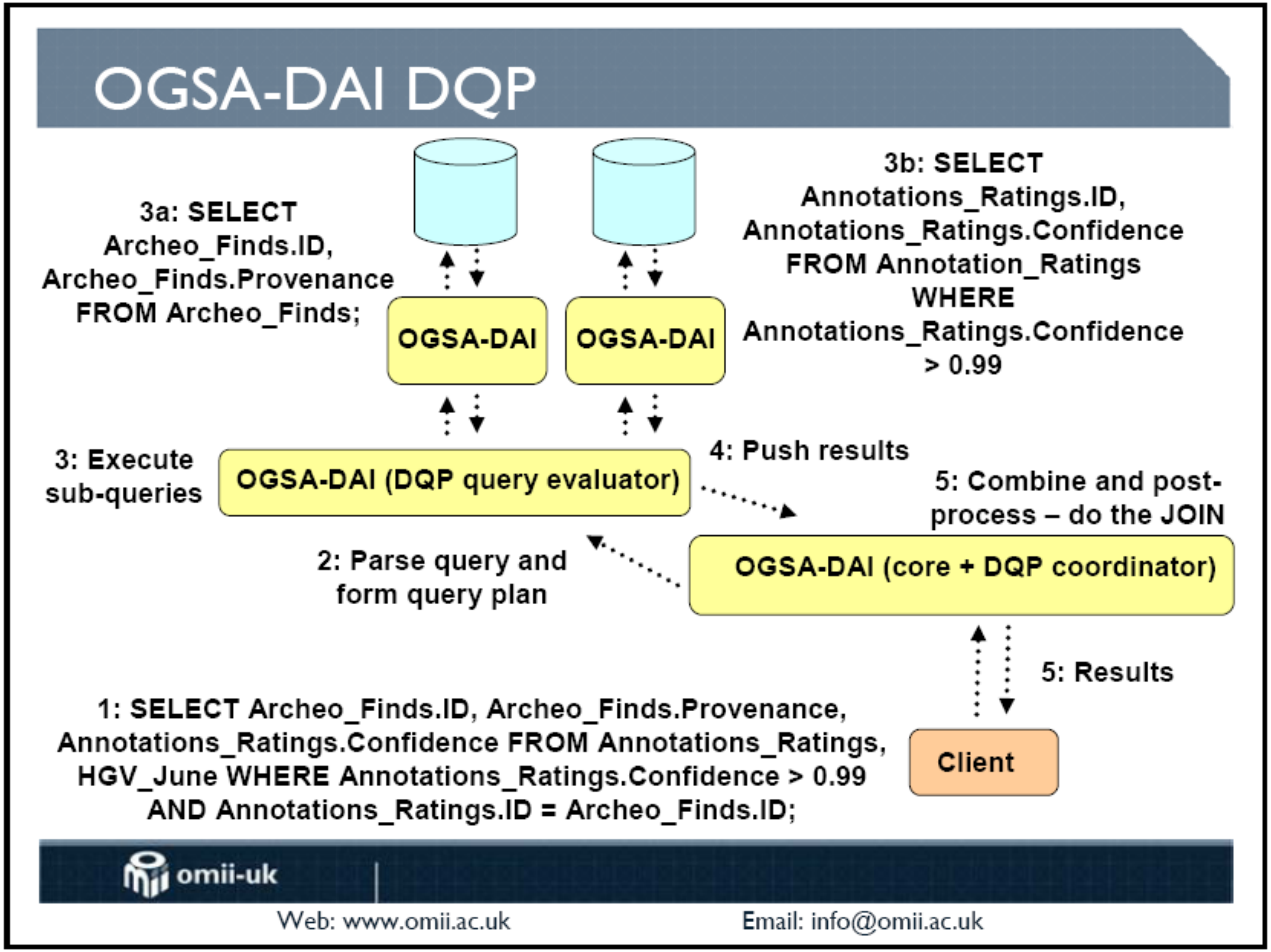
- Example D2R:
  - http://www4.wiwiss.fu-berlin.de/factbook/resource/Russia (URI identifying the non-information resource Russia)
  - http://www4.wiwiss.fu-berlin.de/factbook/data/Russia (information resource with an RDF/XML representation describing Russia)
  - http://www4.wiwiss.fu-berlin.de/factbook/page/Russia (information resource with an HTML representation describing Russia)
- DBPedia
  - http://dbpedia.org/page/Berlin
  - http://dbpedia.org/resource/Berlin
  - http://dbpedia.org/data/Berlin

Table of Contents

Extensions to OGSA-DAI & DQP



OGSA-DQP



OGSA-DAI DQP
3a: SELECT Archeo_Finds.ID, Archeo_Finds.Provenance FROM Archeo_Finds;
3b: SELECT Annotations_Ratings.ID, Annotations_Ratings.Confidence FROM Annotation_Ratings WHERE Annotations_Ratings.Confidence > 0.99
OGSA-DAI
OGSA-DAI
3: Execute sub-queries
OGSA-DAI (DQP query evaluator)
4: Push results
5: Combine and post-process – do the JOIN
2: Parse query and form query plan
OGSA-DAI (core + DQP coordinator)
5: Results
1: SELECT Archeo_Finds.ID, Archeo_Finds.Provenance, Annotations_Ratings.Confidence FROM Annotations_Ratings, HGV_June WHERE Annotations_Ratings.Confidence > 0.99 AND Annotations_Ratings.ID = Archeo_Finds.ID;
Client
omii-uk
Web: www.omii.ac.uk
Email: info@omii.ac.uk

Extensions to OGSA-DAI & DQP

- Currently very poor (or no existing) federation and optimization methods to RDF repositories
  - DARQ [1]
  - Networked Graphs [2]
- Proposal:
  - Extend OGSA-DQP with a new query language: SPARQL
  - SPARQL is "similar" to Relational Algebra
    - Both have the same expressive power
  - Current status: optimising simple SPARQL queries locally

[1] Querying Distributed RDF Data Sources with SPARQL Export, Bastian Quilitz, Ulf Leser. The Semantic Web: Research and Applications (2008), pp. 524-538
[2] Networked graphs: a declarative mechanism for SPARQL rules, SPARQL views and RDF data integration on the web. Simon Schenk and Steffen Staab. WWW '08

Table of Contents

Conclusions

- The Semantic Web is a set of technologies for
  - Adding metadata to resources
    - Data resources, services
  - Integration of data from different sources
  - These technologies allow to reason about data in different levels
    - Basic taxonomy
    - Inference, consistency checking
    - Makes computers more “intelligent”
- Based on standards
- Current hot topic in the community:
  - Linked data
  - SPARQL
- Semantic Web community does not pay much attention to:
  - Distributed Processing problems
  - Database problems
  - Proposal: bring together all

Future Work

- Continue upgrading DQP
- Workflows using all the possible data resources from OGSA-DAI (RDF and Relational mainly)
- Use of relational optimisers in SPARQL
- Study the possibility of using RDF in scientific workflows?
  - Use RDF for scientific data is currently not possible [3]
  - RDF query mechanisms still not optimised
  - RDF stores are significantly slower than databases
- The use of RDF and semantic technologies might be useful for annotating Processing Elements

[3] Gray, Alasdair J. G.; Gray, Norman & Ounis, Iadh: Can RDB2RDF Tools Feasibly Expose Large Science Archives for Data Integration?, ESWC 2009, S. 491-505

Semantic Web & related technologies

Carlos Buil Aranda

Ontology Engineering Group
Facultad de Informática
Universidad Politécnica de Madrid
cbuil@fi.upm.es

NeSC, 25th September 2009