Freie Universität Bozen
Libera Università di Bolzano
Università Liedia de Bulsan
Free University of Bozen · Bolzano

Web-based Graphical Querying of Databases through an Ontology: the WONDER System

Diego Calvanese, C. Maria Keet, Werner Nutt,
Mariano Rodríguez-Muro, **Giorgio Stefanoni**

Free University of Bozen-Bolzano

ACM Symposium on Applied Computing
Sierre (SUI), 23.03.2010

AGENDA

1. Extracting Data from Web-accessible Databases
   - ❖ Issues with standard interfaces
   - ❖ Extending data extraction with OBDA
2. WONDER: Web-based Graphical Querying
   - ❖ Architecture
   - ❖ WONDER Paradigm
3. Evaluation

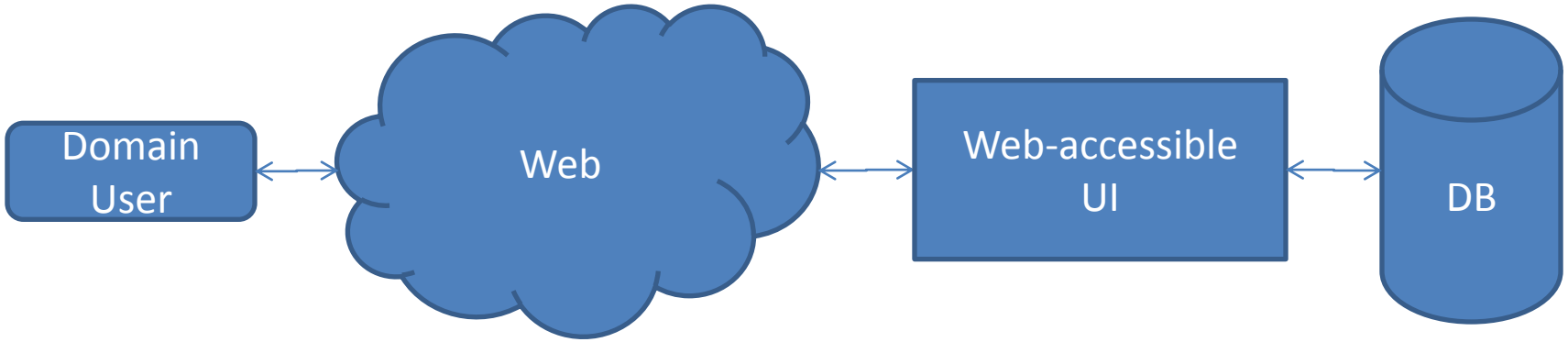
FREIE UNIVERSITÄT BOZEN
LIBERA UNIVERSITÀ DI BOLZANO
UNIVERSITÀ LIEDIA DE BULSAN
FREE UNIVERSITY OF BOZEN · BOLZANO

AGENDA

1. **Extracting Data from Web-accessible Databases**
   - ❖ **Issues with standard interfaces**
   - ❖ **Extending data extraction with OBDA**
2. WONDER: Web-based Graphical Querying
   - ❖ Architecture
   - ❖ WONDER Paradigm
3. Evaluation

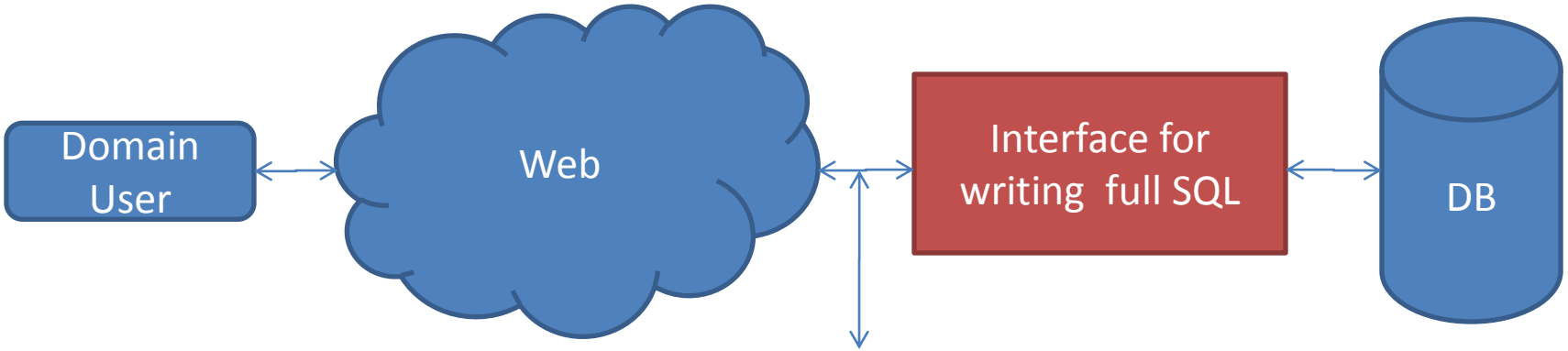
Accessing Databases

1. **Extracting Data from Web-accessible Databases**



Accessing Databases

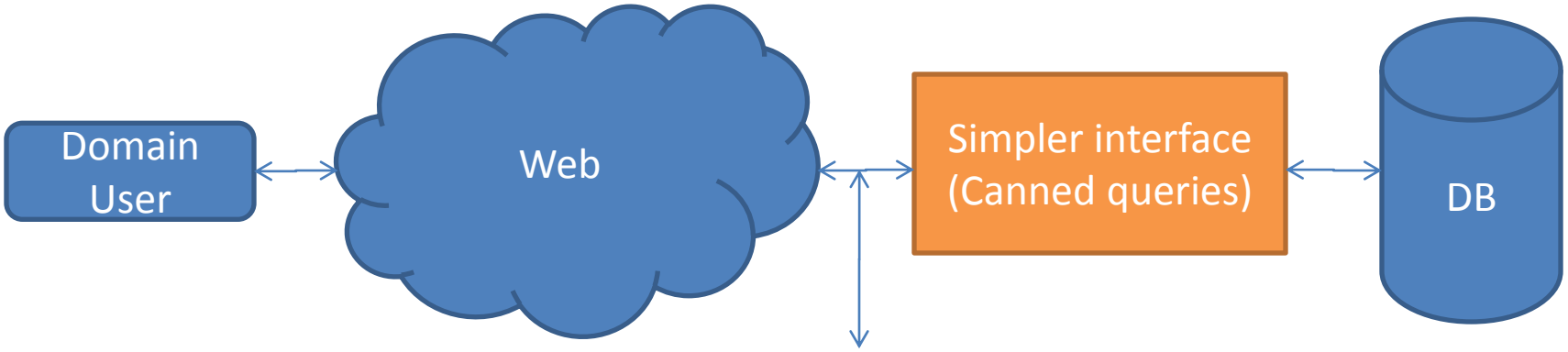
1. **Extracting Data from Web-accessible Databases**



- FAILS TO MEET THE USABILITY REQUIREMENT + MAXIMAL QUERY EXPRESSIVENESS

Accessing Databases

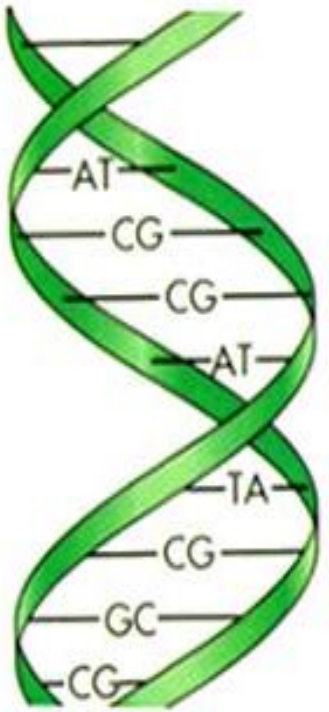
1. **Extracting Data from Web-accessible Databases**



+ MEETS THE USABILITY REQUIREMENT
- NOT ALL QUERIES CAN BE EXPRESSED

Horizontal Gene Transfer DB

- Lots of data made available on the Web by the Life Science field
- Is a web-accessible **genomic database** about prokaryotic organisms
- Contains 477 organisms and 1,445,840 genes
- 4 GB database instance



Sample Information Request

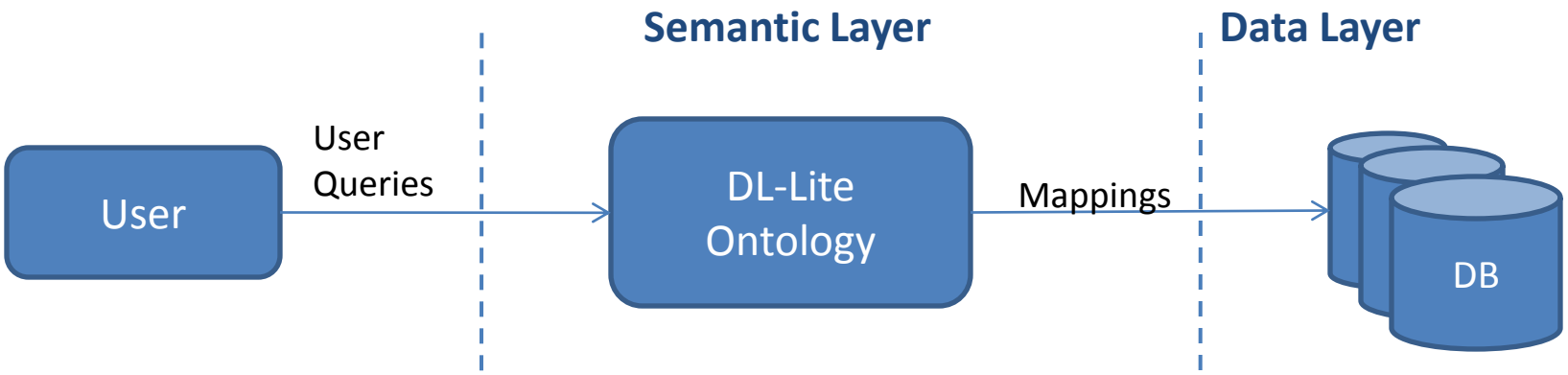
Retrieve all genes of the organisms Neisseria for which horizontal gene transfer is predicted or have a GC3 value > 80

Current Interface

- Offers the possibility to pose **canned queries** and to retrieve text-files of pre-computed queries through a simple text-based HTML interface
- Substantial limitations on expressiveness of queries
  - Domain users **can not extract all the information contained in the database**!

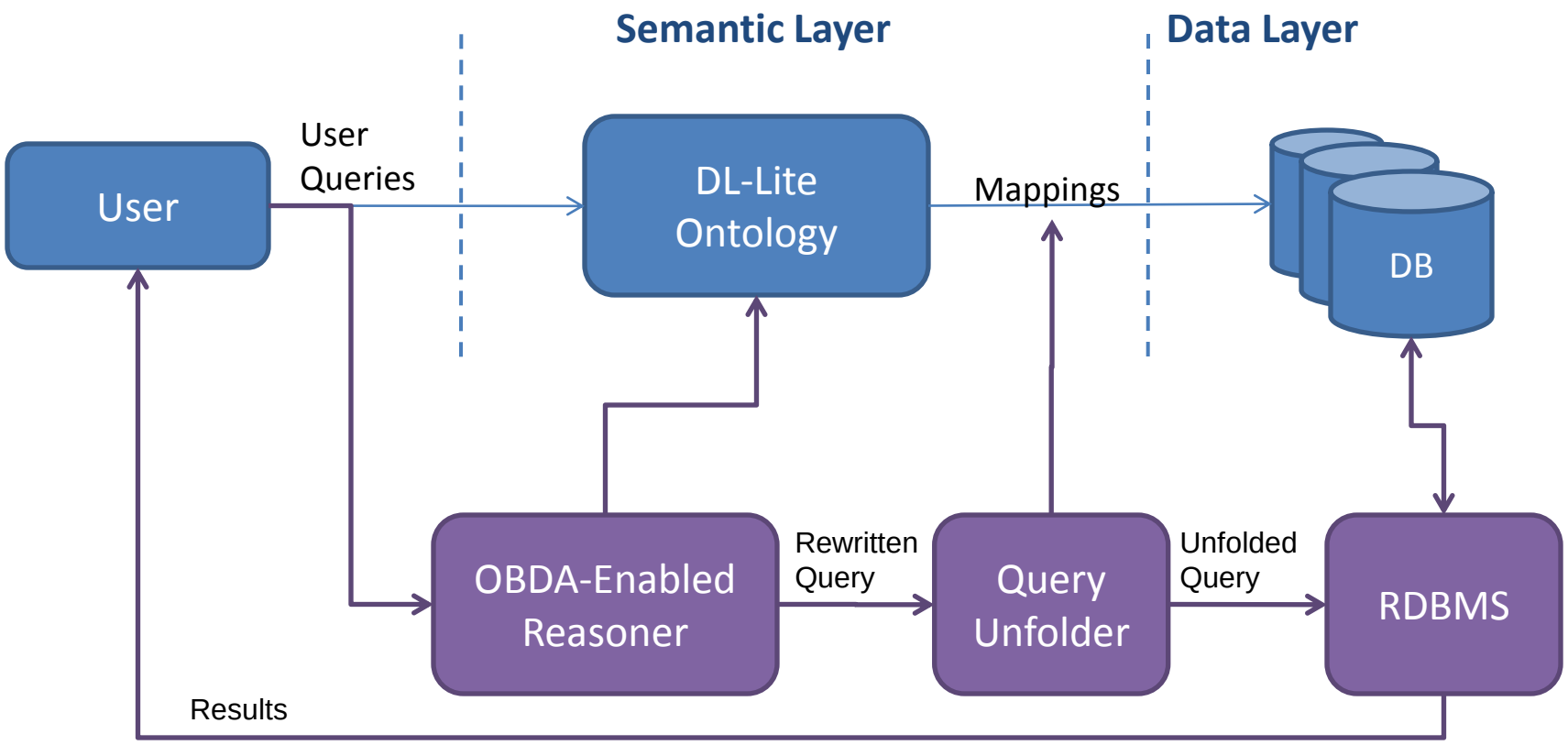
Ontology-Based Data Access

1. **Extracting Data from Web-accessible Databases**



Ontology-Based Data Access

1. **Extracting Data from Web-accessible Databases**



Union of Conjunctive Queries

- A union of conjunctive queries (UCQ) is a disjunction of CQs:

$$q(\vec{x}) \leftarrow \exists \vec{y}.conj(\vec{x}, \vec{y})$$

Atoms: $D(z) \quad S(z, w) \quad z = w$

- CQs corresponds to relational algebra select-project-join (SPJ) queries.

Querying through OBDA

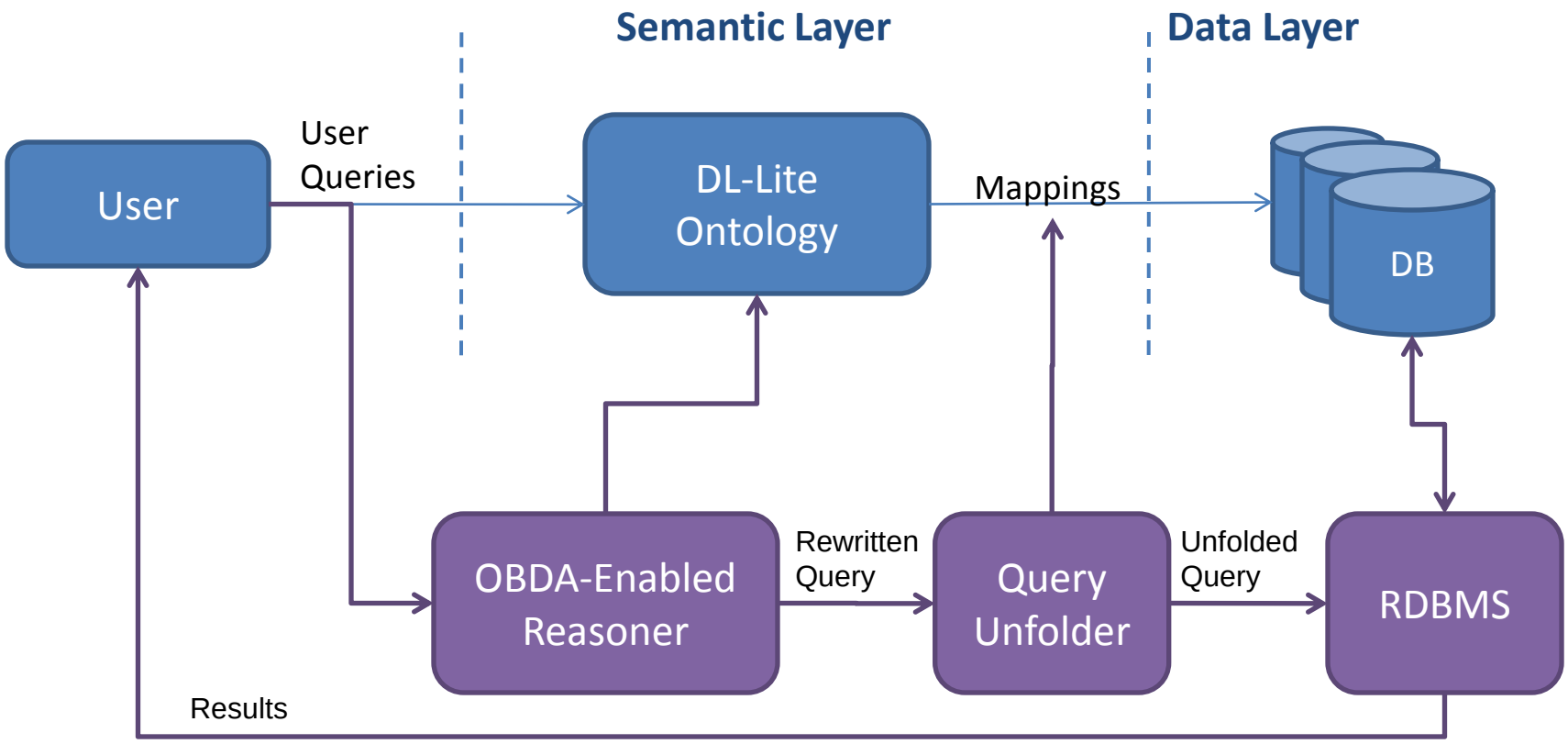
```
SELECT stbl.gene
FROM sparqltable
        (SELECT $gene $orgName $gcVal $predVal
        WHERE {$gene   :GeneHasOrganism $org.
                $org    :OrganismHasOrganismInfo $info.
                $info   :OrganismName $orgName.
                $gene   :GeneHasHGTPredictionGene $pred.
                $pred   :Prediction $predVal.
                $gene   :GeneHasGCstatsGene $gcstats.
                $gcstats :GC3 $gcVal}) stbl
WHERE stbl.orgName LIKE '%Neisseria%' AND
        (stbl.predVal = 'hgt' OR stbl.gcVal > '80')
```

Freie Universität Bozen
Libera Università di Bolzano
Università Liedia de Bulsan
Free University of Bozen · Bolzano

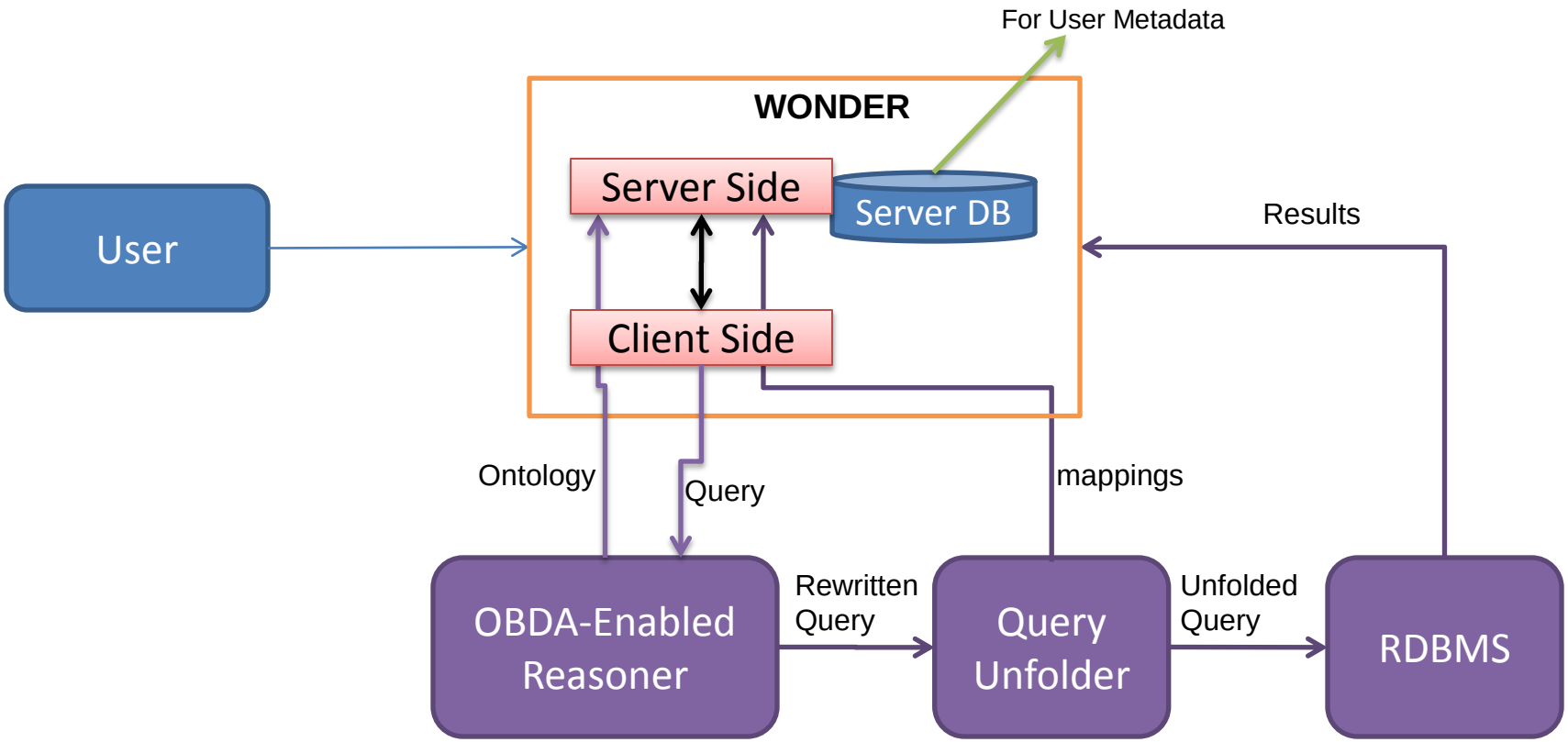
AGENDA

1. Extracting Data from Web-accessible Databases
   - ❖ Issues with standard interfaces
   - ❖ Extending data extraction with OBDA
2. **WONDER: Web-based Graphical Querying**
   - ❖ **Architecture**
   - ❖ **WONDER Paradigm**
3. Evaluation

OBDA Architecture



WONDER Architecture



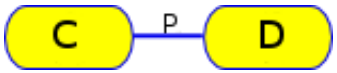
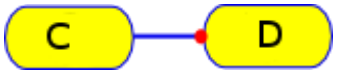
WONDER Paradigm

Accessing information comprises three activities:

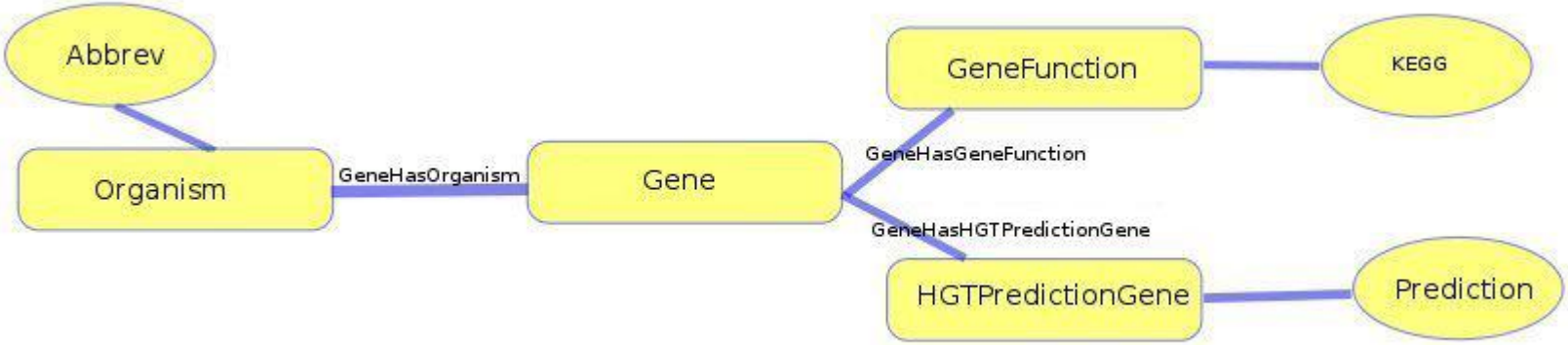
1. **Browsing the ontology**, to understand the structure of the information;
2. **Formulating a query**, to express an information request; and
3. **Retrieving data** that answers the query

The WONDER Web interface consists of a separate component for each of the activities.

Ontology Browser

Construct	Graphical Element	Semantic
Class		$C \sqsubseteq \top$
Object Property		$\exists P \sqsubseteq C$ $\exists P^- \sqsubseteq D$
Data Property		$\delta(A) \sqsubseteq C$ $\rho(A) \sqsubseteq \top_d$
SubClass Relationship		$C \sqsubseteq D$

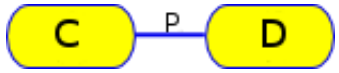
Ontology Browser (Cont.d)



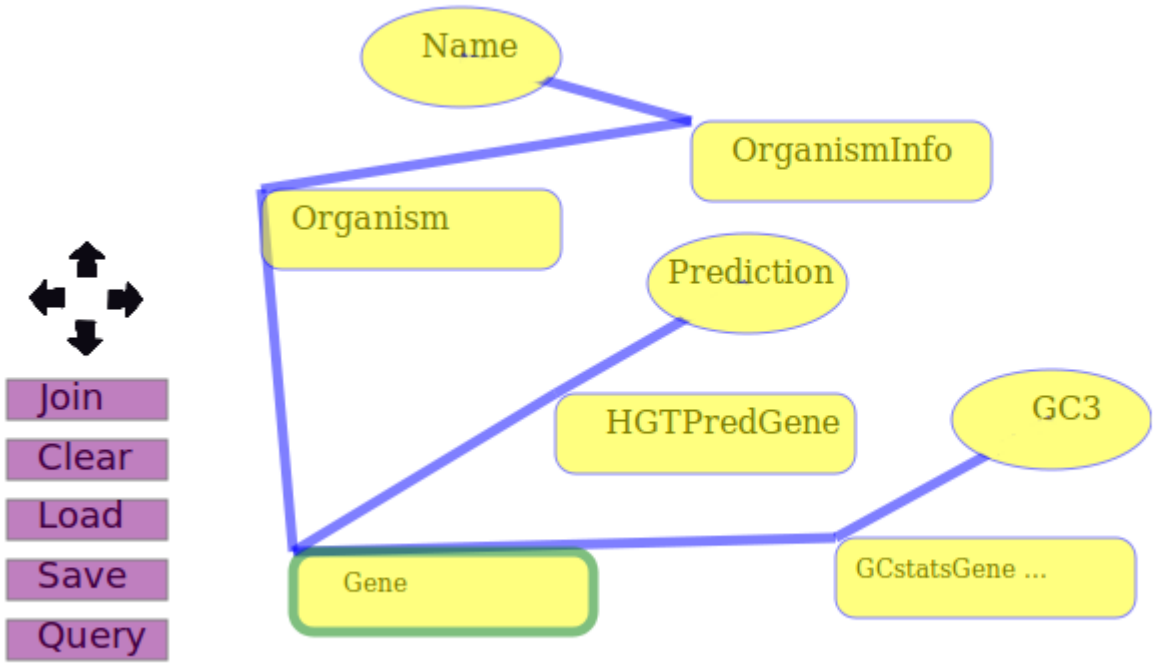
$\delta(Abbrev) \sqsubseteq Organism$	An abbreviation is for an organism
$\rho(Abbrev) \sqsubseteq xsd : string$	An abbreviation is of type string
$Organism \sqsubseteq \delta(Abbrev)$	Each organism has an abbreviation
$(\text{funct } Abbrev)$	Each individual has a single abbreviation
$\exists GeneHasOrganism \sqsubseteq Gene$	Domain of object property

....

Query Pane

Construct	Graphical Element	Semantic
Class node		$C(x), D(x)$
Object Property link		$C(x), P(x,y), D(y)$
Data Property node and link		$C(x), A(x,y)$

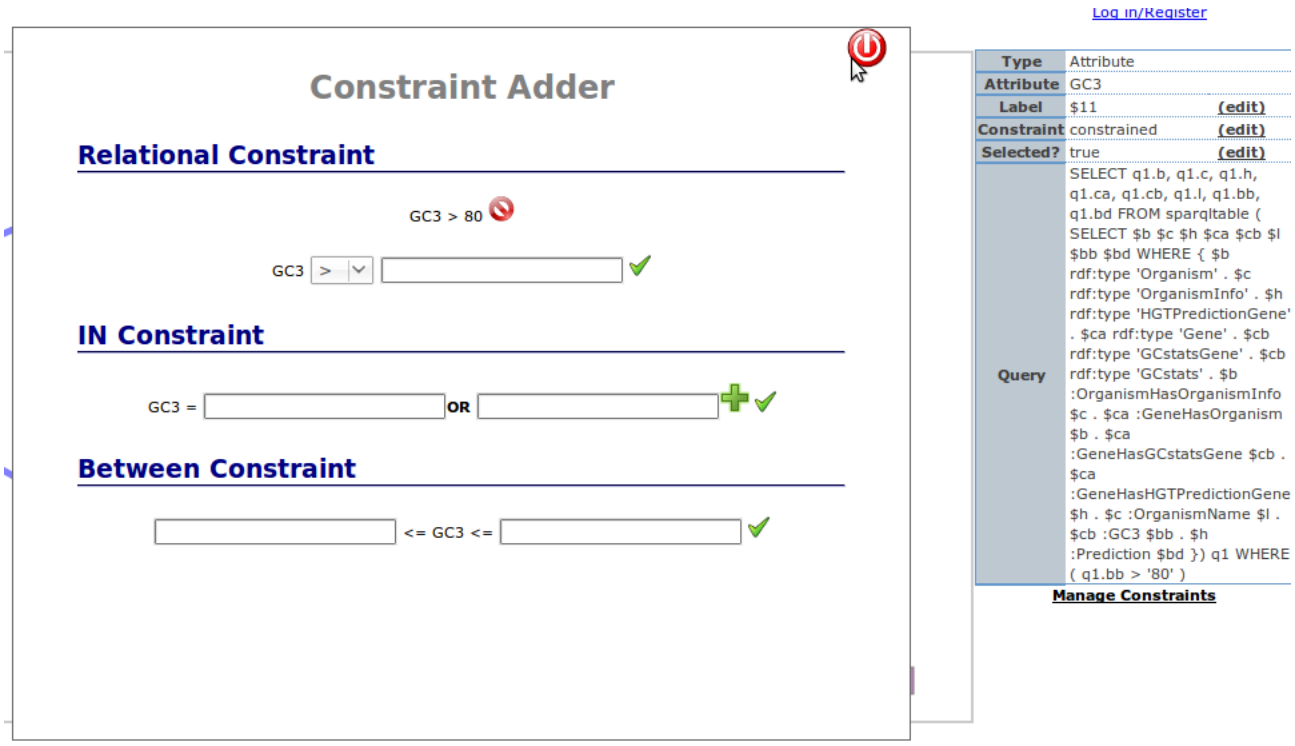
Query Pane (Cont.d)



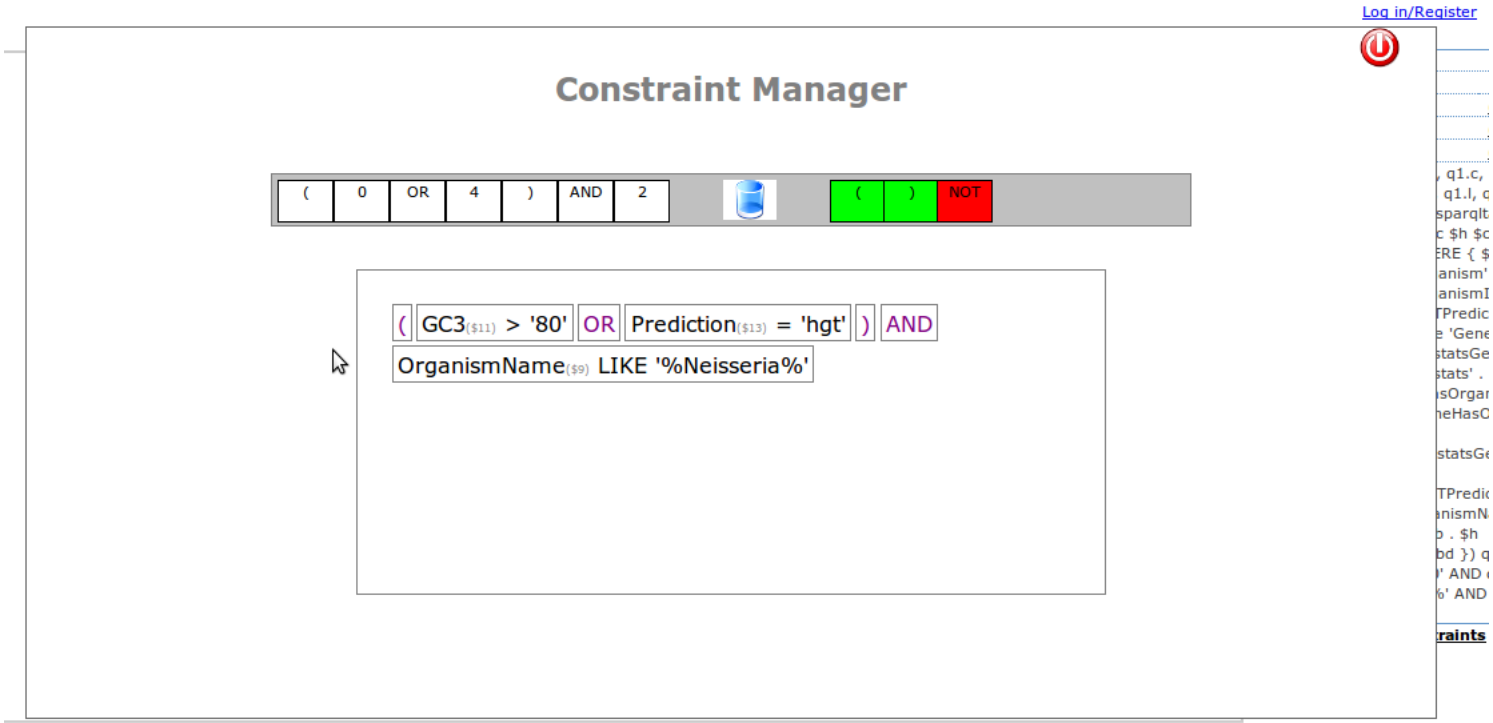
Type	Concept
Concept	OrganismInfo
Label	$3 **(edit)**
Selected?	false **(edit)**
Query	SELECT q1.cb FROM sparqltable (SELECT $cb $e $bb $bi WHERE { $c rdf:type 'Organism' . $d rdf:type 'OrganismInfo' . $bh rdf:type 'HGTPredictionGene' . $bf rdf:type 'GCstatsGene' . $bf rdf:type 'GCstats' . $cb rdf:type 'Gene' . $cb :GeneHasOrganism $c . $c :OrganismHasOrganismInfo $d . $cb :GeneHasGCstatsGene $bf . $cb :GeneHasHGTPredictionGene $bh . $d :OrganismName $e . $bf :GC3 $bb . $bh :Prediction $bi }) q1 WHERE (q1.e LIKE '%Neisseria%' AND (q1.bi = 'hgt' OR q1.bb > '80'))

Manage Constraints

Constraint Adder



Constraint Manager



Result Pane

Ontology Browser | Query Pane | Res

Hi Giorgio Stefanoni
Logout

This table contains 100/787 results
Please follow this link in order to download the entire result set: [Click here]

the Gene
getGene(ngon_1035)
getGene(ngon_1063)
getGene(ngon_1066)
getGene(ngon_1154)
getGene(ngon_1260)
getGene(ngon_1442)
getGene(ngon_1443)
getGene(ngon_1450)
getGene(ngon_1499)
getGene(ngon_1602)
getGene(ngon_1780)
getGene(ngon_208)
getGene(ngon_21)
getGene(ngon_23)
getGene(nmen1_1341)
getGene(nmen1_1680)
getGene(nmen1_1685)
getGene(nmen1_1692)
getGene(nmen1_1694)
getGene(nmen1_1700)
getGene(nmen1_1703)
getGene(nmen1_1797)
getGene(nmen1_1977)
getGene(nmen1_1978)
getGene(nmen1_2014)
getGene(nmen1_227)
getGene(nmen1_484)
getGene(nmen1_494)
getGene(nmen1_631)
getGene(nmen1_67)
getGene(nmen1_71)
getGene(nmen1_830)
getGene(nmen1_95)

AGENDA

1. Extracting Data from Web-accessible Databases
   - ❖ Issues with standard interfaces
   - ❖ Extending data extraction with OBDA
2. WONDER: Web-based Graphical Querying
   - ❖ Architecture
   - ❖ WONDER Paradigm
3. **Evaluation**

Query Formulation Results

- **Syntactical correctness of the query** is ensured by the formal foundation of the interface.
- Domain users have **more freedom** in constructing the queries and thanks to the query loading/saving feature, the overall service results more usable.
- While using the WONDER interface, domain users came up with **new queries** that are interesting for their studies.
- The **user is helped** in the formulation of complex constraints over the queries (Constraint Manager)

Technological Results

- The overhead caused by the graphical interface is negligible wrt. the standard OBDA setting.
- The approach result to be scalable enough to deal with pretty large database (> 4GB).
- Achieved seamless integration of different (Semantic) Web Technologies: OWL 2, AJAX, JavaScript, SVG and XSLT

AGENDA

1. Extracting Data from Web-accessible Databases
   - ❖ Issues with standard interfaces
   - ❖ Extending data extraction with OBDA
2. WONDER: Web-based Graphical Querying
   - ❖ Architecture
   - ❖ WONDER Paradigm
3. Evaluation

Conclusions

- To make data on the Web fully accessible to domain experts, query interfaces must go beyond forms
- Visual querying can bridge the gap between ontologies and the end user
- Web technology is now mature enough to support visual querying through a browser

Acknowledgment. We would like to thank: Dr. van Passel and Prof. Garcia-Vallvé, who have participated in the project.

Future Work

- **Improving the interface:**
  - Allow comparison between two different attribute values.
  - Extend the Query Pane to help the user to create the query by reasoning over the ontology,
- **Broaden our horizons**
  - Exploit OBDA infrastructure to cope with data integration and data incompleteness