4th International SWAT4LS Workshop
Semantic web applications and tools for life sciences

December 7-9th, 2011
London, UK

Bioqueries: A Social Community Sharing Experiences while Querying Biological Linked Data

(http://bioqueries.uma.es)

María Jesús García-Godoy, Ismael Navas-Delgado, José Francisco Aldana Montes

Computing Languages and Computing Science Department, University of Malaga, Malaga 29071, Spain

ismael@lcc.uma.es

Khaos INVESTIGACIÓN

LENGUAJES Y CIENCIAS DE LA COMPUTACIÓN UNIVERSIDAD DE MÁLAGA

GISUM

UNIVERSITAS MALACITANA

The Life Science Information

- Life Sciences have emerged as a key domain in the Web of Data.

- There is a big amount and diversity of semantic data and formats available by a variety of biological databases and Web technologies.

Linked Data Tools in the Life Sciences Domain

- Bio2RDF project is a Semantic Web Application designed to solve problems of integration of data.

Bio2RDF.org

Select an ID example | Submit

- Linked Life Data is an integration platform that stores billions of RDF statements that can be queried by the endpoint provided by the application. It is a good trial to close the Linked Data technology and biologists.

linked life data

a semantic data integration platform for the biomedical domain

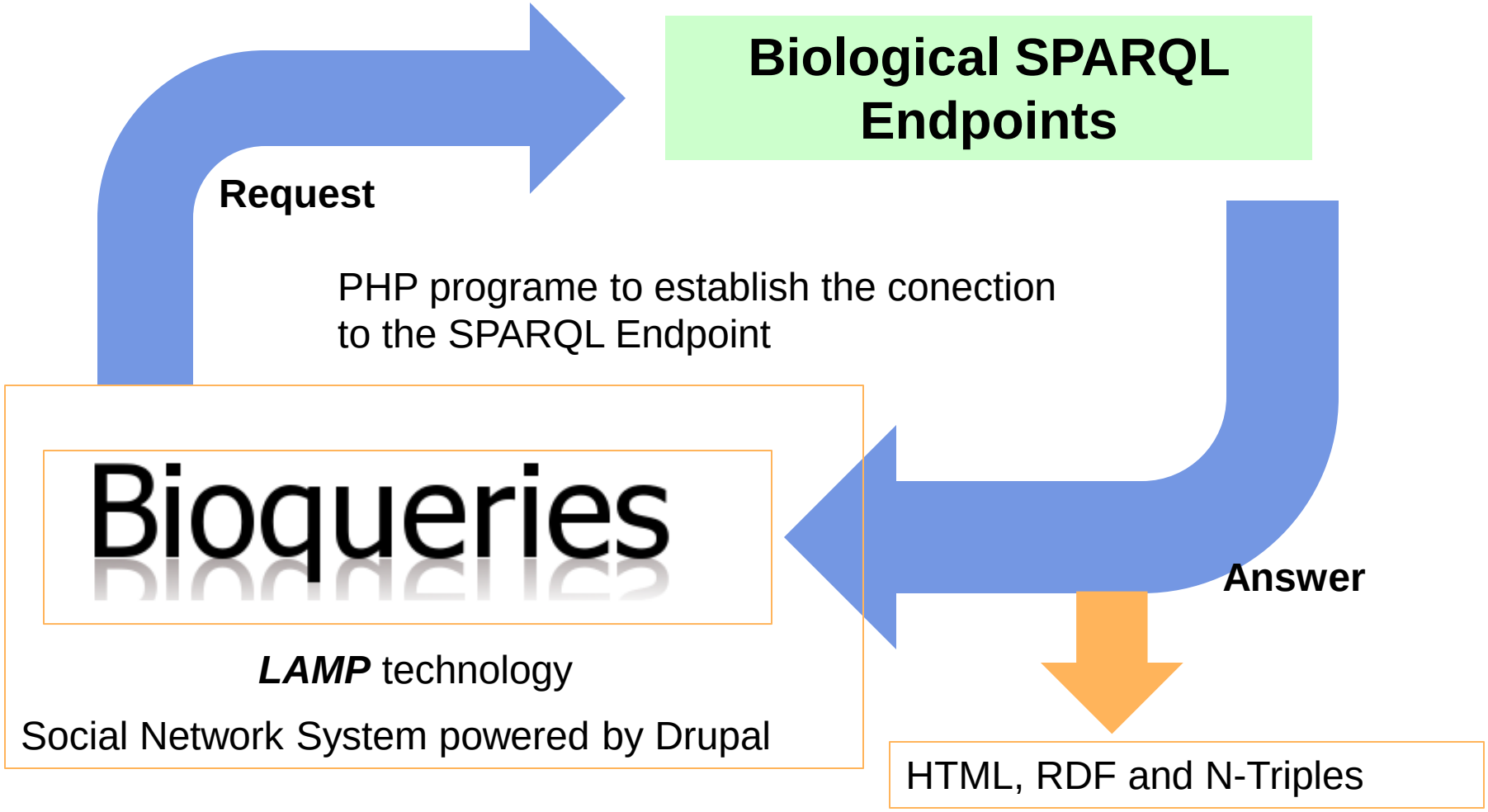
Bioqueries

Bioqueries

Bioqueries aim to help biological end-users to start using Linked Data

- A Web portal to design, share and execute SPARQL queries.
- Community building as a way to approximate Linked Data to biological users.
- Easy to use interfaces to take real advantages of SPARQL Endpoints that were designed to implement applications for the consumption and the reusability of Linked Data in the domain of the Life Sciences domain.

The Bioqueries Portal



Khaos

Community Building

- To make easier the use of this application we have included a set of 116 SPARQL queries as an initial seed for the community.
- New SPARQL Endpoints can be added by the users.
- The queries are classified manually into the following categories: Annotations, Biological process, Chemical compounds, Diseases, Enzimology, Gene expression, Genetic, Metabolism, Proteomic, Scientific texts and Taxonomy.

The Bioqueries Portal

- A Wiki-based portal developed extending Drupal with the execution of parameterized SPARQL queries. These queries are evaluated and the results of the execution of these queries are displayed hiding the RDF.

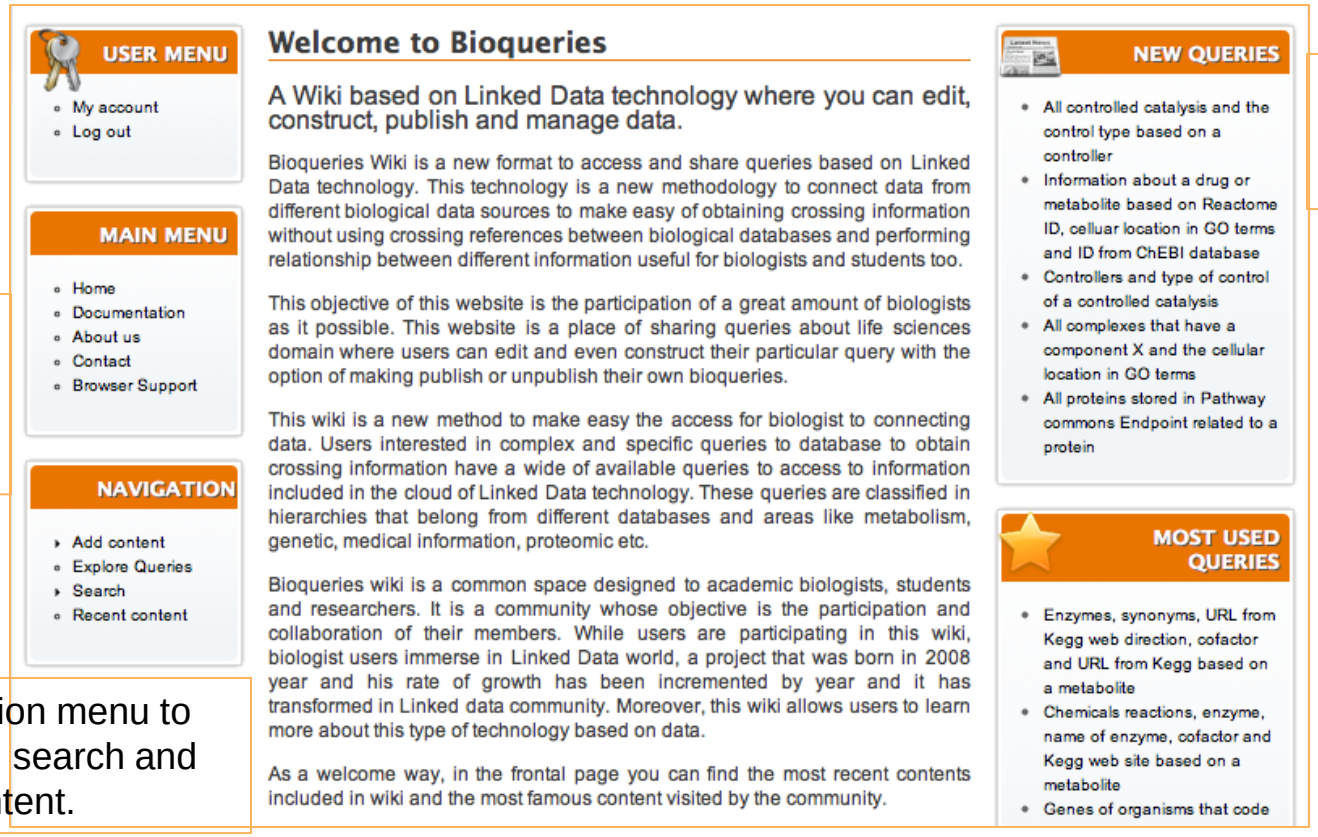
User menu for registration

Main Menu for documentation and contact information

Navigation menu to explore, search and add content.

New Queries introduced in the system

The most used queries



Khaos

http://bioqueries.uma.es

Registration

CREATE NEW ACCOUNT | LOG IN | REQUEST NEW PASSWORD

User account

TO FILL BY USER

Username *

Spaces are allowed; punctuation is not allowed except for periods, hyphens, apostrophes, and underscores.

E-mail address *

A valid e-mail address. All e-mails from the system will be sent to this address. The e-mail address is not made public and will only be used if you wish to receive a new password or wish to receive certain news or notifications by e-mail.

CREATE NEW ACCOUNT

Lux,

Thank you for registering at BioQueries. You may now log in by clicking this link or copying and pasting it to your browser:

http://bioqueries.uma.es/user/reset/52/1322756485/DUYbmgZm8waFkNEbwxUPuxpQBVTRE_KlbRguNRZEJbs

This link can only be used once to log in and will lead you to a page where you can set your password.

After setting your password, you will be able to log in at http://bioqueries.uma.es/user in the future using:

username: Lux
password: Your password

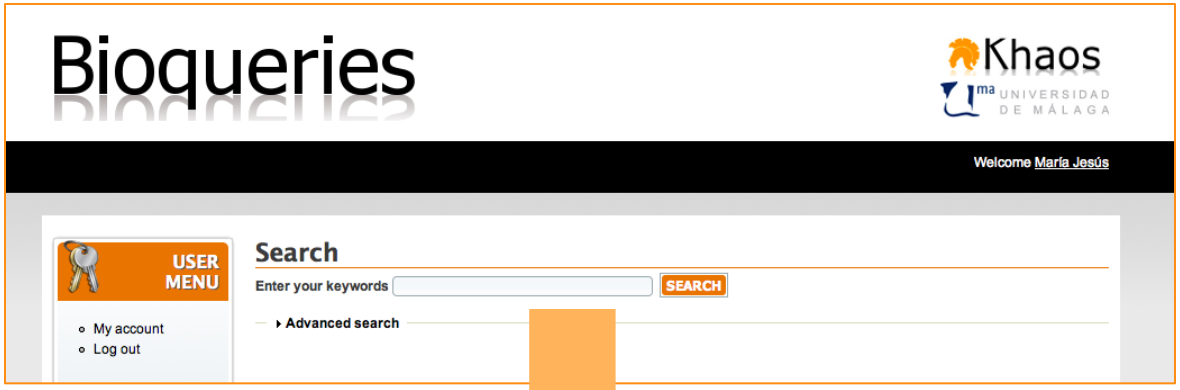
We are making an usability test of BioQueries system so, please, when you log in and use this application, do not forget to fill the following questionary (http://bioqueries.uma.es/sites/default/files/SUS.doc) and send it to bioqueries@lcc.uma.es. Thank you.

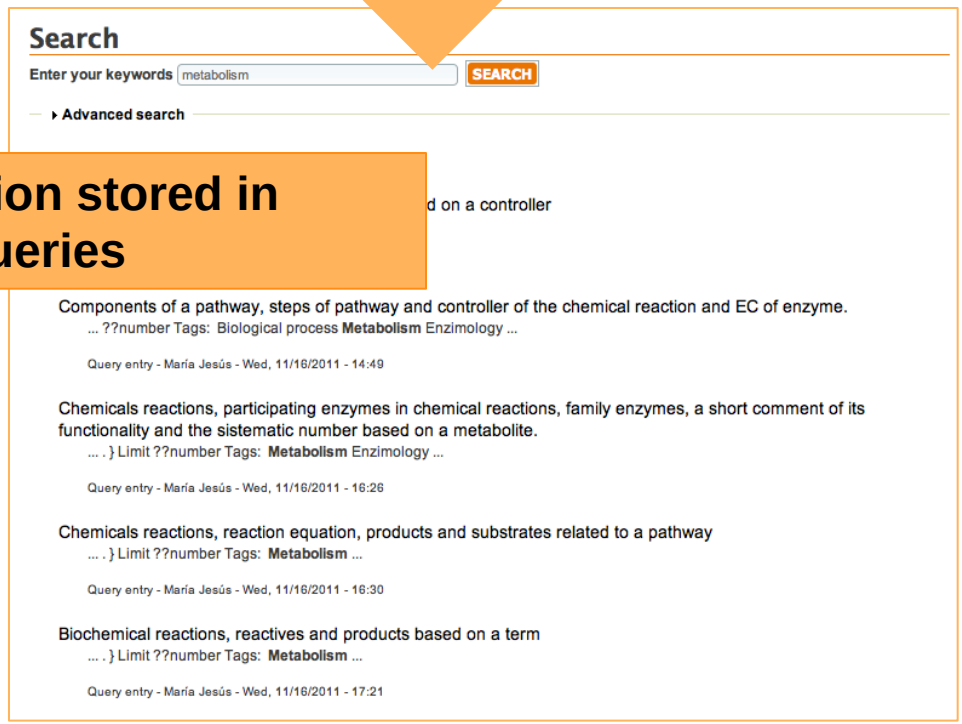
-- BioQueries team

Welcome e-mail

Usability test for evaluation

Search Engine

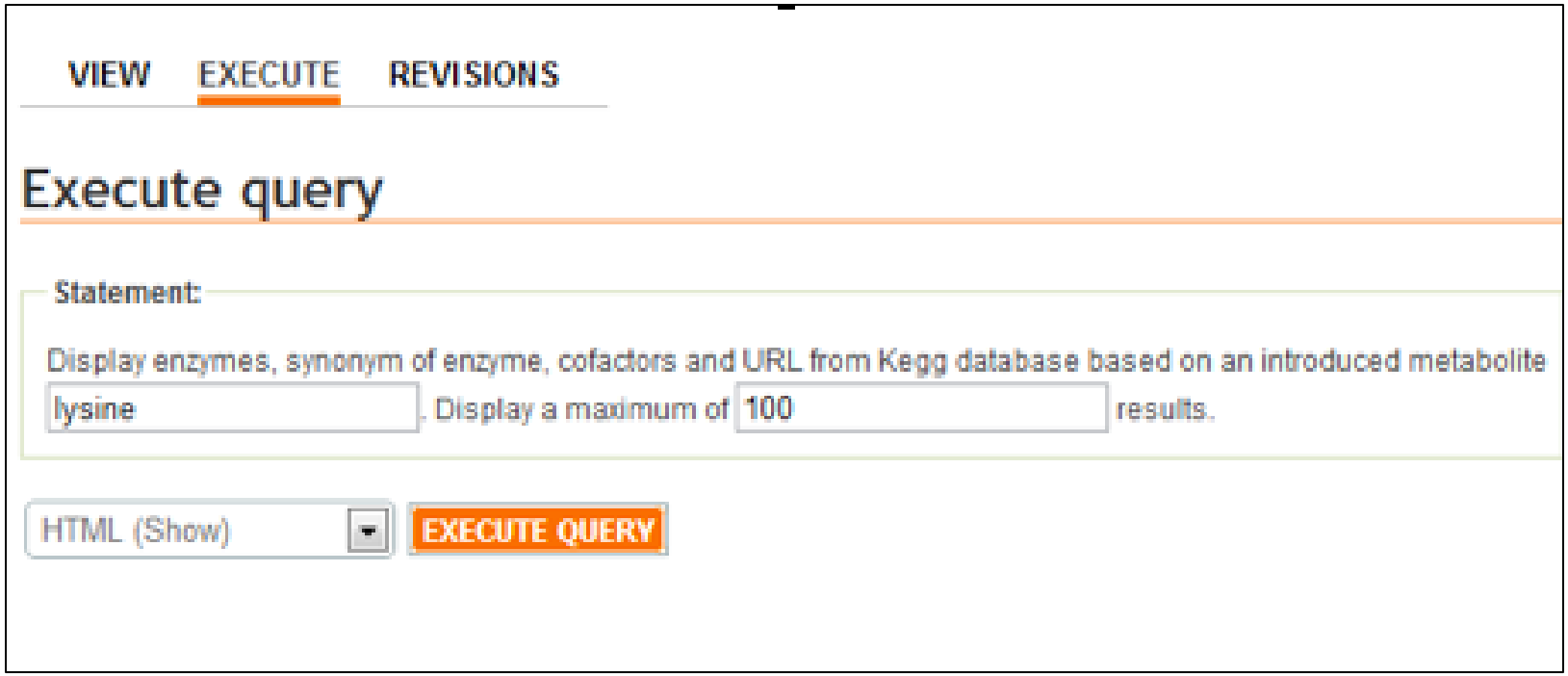




All information stored in Bioqueries

Executing Query

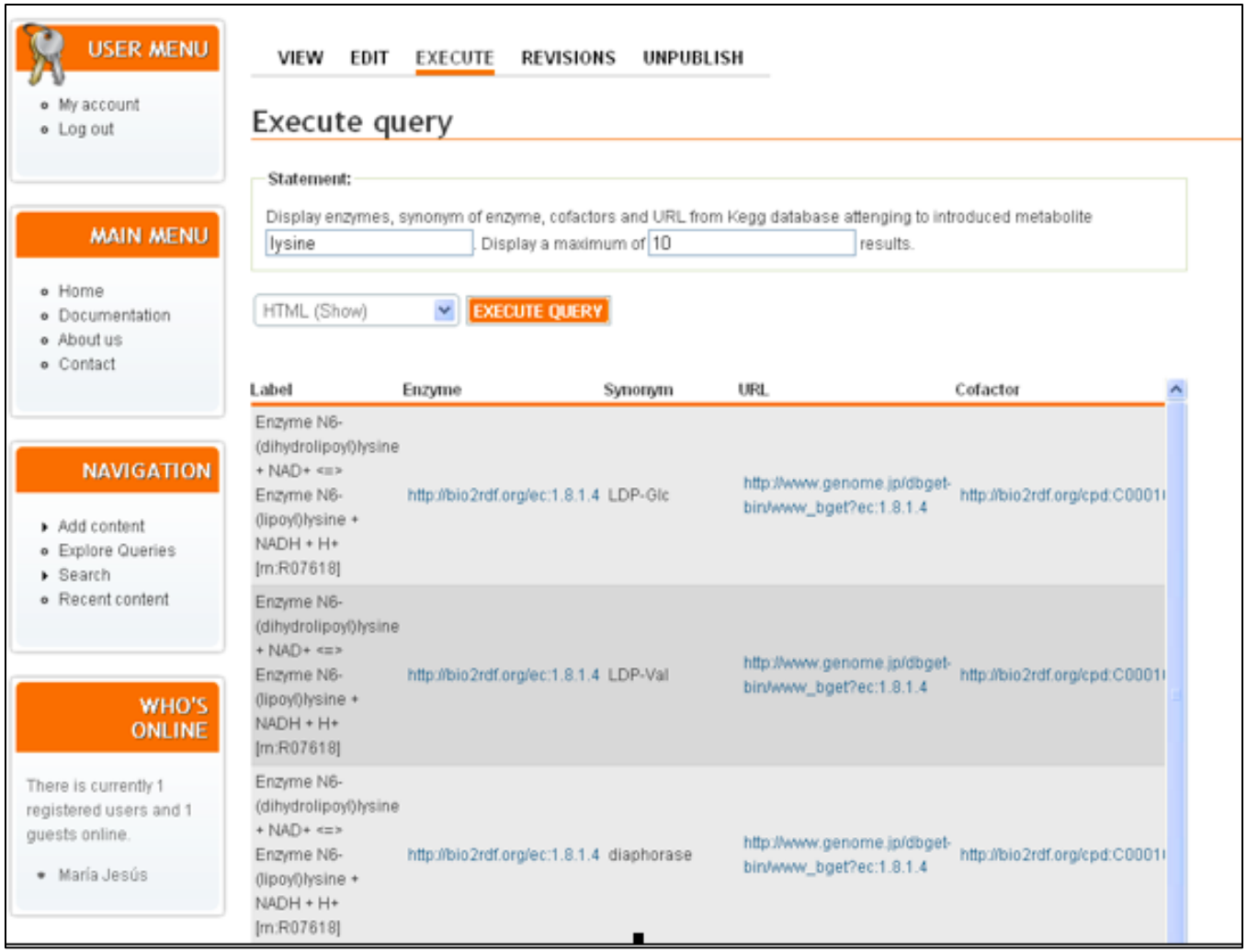
- How to execute a *SPARQL query*



This query displays information about all chemical reactions, enzymes and synonyms of enzymes, cofactors and URL from the Kegg database based on the metabolite introduced (e.j. Lysine). The numbers of results per page are 100.

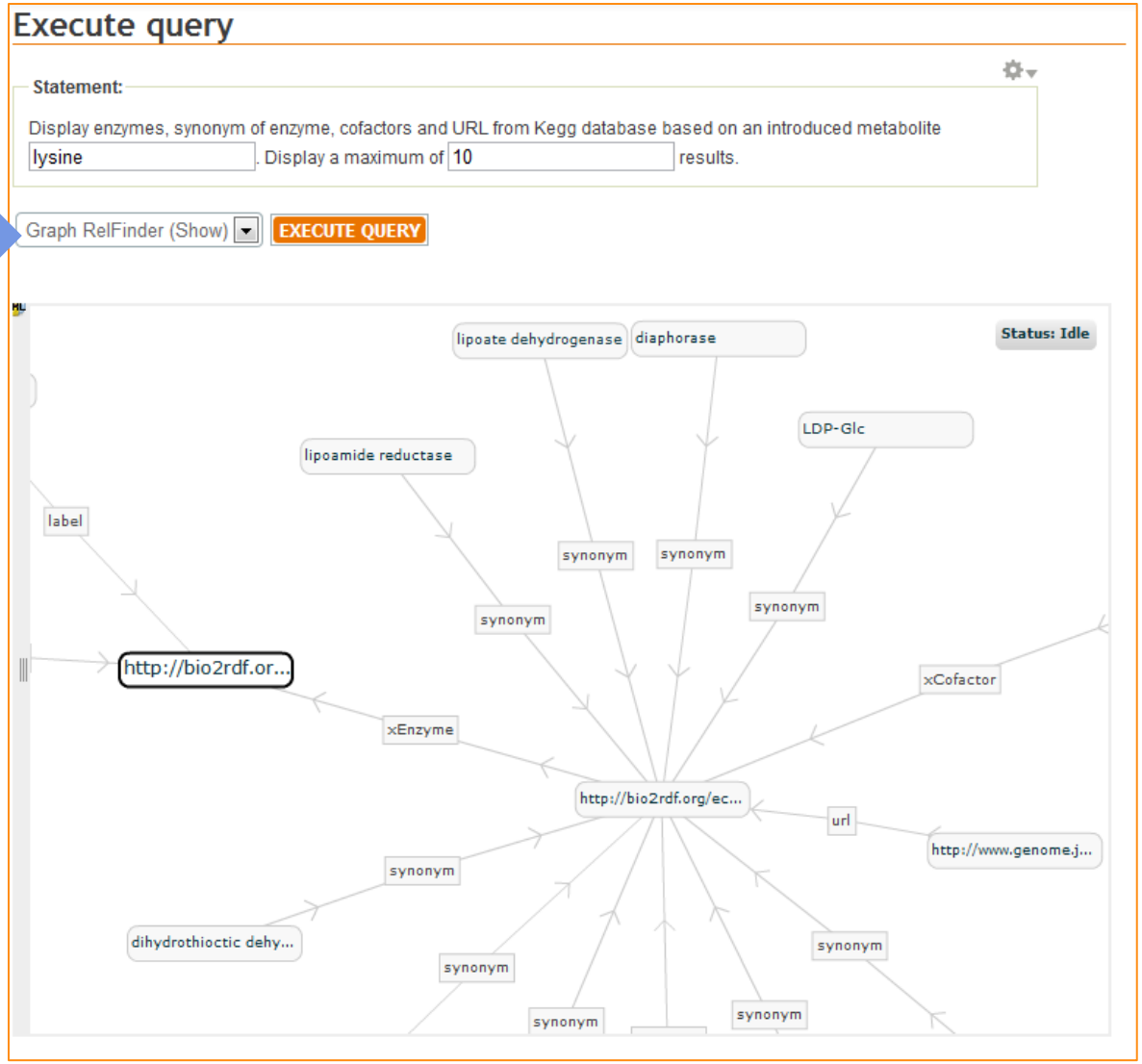
Results

- How the results are displayed:



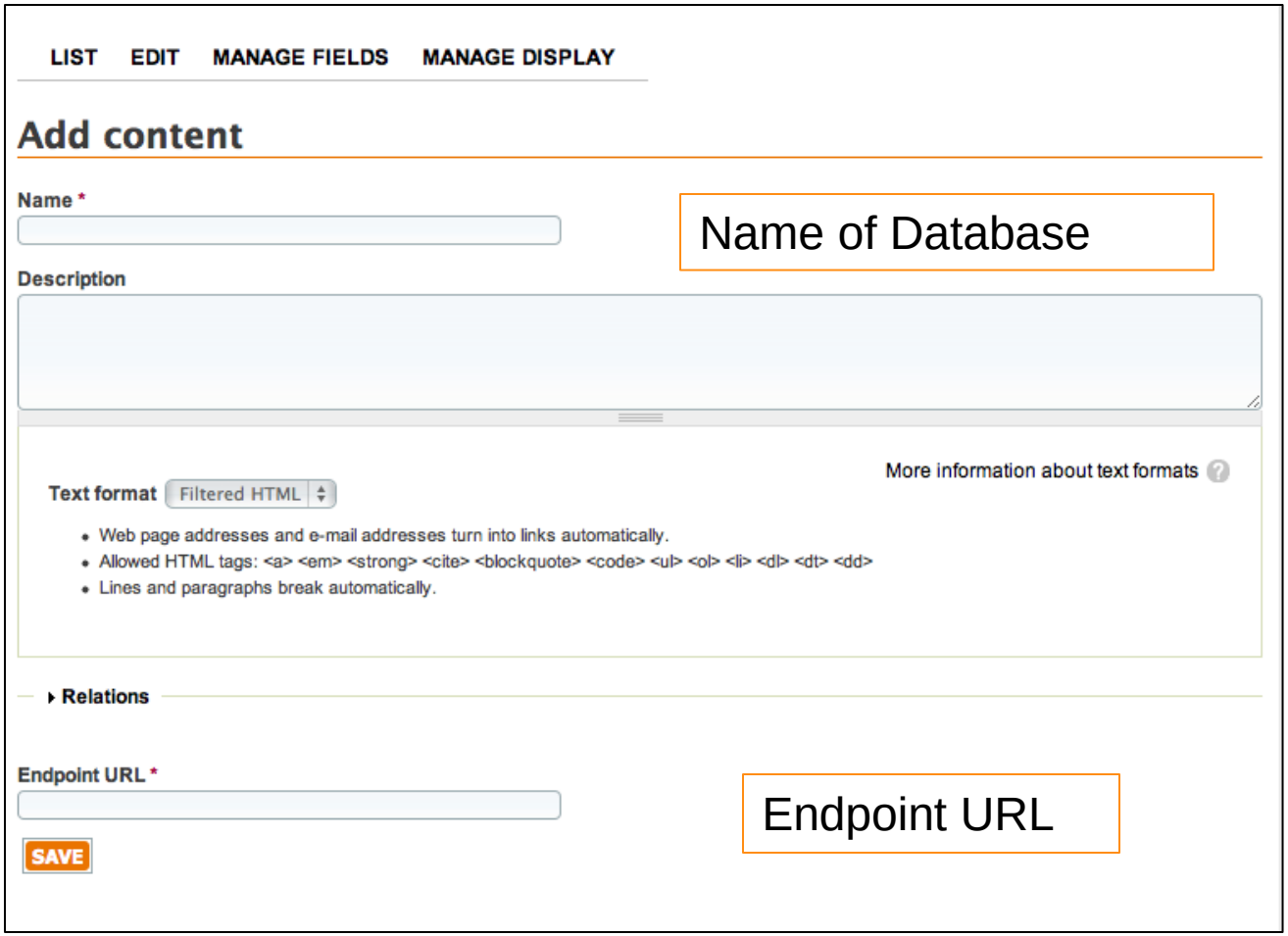
Results

HTML
Relfinder
RDF
N-triples



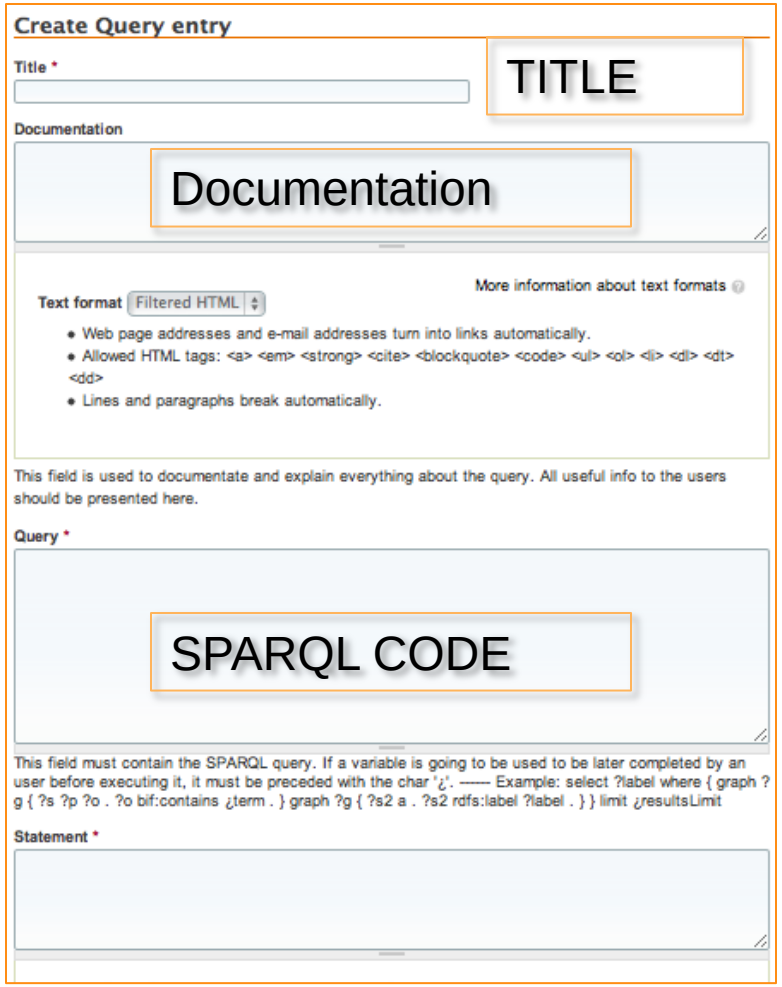
Introducing an Endpoint

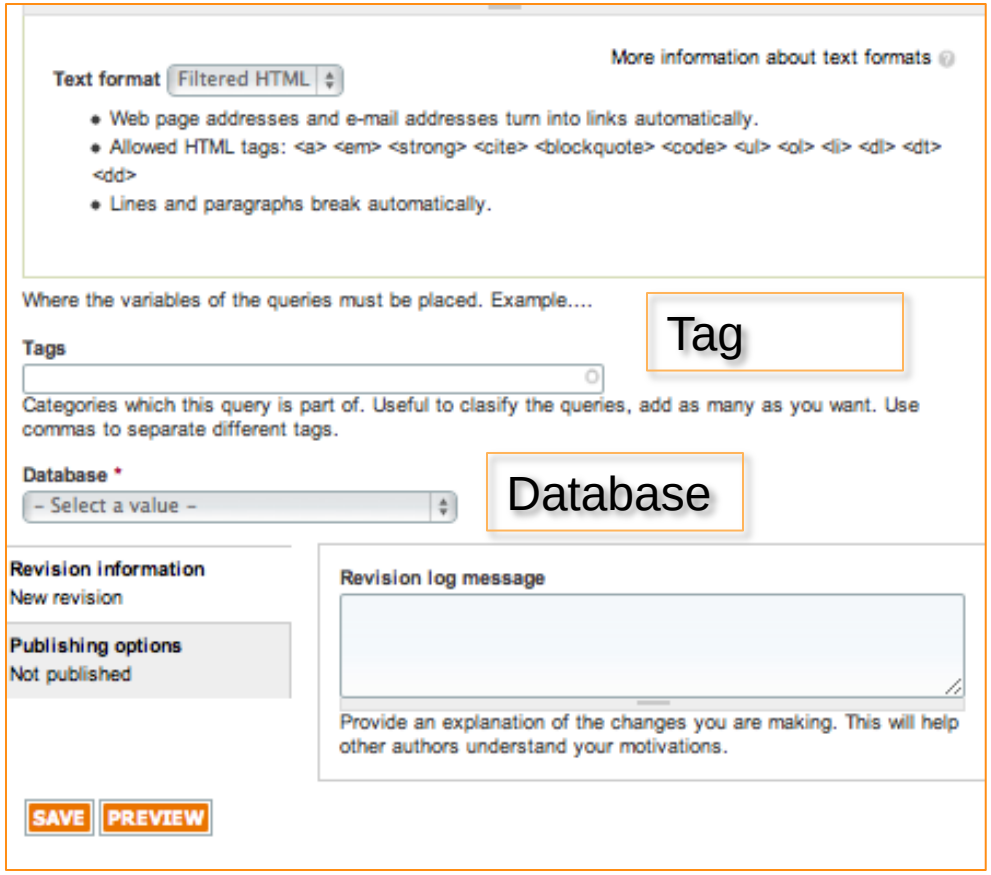
- How to add a new Endpoint



Creating the Query

- How to add a SPARQL query





Editing Query

- How to edit a SPARQL query

VIEW EDIT EXECUTE REVISIONS UNPUBLISH

Edit Query entry **All information related to a disease from Quebec**

Title *

All information related to a disease from Quebec

Documentation

Text format Filtered HTML

More information about text formats

- Web page addresses and e-mail addresses turn into links automatically.
- Allowed HTML tags: <a>   <cite> <blockquote> <code>   <dl> <dt> <dd>
- Lines and paragraphs break automatically.

This field is used to documentate and explain everything about the query. All useful info to the users should be presented here.

Query *

```
SELECT ?label ?InformationofDisease
WHERE {
  ?s1 ?p1 ?InformationofDisease .
  ?s1 <http://www.w3.org/2000/01/rdf-schema#label> ?label .
  ?InformationofDisease bif:contains "'??disease'" .
}
Limit ??number
```

This field must contain the SPARQL query. If a variable is going to be used to be later completed by an user before executing it, it must be preceded with the char '¿'. ----- Example: select ?label where { graph ?g { ?s ?p ?o . ?o bif:contains ¿term . } graph ?g { ?s2 a . ?s2 rdfs:label ?label . } } limit ¿resultsLimit

Statement *

Display all information related to a disease based on a disease (e.g. glaucoma) $disease$. Display a maximum of (e.g.100) $number$ results.

SAVE PREVIEW DELETE

History of Revision

- Queries can be improved by the community and changes can be known

Revisions for *Enzymes, synonyms, URL from Kegg web direction, cofactor and URL from Kegg based on a metabolite*

Revision	Operations
11/16/2011 - 16:28 by Esteban	*current revision*
10/19/2011 - 04:07 by María Jesús	
10/19/2011 - 03:55 by María Jesús	
10/18/2011 - 10:30 by María Jesús	
10/18/2011 - 10:30 by María Jesús	
10/17/2011 - 17:51 by Esteban	
10/11/2011 - 10:13 by María Jesús	
10/03/2011 - 16:24 by Esteban	
09/29/2011 - 17:47 by María Jesús	
09/28/2011 - 14:22 by María Jesús	
09/28/2011 - 14:15 by María Jesús	
09/28/2011 - 14:13 by María Jesús	
09/28/2011 - 14:12 by María Jesús	
09/28/2011 - 13:01 by María Jesús	
09/28/2011 - 13:01 by María Jesús	
09/28/2011 - 12:57 by María Jesús	
09/28/2011 - 12:56 by María Jesús	
09/28/2011 - 12:55 by María Jesús	
09/28/2011 - 12:55 by María Jesús	

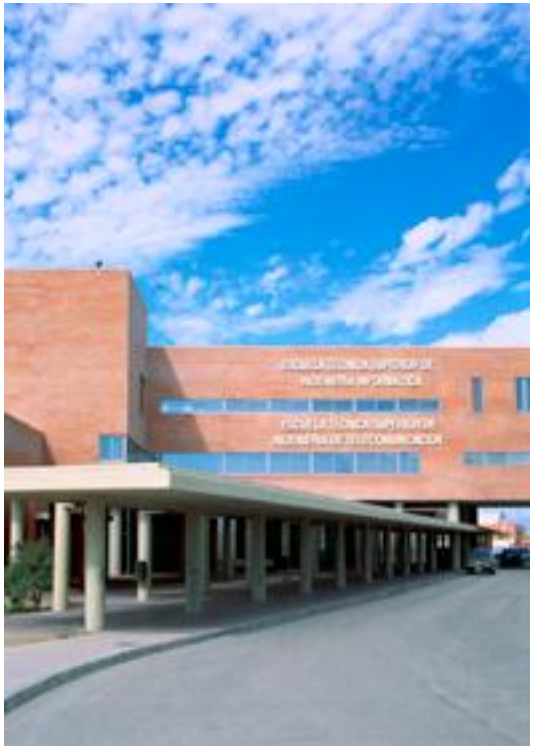
Conclusions

- The main goal is the development of a tool to enable the building of an active community of biologists around the consume of Biological Linked Data.
- This system is designed for two profiles: a bioinformatics profile to add new SPARQL queries and a biological profile to use them.
- The system includes an administration module to build *a group of administrators to organize, maintain and order the information stored in Bioqueries*.

Future and Present Works

- Including federated queries in such a way that users can query different biological Endpoints simultaneously.
- Improving the user interface to make it more intuitive in the query design using Relfinder.
- Automated classification of the queries based on classes and properties they use.

The group



JOSÉ F. ALDANA MONTES

Professor belonging to Department of Computer Languages and Computing Sciences from the University of Málaga. More than 20 years of expertise researching and teaching about databases field. His particular fields of interest are Semantic Middleware, Semantic Web, Semantic Data and Application Integration and Database Extension using formal semantics.

Job:
Group Leader

e-mail:
jfam@lcc.uma.es

Phone:
+34 952 13 28 13

Postal address:
Room 3.2.5 - E.T.S. de Ingeniería Informática - Campus de Teatinos (Málaga)

ESTEBAN LÓPEZ CAMACHO

Job:
Software Engineer

e-mail:
esteban@lcc.uma.es

Phone:
+34 952 13 28 65

Postal address:
Room 3.3.2 - E.T.S. de Ingeniería Informática - Campus de Teatinos (Málaga)

MARÍA JESÚS GARCÍA GODOY

She obtained her Barchelor degree in Biology at University of Malaga in 2009. She has finished Molecular and Celular Biology Master degree in the year 2010. She is working on Software Engineering and Artificial Intelligence Master Thesis and her researches are related to Bioinformatic.

Job:
Assistant Researcher

e-mail:
mjgarciag@lcc.uma.es

Phone:
+34 952 13 28 65

Postal address:
Room 3.3.2 - E.T.S. de Ingeniería Informática - Campus de Teatinos (Málaga)

ISMAEL NAVAS DELGADO

Ismael Navas Delgado finished his PhD in Computer Science in the University of Málaga in 2009. His main interests are: Systems Biology and Semantic Technologies.

Job:
Researcher

e-mail:
ismael@lcc.uma.es

Phone:
+34 952 13 63 14

Postal address:
Room 3.2.49 - E.T.S. de Ingeniería Informática - Campus de Teatinos (Málaga)

4th International SWAT4LS Workshop
Semantic web applications and tools for life sciences

December 7-9th, 2011
London, UK

Bioqueries: A Social Community Sharing Experiences while Querying Biological Linked Data

http://bioqueries.uma.es

María Jesús García-Godoy, Ismael Navas-Delgado, José Francisco Aldana Montes

Computing Languages and Computing Science Department, University of Malaga, Malaga 29071, Spain

ismael@lcc.uma.es