



OPPL-Galaxy: Enhancing ontology exploitation in Galaxy with OPPL

Mikel Egaña Aranguren (megana@fi.upm.es)
Ontology Engineering Group
School of Computer Science
UPM, Spain

Jesualdo Tomás Fernández-Breis (jfernand@um.es)
School of Computer Science
UM, Spain

Erick Antezana (erick.antezana@bio.ntnu.no)
Department of Biology
NTNU, Norway

<http://www.slideshare.net/MikelEganaAranguren/opplgalaxy-enhancing-ontology-exploitation-in-galaxy-with-oppl>



What is OPPL-Galaxy?

Use cases

Implementation

Availability

Conclusions

What is OPPL-Galaxy?

Ontology Pre Processor Language (OPPL)

Scripting Language for automating ontology manipulation

Ontology

```
?whole:CLASS,  
?part:CLASS  
SELECT ?part SubClassOf part_of some ?whole  
WHERE ?part != Nothing  
BEGIN  
ADD ?part SubClassOf part_of only ?whole  
END;
```

New ontology

Galaxy

The screenshot displays the Galaxy web interface. The main content area shows the configuration for a 'Filter' tool. The filter is set to '1: s1' and the condition is 'c5/c10 >= 0.95'. Below the configuration, there are several tips and a syntax section. The history panel on the right shows a list of previous tool runs, including 'Draw phylogeny on data 12', 'Find lowest diagnostic rank on data 10', and 'Fetch taxonomic representation on data 8'. The current tool run is highlighted in green.

Filter

Filter:

Query missing? See TIP below.

With following condition:

Double equal signs, ==, must be used as shown above. To filter for an arbitrary string, use the Select tool.

TIP: Attempting to apply a filtering condition may throw exceptions if the data type (e.g., string, integer) in every line of the columns being filtered is not appropriate for the condition (e.g., attempting certain numerical calculations on strings). If an exception is thrown when applying the condition to a line, that line is skipped as invalid for the filter condition. The number of invalid skipped lines is documented in the resulting history item as a "Condition/data issue".

TIP: If your data is not TAB delimited, use *Text Manipulation* -> *Convert*

Syntax

The filter tool allows you to restrict the dataset using simple conditional statements.

- Columns are referenced with **c** and a **number**. For example, **c1** refers to the first column of a tab-delimited file
- Make sure that multi-character operators contain no white space (e.g., **<=** is valid while **< =** is not valid)
- When using 'equal-to' operator **double equal sign '==' must be used** (e.g., **c1==chr1**)
- Non-numerical values must be included in single or double quotes (e.g., **c6=="+"**)
- Filtering condition can include logical operators, but **make sure operators are all lower case** (e.g., **(c1='chrX' and c1!='chrY')** or not **c6=="+"**)

Example

- c1=='chr1'** selects lines in which the first column is chr1
- c3-c2<100*c4** selects lines where subtracting column 3 from column 2 is less than the value of column 4 times 100
- len(c2.split("\n")) < 4** will select lines where the

History

14: Draw phylogeny on data 12

13: Draw phylogeny on data 11

12: Find lowest diagnostic rank on data 10

11: Find lowest diagnostic rank on data 9

10: Fetch taxonomic representation on data 8

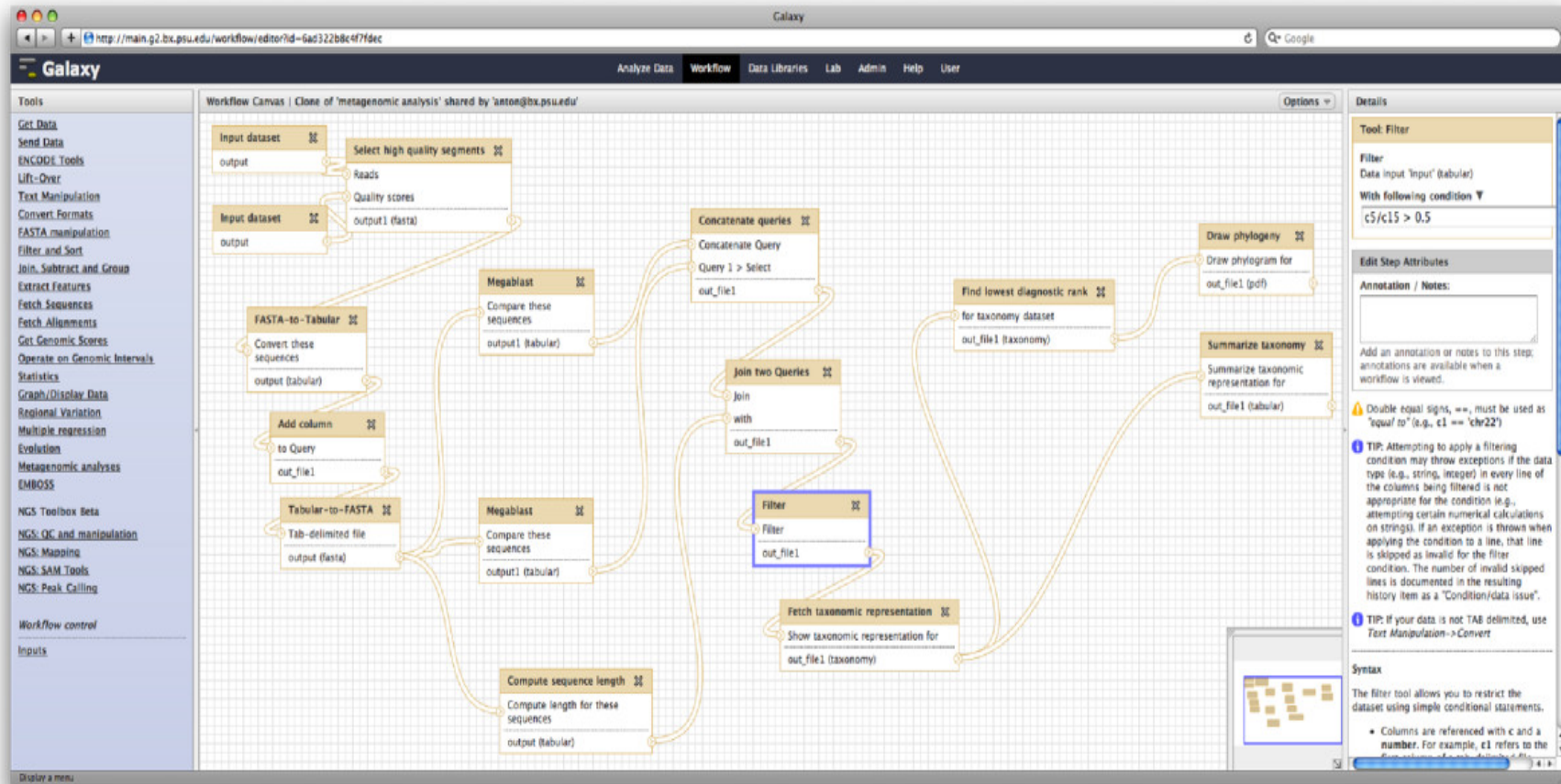
9: Fetch taxonomic representation on data 7

8: s234 within 5% of max
62,647 lines, format: tabular, database: ?
Info: Filtering with c5/c10 >= 0.95, kept 1.75% of 3581932 lines.


7: s1 within 5% of max
54,458 lines, format: tabular, database: ?
Info: Filtering with c5/c10 >= 0.95, kept 1.43% of 3812372 lines.

Jeremy Goecks, Anton Nekrutenko, James Taylor, and The Galaxy Team. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biology*, 11(8):R86+, 2010.

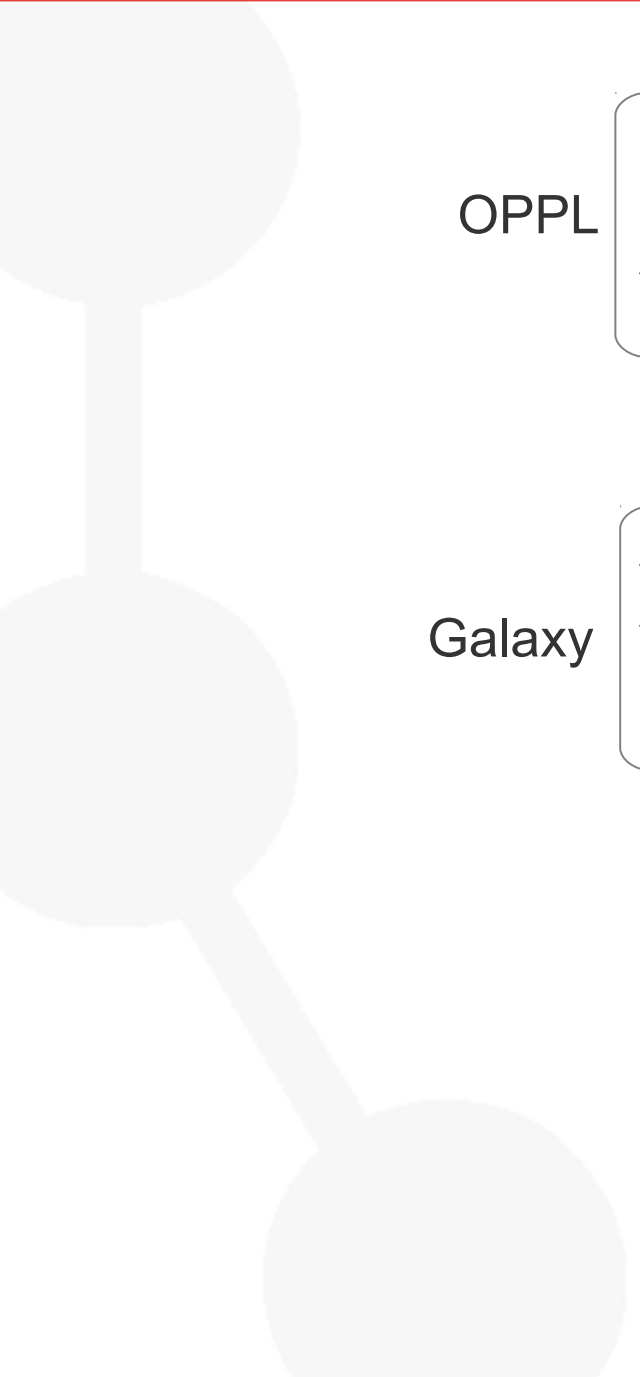
Galaxy



Jeremy Goecks, Anton Nekrutenko, James Taylor, and The Galaxy Team. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biology*, 11(8):R86+, 2010.



OPPL {
Ontology refactoring (Axiomatic enrichment, ODPs, ...)
Detection of structures (Antipatterns, ...)
Advanced querying (Reasoning+regexps, ...)
Ontology trimming

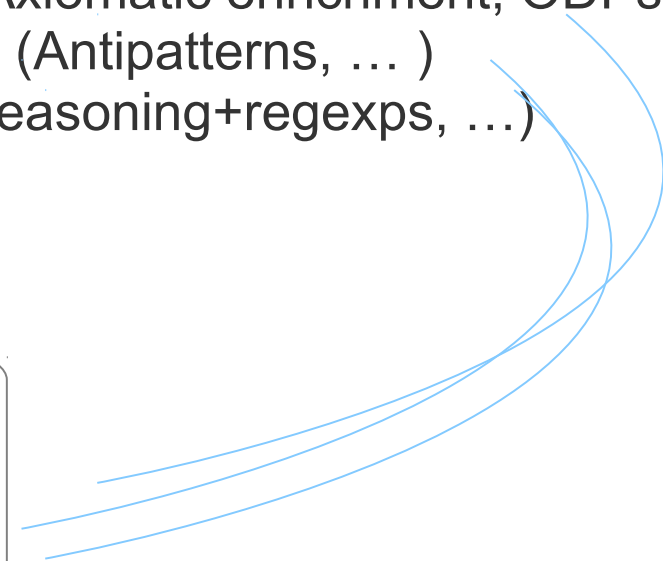


OPPL {
Ontology refactoring (Axiomatic enrichment, ODPs, ...)
Detection of structures (Antipatterns, ...)
Advanced querying (Reasoning+regexps, ...)
Ontology trimming

Galaxy {
Analysis history
Analysis shareability
Combinations of tools
Complex workflows

OPPL {
Ontology refactoring (Axiomatic enrichment, ODPs, ...)
Detection of structures (Antipatterns, ...)
Advanced querying (Reasoning+regexps, ...)
Ontology trimming

Galaxy {
Analysis history
Analysis shareability
Combinations of tools
Complex workflows



What is OPPL-Galaxy

Ontology ← User → OPPL script

Galaxy Analyze Data Workflow Shared Data Help User

Tools Options

- Lift-Over
- Text Manipulation
- Filter and Sort
- Join, Subtract and Group
- Convert Formats
- Extract Features
- Fetch Sequences
- Fetch Alignments
- Get Genomic Scores
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- Regional Variation
- Multiple regression
- Multivariate Analysis
- Evolution
- Motif Tools
- Multiple Alignments
- Metagenomic analyses
- FASTA manipulation
- NCBI BLAST+

Execute an OPPL file against an ontology

Tick the box if the input ontology includes OWL imports:

Input ontology file:
268: gene_ontology_ext.owl

OPPL file:
237: Disjoint_localisation.oppl

Choose ontology output format:
OWL

Add inferred subsumption axioms to output ontology:

To merge ontologies (If there are imports), type an URI for the new ontology (Leave NoMerge for not merging):
NoMerge

Choose reasoner:
Pellet

Execute

History Options

Unnamed history 2.0 Gb

- 274: Execute an OPPL file against an ontology on data 269, data 267, and data 268
- 273: Execute an OPPL file against an ontology on data 269, data 267, and data 268
- 272: Execute an OPPL file against an ontology on data 263 and data 264
- 271: Execute an OPPL file against an ontology on data 263 and data 264
- 270: Execute an OPPL file against an ontology on data 268, data 267, and data 269
- 269: test URI mappings
- 268: gene_ontology_ext.owl

Ontology → User

Use cases

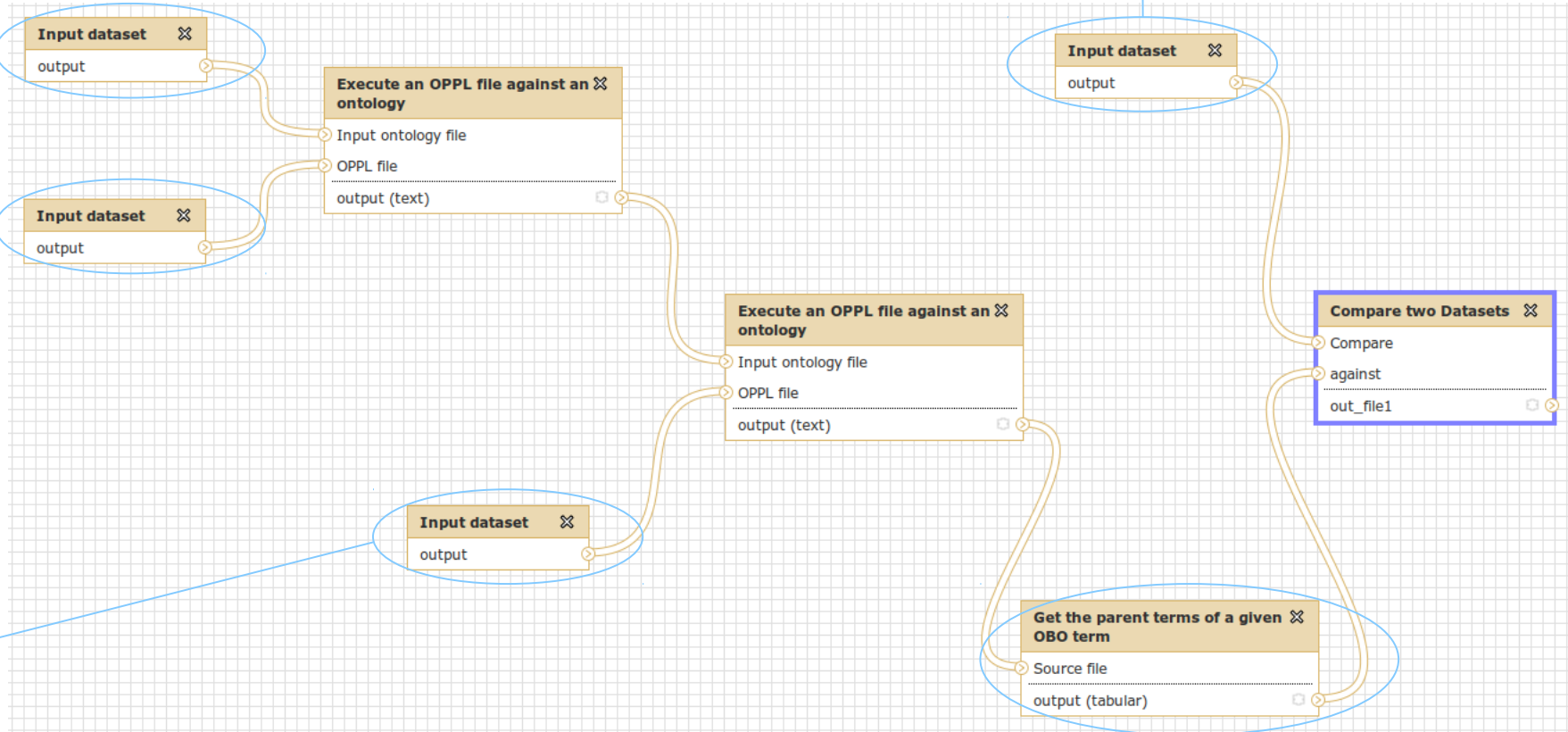
<http://miuras.inf.um.es/OPPL-Galaxy>

Gene Ontology (GO)

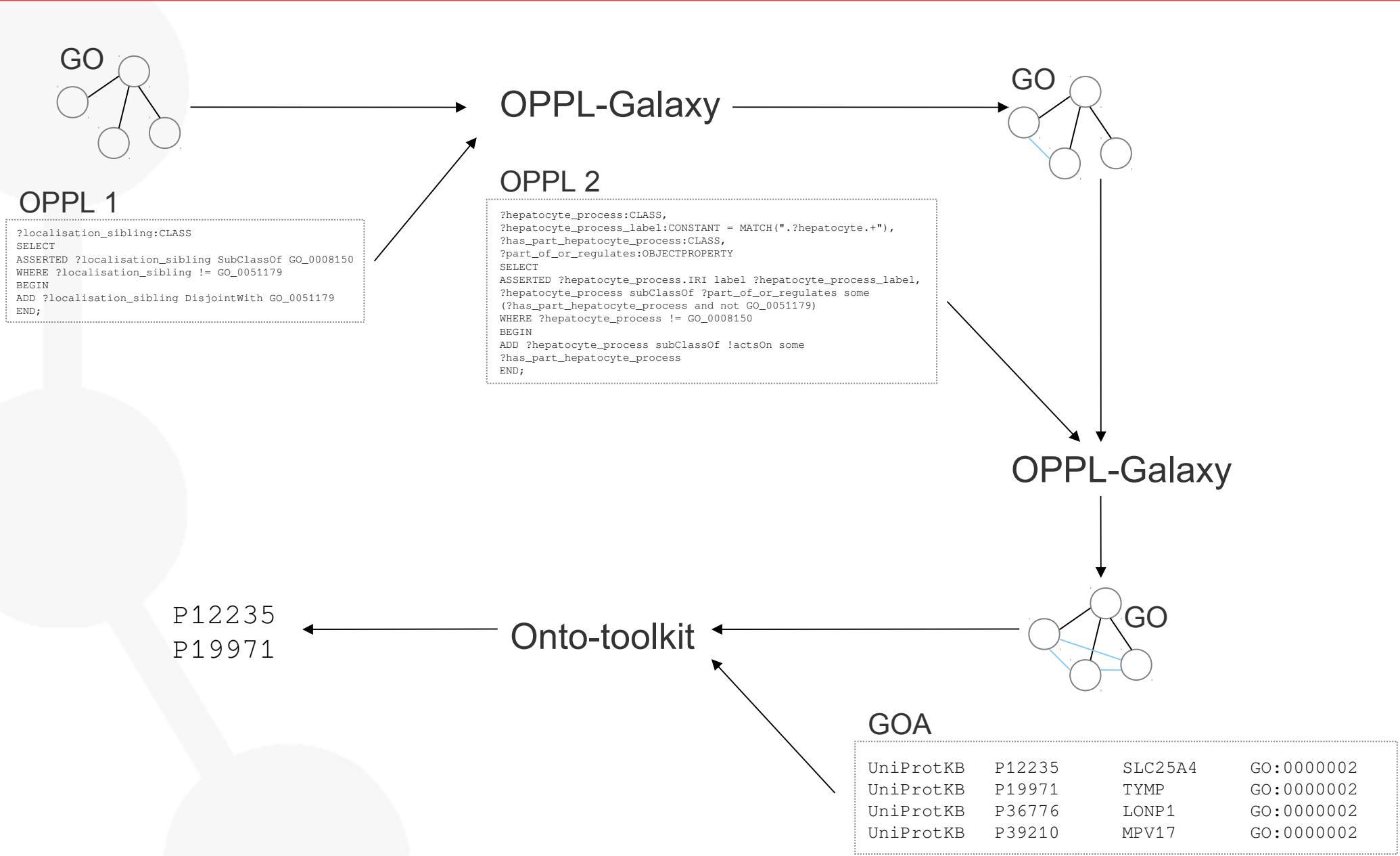
Gene Ontology Annotation file (GOA)

OPPL script 1

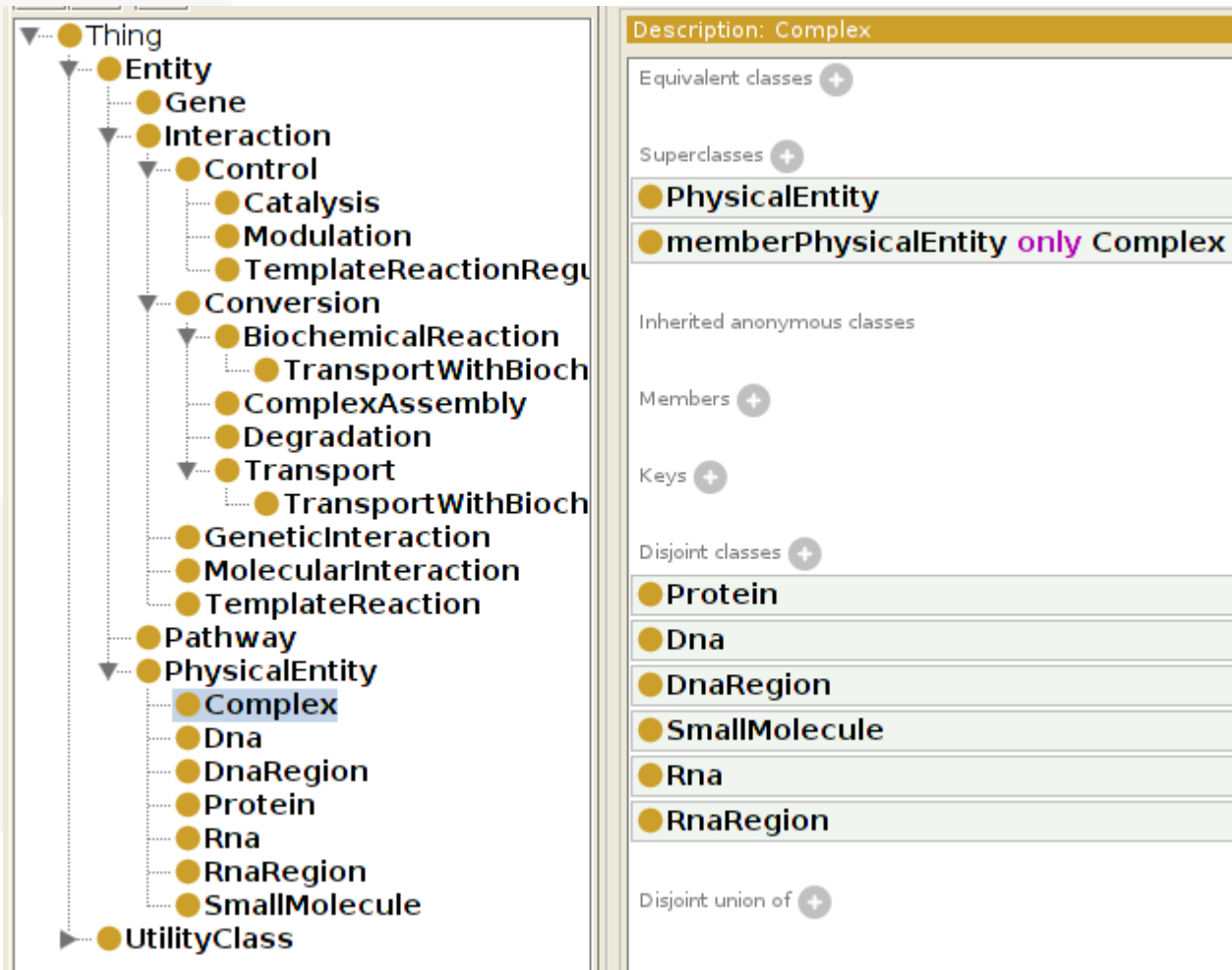
OPPL script 2



Onto-toolkit



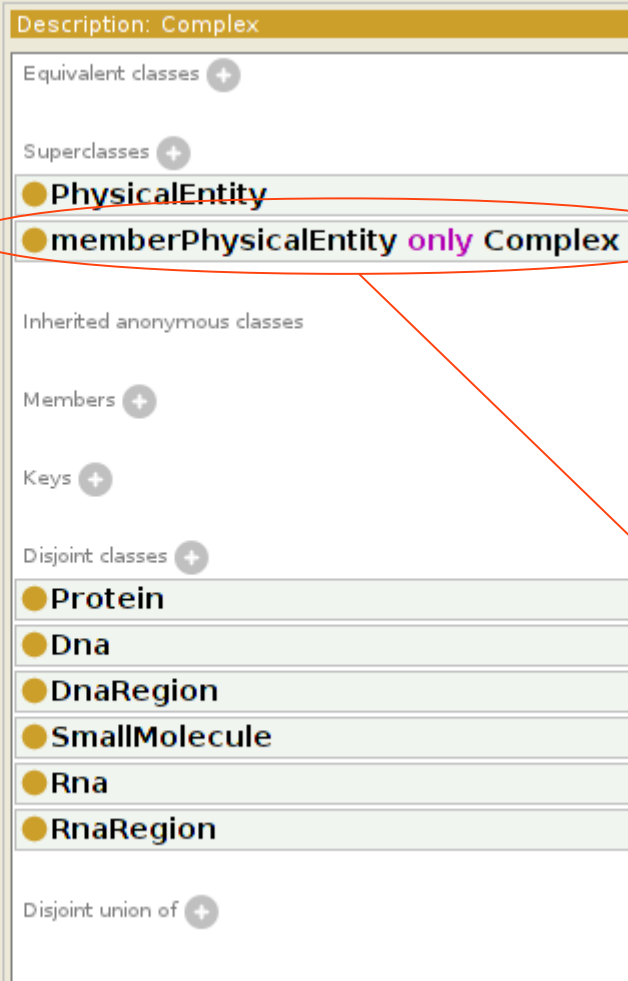
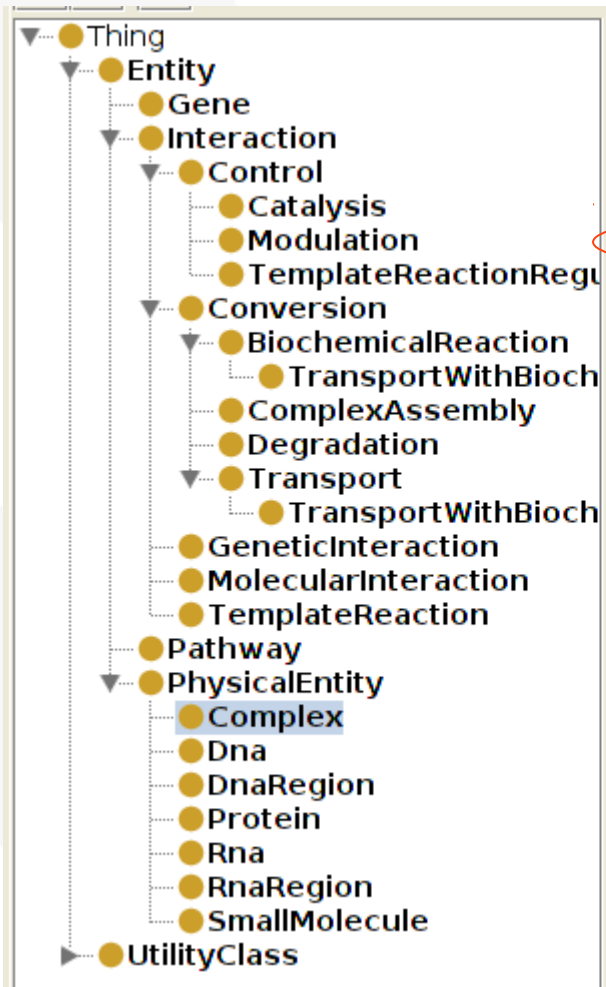
BioPAX



```
?target:CLASS,
?prop:OBJECTPROPERTY,
?filler:CLASS
SELECT ASSERTED ?target SubClassOf
?prop only ?filler
WHERE FAIL ?target SubClassOf
?prop some ?filler
BEGIN
ADD ?target SubClassOf
!OnlyBadPracticeResult
END;
```

Alan Rector, Nick Drummond, Matthew Horridge, Jeremy Rogers, Holger Knublauch, Robert Stevens, Hai Wang, and Chris Wroe. OWL pizzas: Practical experience of teaching OWL-DL: Common errors and common patterns. In LNCS 3257, 2004.

BioPAX

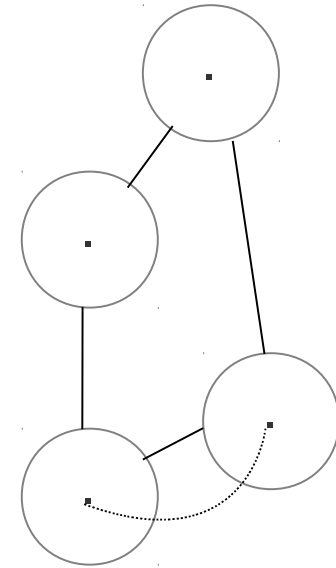
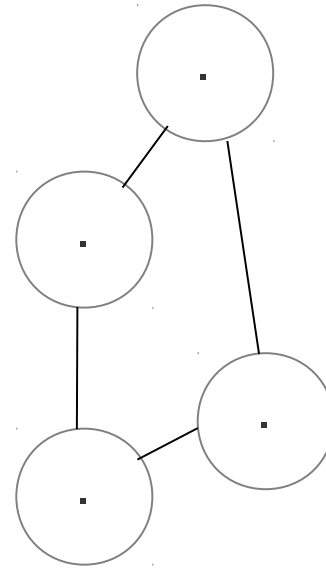
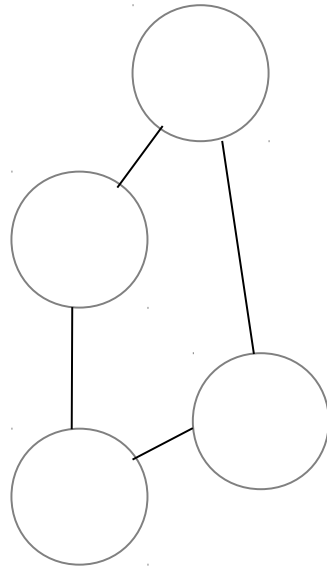
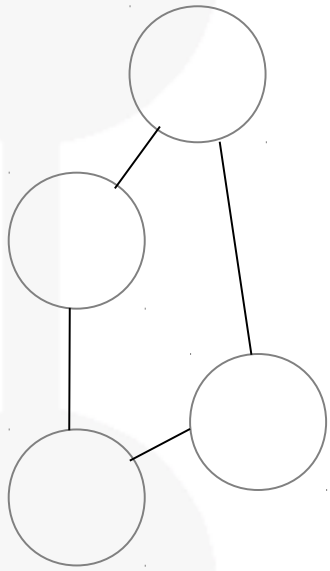


```
?target:CLASS,  
?prop:OBJECTPROPERTY,  
?filler:CLASS  
SELECT ASSERTED ?target SubClassOf  
?prop only ?filler  
WHERE FAIL ?target SubClassOf  
?prop some ?filler  
BEGIN  
ADD ?target SubClassOf  
!OnlyBadPracticeResult  
END;
```



Alan Rector, Nick Drummond, Matthew Horridge, Jeremy Rogers, Holger Knublauch, Robert Stevens, Hai Wang, and Chris Wroe. OWL pizzas: Practical experience of teaching OWL-DL: Common errors and common patterns. In LNCS 3257, 2004.

Refactoring: OBO2OWL puning



OBO

OWL

OWL (Punned)

OWL (Punned, triples)

```
?x:CLASS,  
?y:INDIVIDUAL = create(?x.RENDERING)  
SELECT ?x SubClassOf Thing  
WHERE ?x != Nothing, ?x != Thing  
BEGIN  
ADD ?y Type ?x  
END;
```

```
?x:CLASS,  
?y:INDIVIDUAL,  
?z:CLASS,  
?w:INDIVIDUAL,  
?p:OBJECTPROPERTY  
SELECT ASSERTED ?x SubClassOf ?p some ?z,  
ASSERTED ?y Type ?x, ASSERTED ?w Type ?z  
WHERE ?x != Nothing, ?x != Thing  
BEGIN  
ADD ?y ?p ?w  
END;
```


Refactoring: application of ODPs

Person subClassOf hasheight some high

```
?x:CLASS,  
?y:OBJECTPROPERTY = MATCH("has((\\w+))"),  
?z:CLASS,  
?feature:CLASS = create(?y.GROUPS(1))  
SELECT ASSERTED ?x subClassOf ?y some ?z  
BEGIN  
REMOVE ?x subClassOf ?y some ?z,  
ADD ?x subClassOf !hasFeature some  
(?feature and !hasValue some ?z)  
END;
```

Person subClassOf hasFeature some (height and hasValue some high)

Ondrej Sváb-Zamazal, Vojtech Svátek, Luigi Iannone: Pattern-Based Ontology Transformation Service Exploiting OPPL and OWL-API. EKAW 2010

http://www.gong.manchester.ac.uk/odp/html/Entity_Feature_Value.html

Implementation

URI mapping file (If imports)

Input ontology: OWL, OBO, ...

OPPL flat file

Output ontology: OWL (RDF/XML) or OBO

Add inferred subsumptions as assertions

Merge ontologies (If imports)

Reasoner: Pellet, FaCT++, Hermit

Execute an OPPL file against an ontology

Tick the box if the input ontology includes OWL imports:

Input ontology file:
268: gene_ontology_ext.owl

OPPL file:
237: Disjoint_localisation.oppl

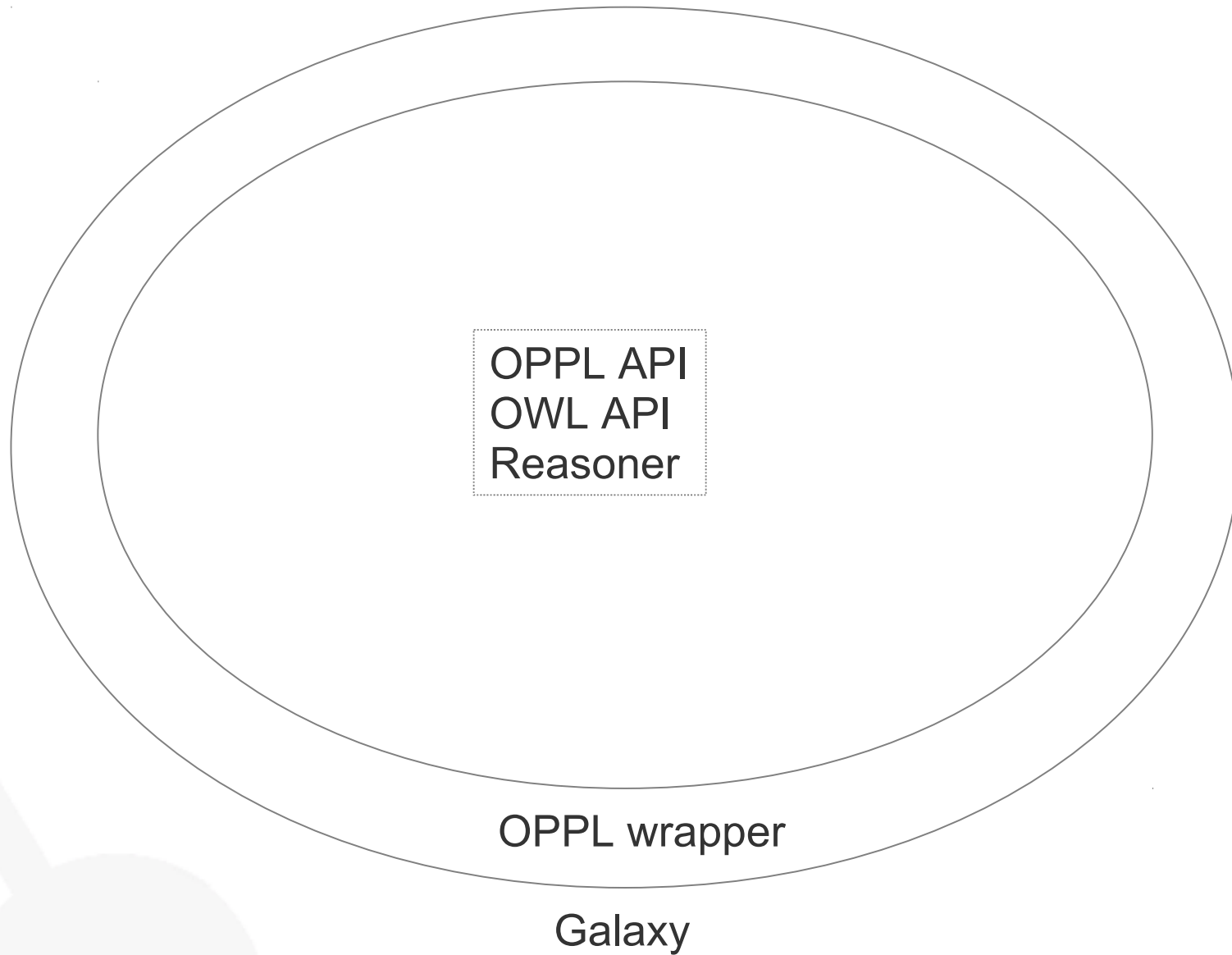
Choose ontology output format:
OWL

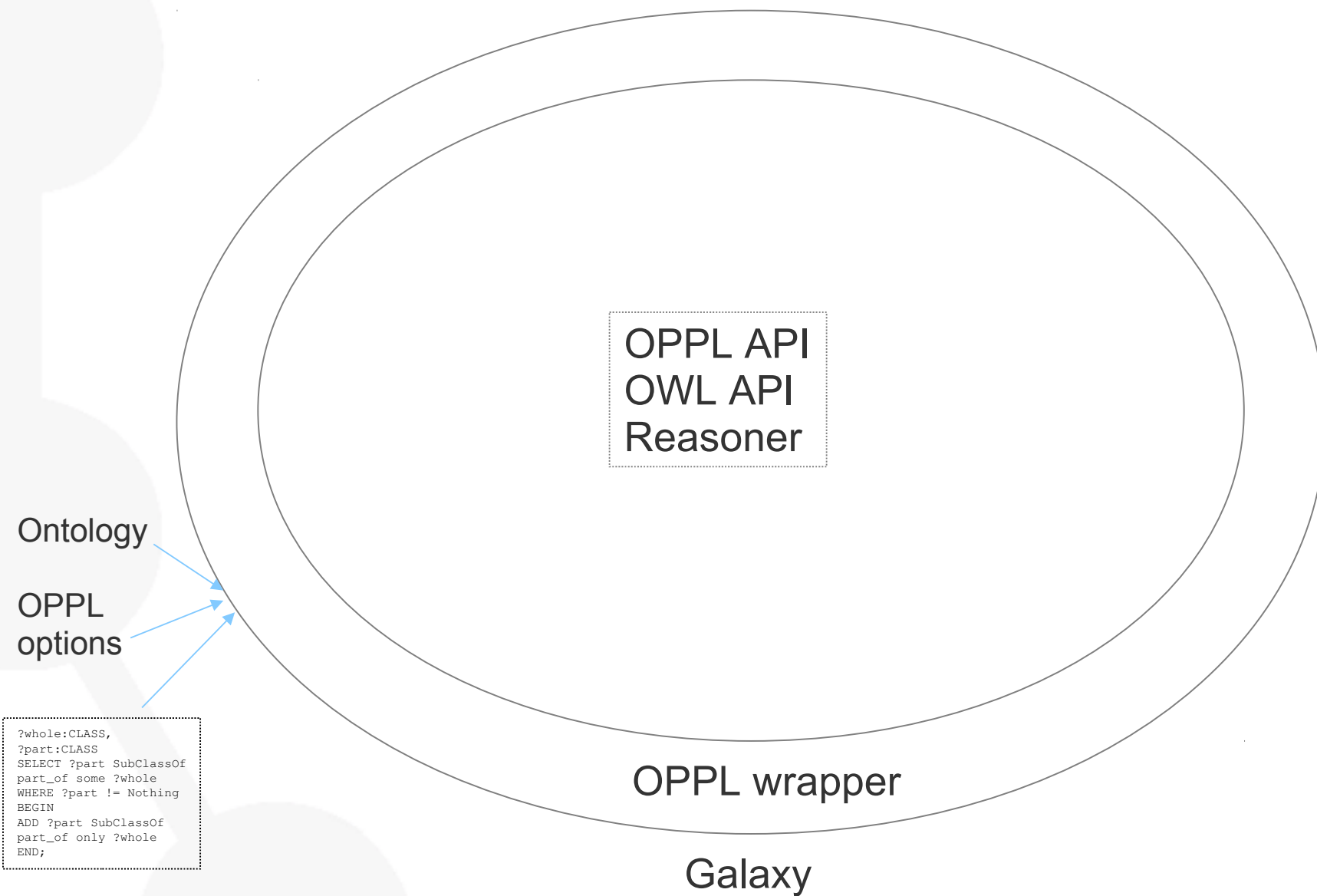
Add inferred subsumption axioms to output ontology:

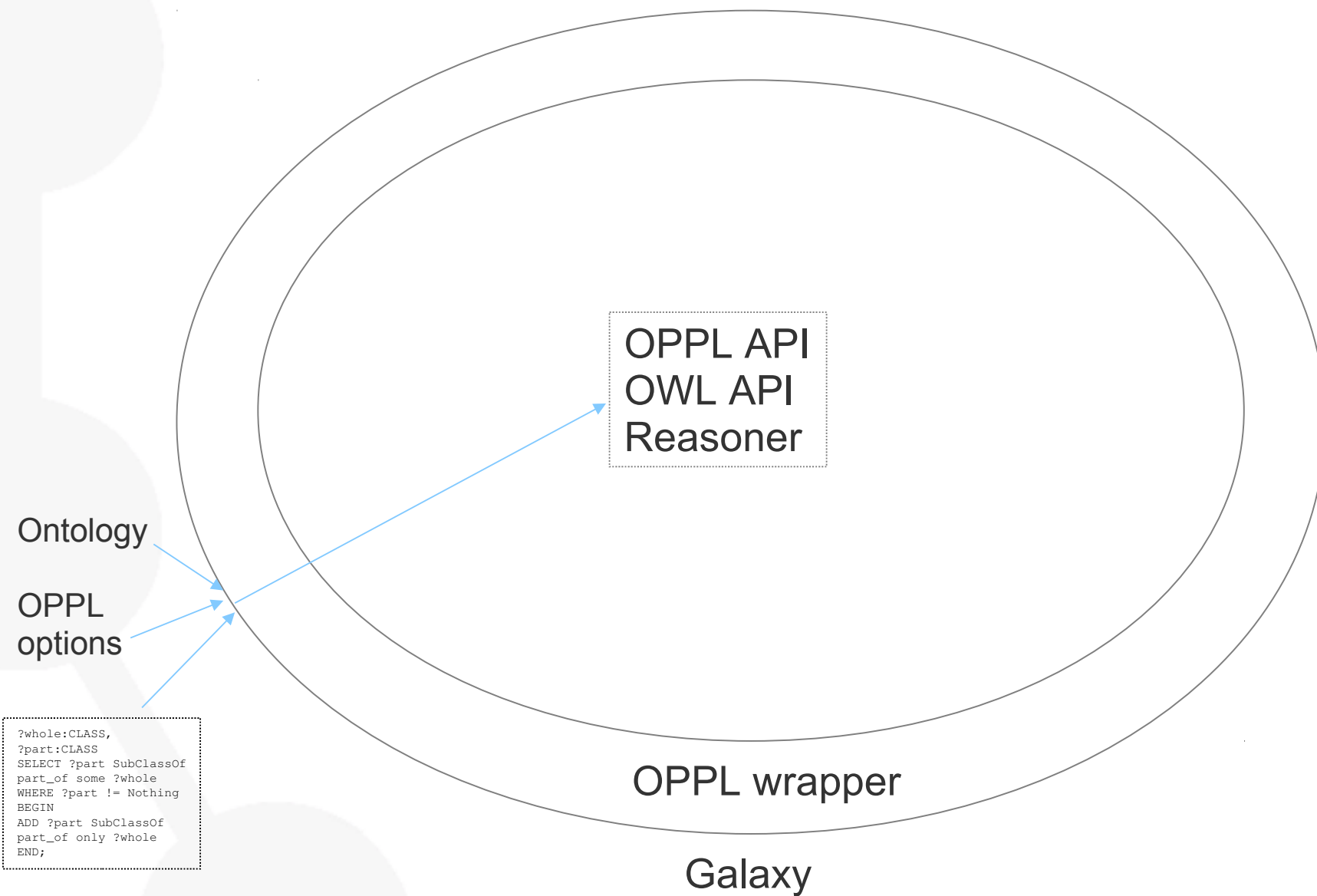
To merge ontologies (If there are imports), type an URI for the new ontology (Leave NoMerge for not merging):
NoMerge

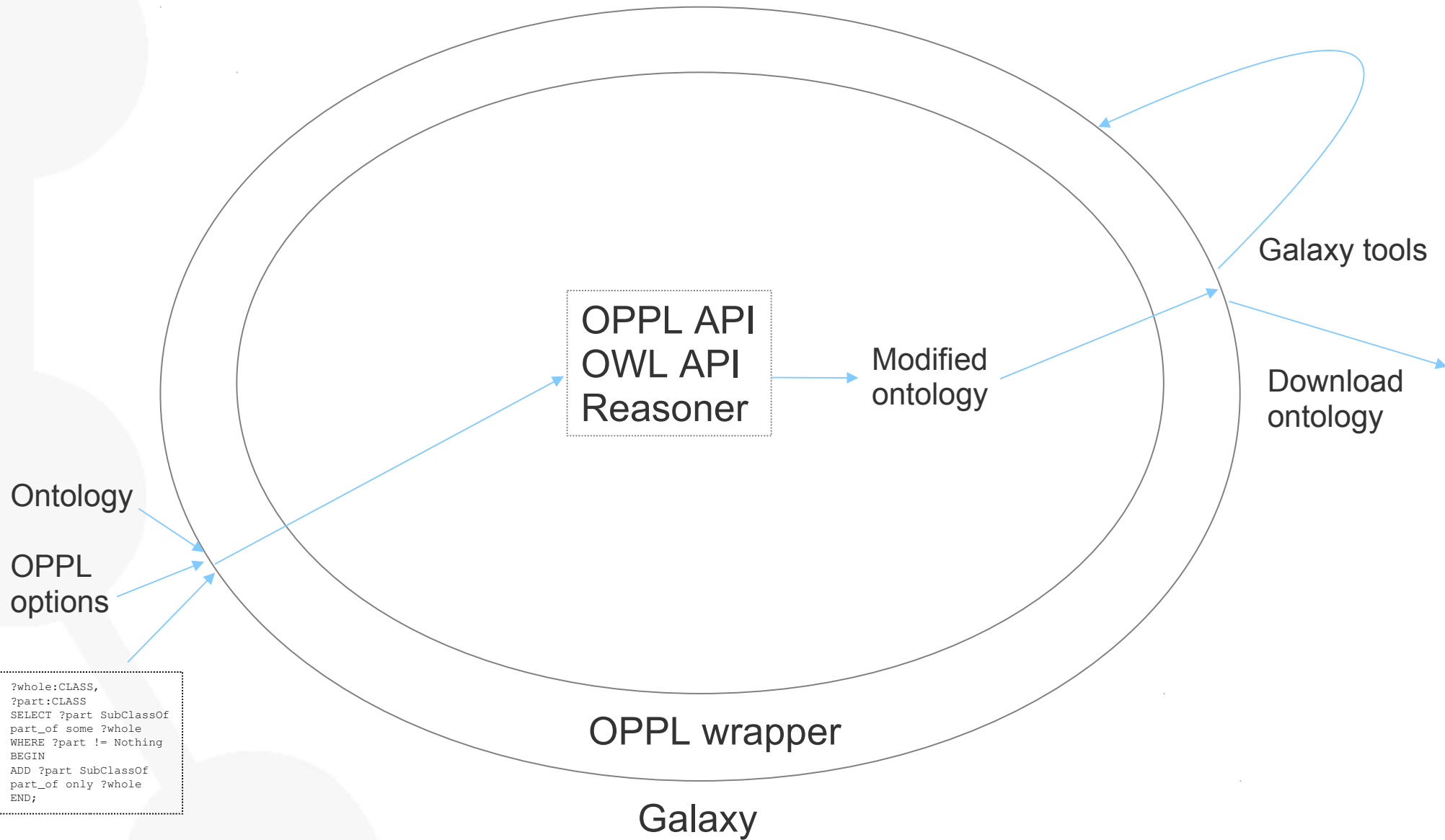
Choose reasoner:
Pellet

Execute









Availability

OPPL-Galaxy at Galaxy tool shed

Ontology manipulation: <http://toolshed.g2.bx.psu.edu/>

OPPL-Galaxy source

<http://toolshed.g2.bx.psu.edu/repos/mikel-egana-aranguren/oppl>

Public Galaxy instance with OPPL-Galaxy

<http://sele.inf.um.es:8080/>

OPPL scripts:

<http://oppl2.sourceforge.net/taggedexamples/>

OPPL-Galaxy works in UNIX (Mac OS X, GNU/Linux)

OPPL-Galaxy is GPL

Conclusions

OPPL: automate ontology manipulation

OPPL-Galaxy: exploit automated ontology manipulation:

In combination with other genomics tools

In complex workflows

History, shared execution, reproduce execution, ...

Ontology refactoring, complex querying, ontology pruning, ...

Inference

Ultimately, automate ontology manipulation in a “natural” environment for biologists

Mikel Egaña Aranguren is funded by the Marie Curie Cofund programme (FP7)

Jesualdo Tomás Fernández Breis is funded by the Spanish Ministry of Science and Innovation (Grant TIN2010-21388-C02-02)