

The Cell Cycle Ontology

Erick Antezana

Dept. of Plant Systems Biology

Flanders Institute for Biotechnology (VIB) / Ghent University

Ghent - BELGIUM

erant@psb.ugent.be

<http://www.CellCycleOntology.org>

(Some) Motivating questions

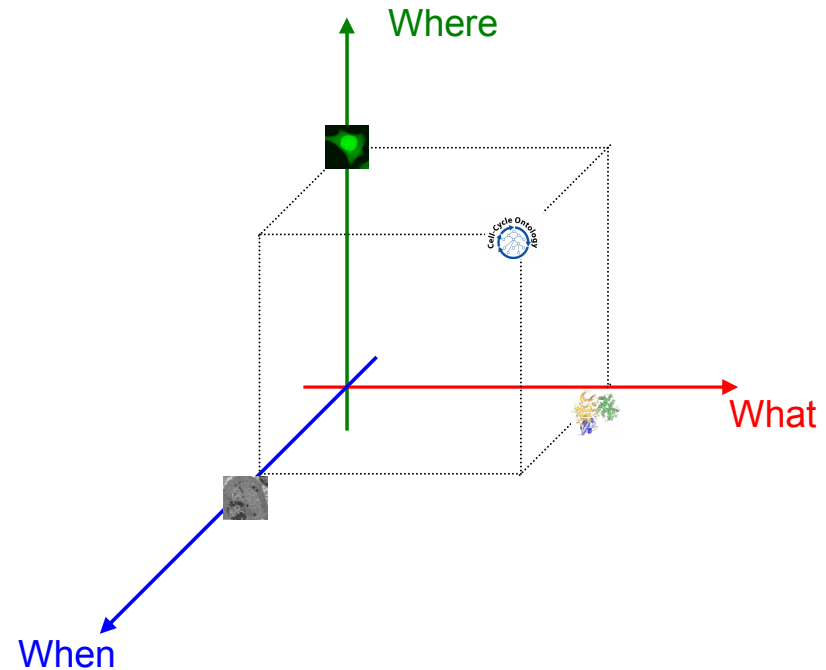
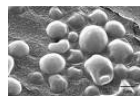
- I'm working with **AT5g35520**, in which interactions this gene play a role?
- From my microarray experiment I've got this gene **X**, is this gene involved in the cell cycle, ..., ?
- Verify my models of genetic, metabolic and product interaction networks
- ...



CCO in a nutshell

- Capture the knowledge of the CC process
- dynamic aspects of terms and their interrelations
- promote sharing, reuse and enable better computational integration with existing resources
- Issues: *synonymy*, *polysemy*

ORGANISMS:



“Cyclin B (*what*) is located in Cytoplasm (*where*) during Interphase (*when*)”

<http://www.CellCycleOntology.org>

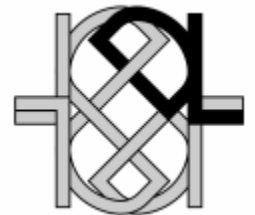
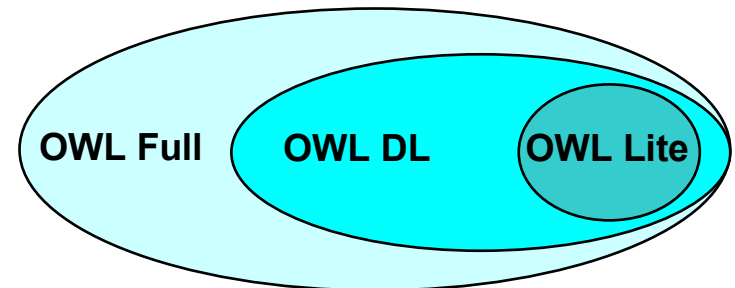
Antezana et al. LNBI, 2006

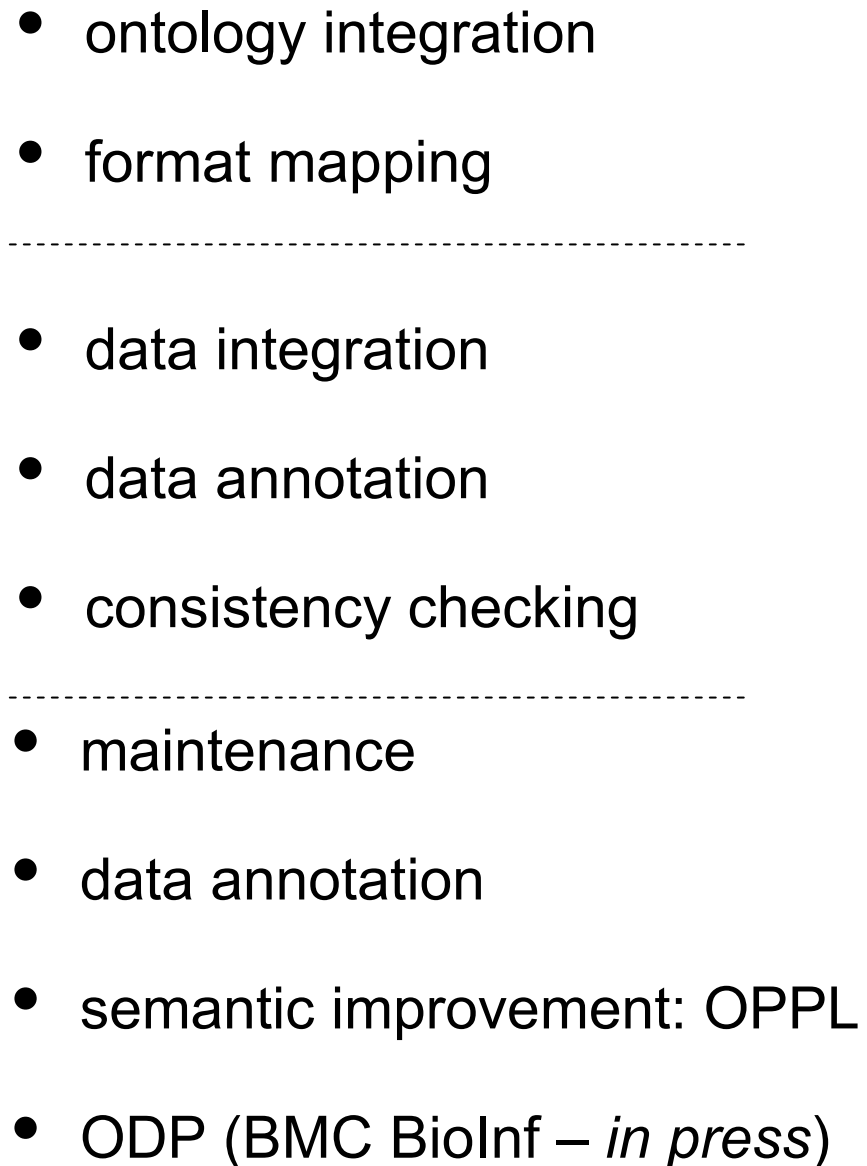
Knowledge representation

- Why OBO?
 - “Human readable”
 - Standard
 - Tools (e.g. OBOEdit)
 - <http://obo.sourceforge.net>

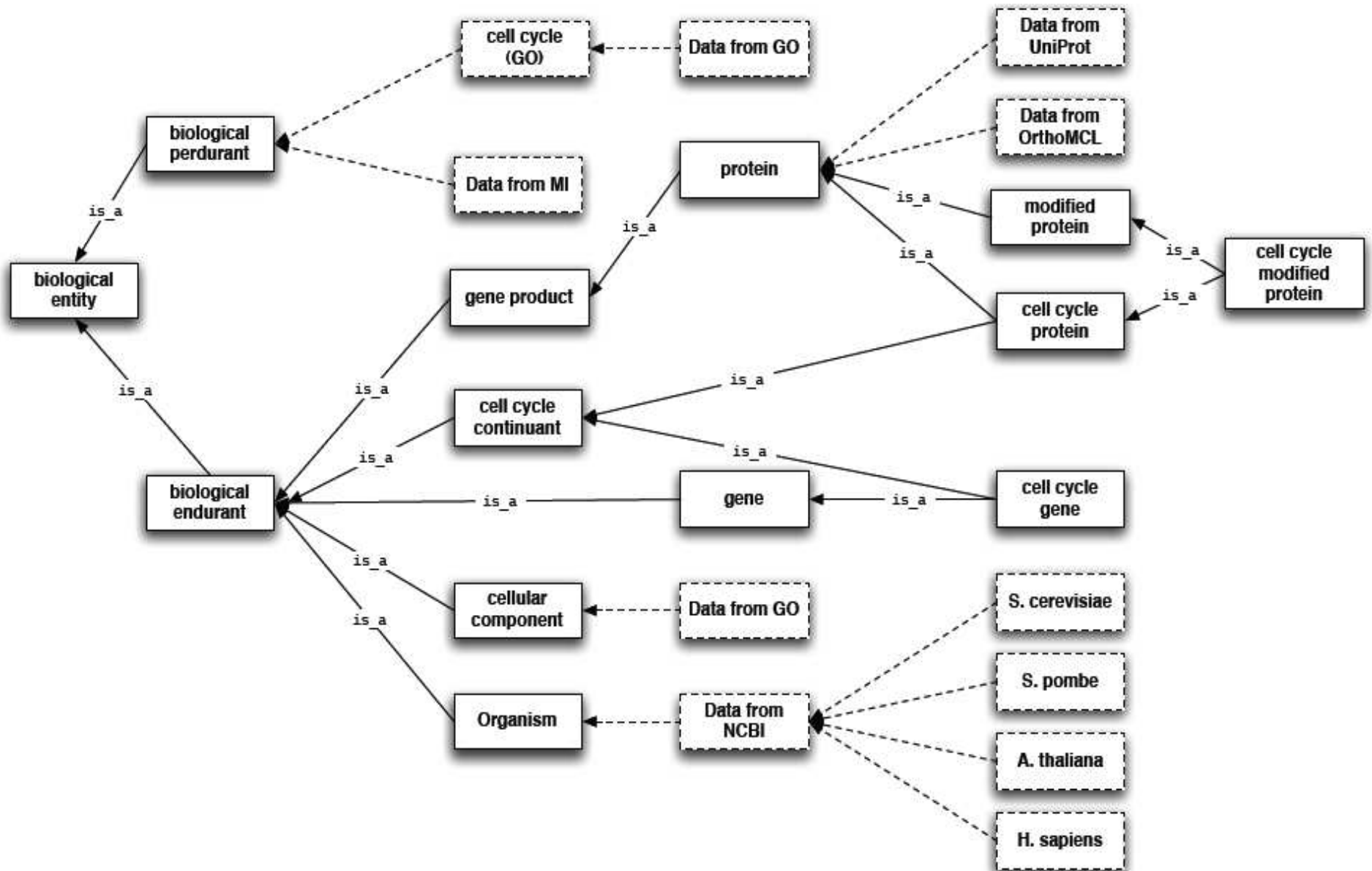


- Why OWL?
 - “Computer readable”
 - Reasoning capabilities vs. computational cost ratio
 - Formal foundation (Description Logics: <http://dl.kr.org/>)
 - <http://www.w3c.org/TR/2004/REC-owl-features-20040210>
 - **Reasoning:** RACER, Pellet, FaCT++





ULO



CCO accession number

CCO:[CPFRTIBGOU]nnnnnnnn

namespace

sub-namespaces

7 digits

C: cellular component
P: biological process
F: molecular function
R: reference
T: taxon
I: interaction
B: protein
G: gene
G: ortholog
U: upper-level term

•Examples in CCO:

CCO: P0000056 ↔ “cell cycle”

CCO: B0000046 ↔ “CYCA3;2”

•In other ontologies:

OBO_REL: has_participant

GO:0007049 ↔ “cell cycle”

Sample entry in OBO

```
[Term]
id: CCO:B0002060
name: NEB2_HUMAN
def: "Neurabin-2" [UniProt:Q96SB3]
synonym: "Neurabin-II" EXACT [UniProt:Q96SB3]
xref: UniProt:Q8TCR9
is_a: CCO:B0000000 ! core cell cycle protein
relationship: belongs_to CCO:T0000004 ! Homo sapiens organism
relationship: encoded_by CCO:G0005171 ! PPP1R9B_human
relationship: participates_in CCO:I0006401 ! aah62584-q96sb3 physical interaction
relationship: transforms_into CCO:B0013139 ! NEB2_HUMAN-Phosphoserine15
```

OBO2OWL Mapping: http://www.bioontology.org/wiki/index.php/OboInOwl:Main_Page
Tool: ONTO-PERL (*Antezana et al. Bioinformatics 2008*)

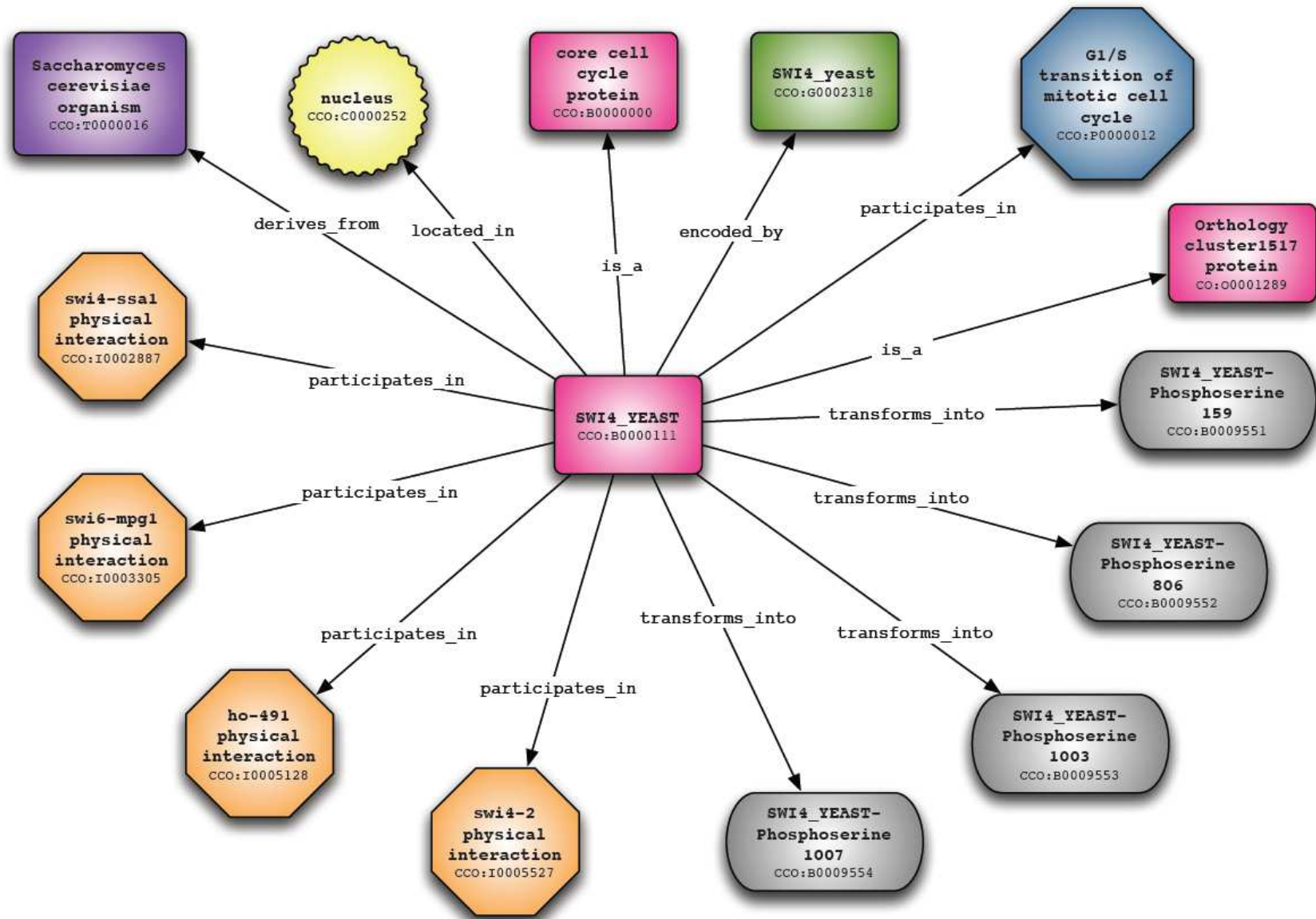
Some figures

Entity	Ontology				
	At	Hs	Sc	Sp	CCO
Proteins	252	5829	7069	930	24541
Genes	222	1806	3148	852	6028
Interactions	76	2394	5162	399	8031
Orthology groups	—	—	—	—	1649

CCO is the composite ontology = At + Hs + Sc + Sp + orthology

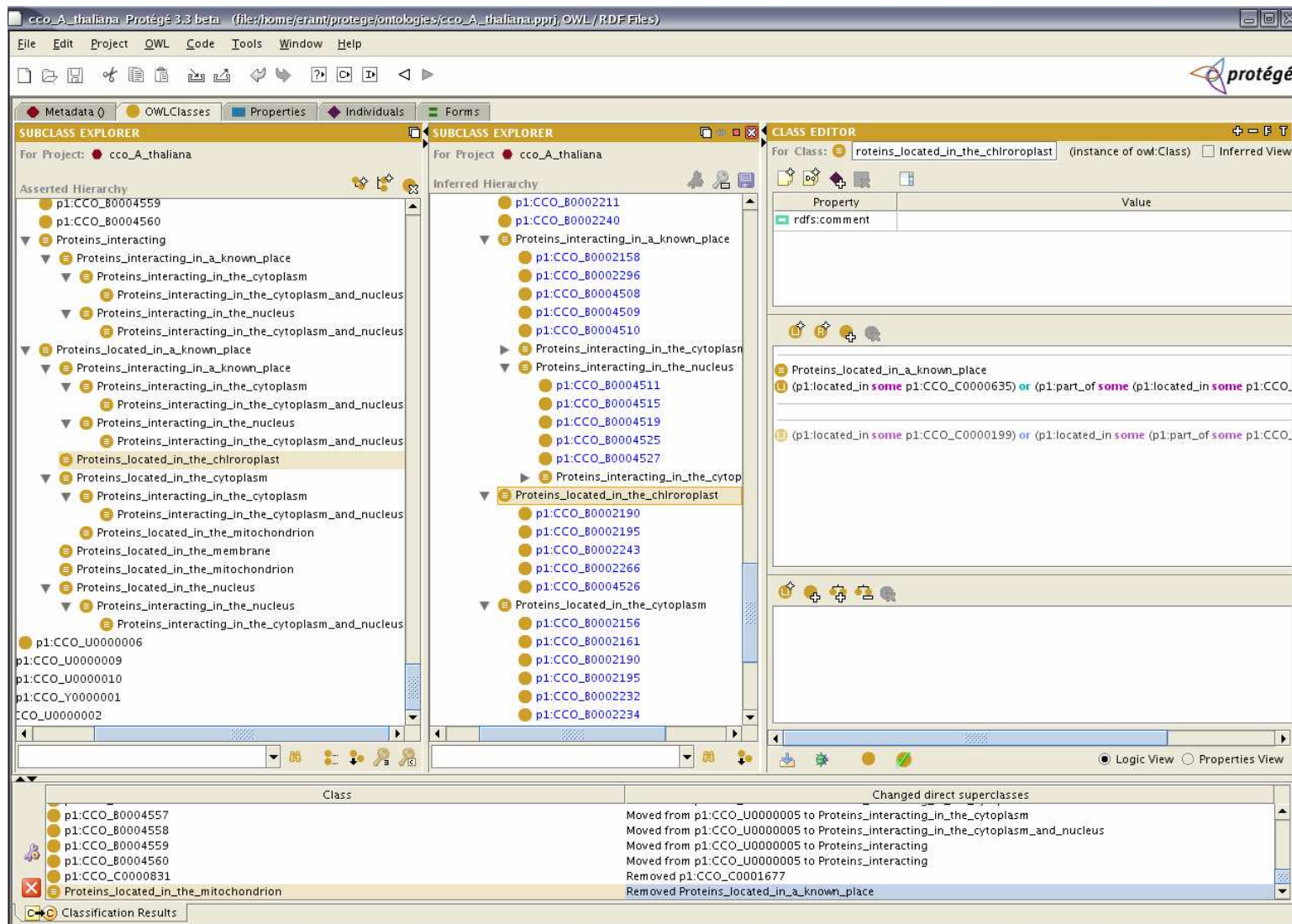
2008-03-07: 49226 terms in CCO

Current knowledge



Cellular localization checks

- Query: “If a protein is cell cycle regulated, it must **not** be located in the chloroplast (IDEM: mitochondria)” (RACER*)



The screenshot displays the Protégé 3.3 beta interface for the project 'cco_A_thaliana'. The main window shows the 'SUBCLASS EXPLORER' with two panes: 'Asserted Hierarchy' and 'Inferred Hierarchy'. The 'Asserted Hierarchy' pane shows a tree of classes, including 'Proteins_interacting_in_a_known_place', 'Proteins_interacting_in_the_cytoplasm', 'Proteins_interacting_in_the_cytoplasm_and_nucleus', 'Proteins_interacting_in_the_nucleus', 'Proteins_interacting_in_the_cytoplasm_and_nucleus', 'Proteins_located_in_a_known_place', 'Proteins_located_in_the_cytoplasm', 'Proteins_located_in_the_chloroplast', 'Proteins_located_in_the_mitochondrion', 'Proteins_located_in_the_membrane', 'Proteins_located_in_the_nucleus', and 'Proteins_interacting_in_the_nucleus'. The 'Inferred Hierarchy' pane shows a similar tree, but with additional classes like 'Proteins_interacting_in_the_cytoplasm_and_nucleus' and 'Proteins_located_in_the_cytoplasm'.

The 'CLASS EDITOR' pane on the right shows the class 'Proteins_located_in_the_chloroplast' with its properties and values. The 'Properties' tab is selected, showing a table with columns 'Property' and 'Value'. The 'Values' tab is also visible, showing a list of values for the class.

The bottom pane shows the 'Classification Results' table, which lists the classes and their direct superclasses. The table has two columns: 'Class' and 'Changed direct superclasses'. The 'Class' column lists the classes, and the 'Changed direct superclasses' column lists the superclasses that have been changed or removed.

Class	Changed direct superclasses
p1:CCO_80004557	Moved from p1:CCO_U00000005 to Proteins_interacting_in_the_cytoplasm
p1:CCO_80004558	Moved from p1:CCO_U00000005 to Proteins_interacting_in_the_cytoplasm_and_nucleus
p1:CCO_80004559	Moved from p1:CCO_U00000005 to Proteins_interacting
p1:CCO_80004560	Moved from p1:CCO_U00000005 to Proteins_interacting
p1:CCO_C0000831	Removed p1:CCO_C0001677
Proteins_located_in_the_mitochondrion	Removed Proteins_located_in_a_known_place



VisANT::File opened:http://www.CellCycleOntology.org/incubator/cco.xml

File Edit View Filters Layout Node(s) Edges MetaGraph Topology Options Expression Plugins

Weight Cutoff: 0 to 1 ☐ Edge Optimization -1.4 1.1 No Exp Exp 255 ☐ Expression Plot Play

☐ Labels ☒ Auto Fit
☒ Quick Tip Help
Select Zoom In
Pan Zoom Out
Clear Zoom Back
Reset Fit to Page

Search Compound, Pathway and
Protein/Gene: Search
SNP33_ARATH

in species
Arabidopsis thaliana

Please Register
If you don't have an account

Login

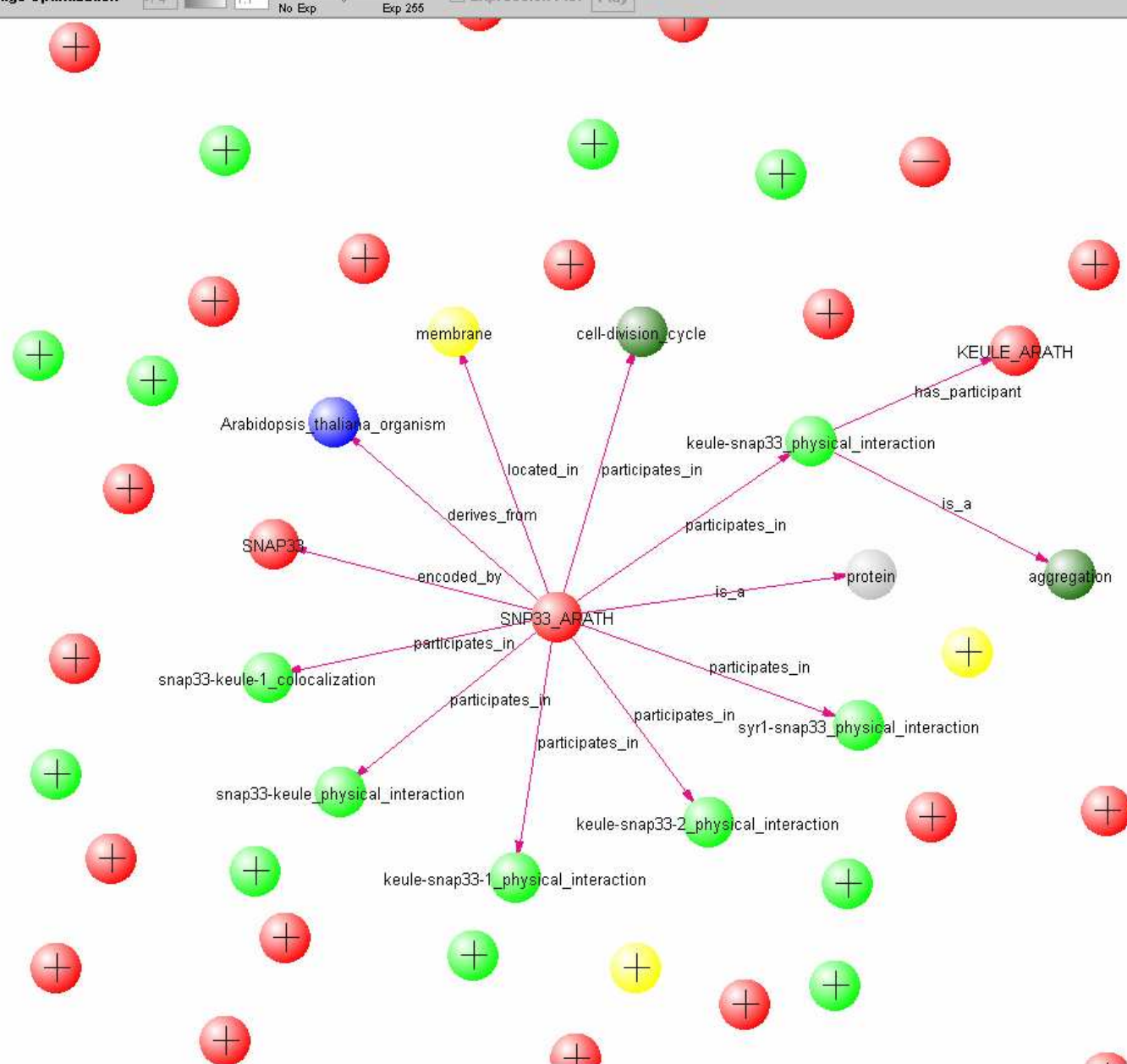
Save As Save

Delete Open

Available Files: Share

Shared Files: Open

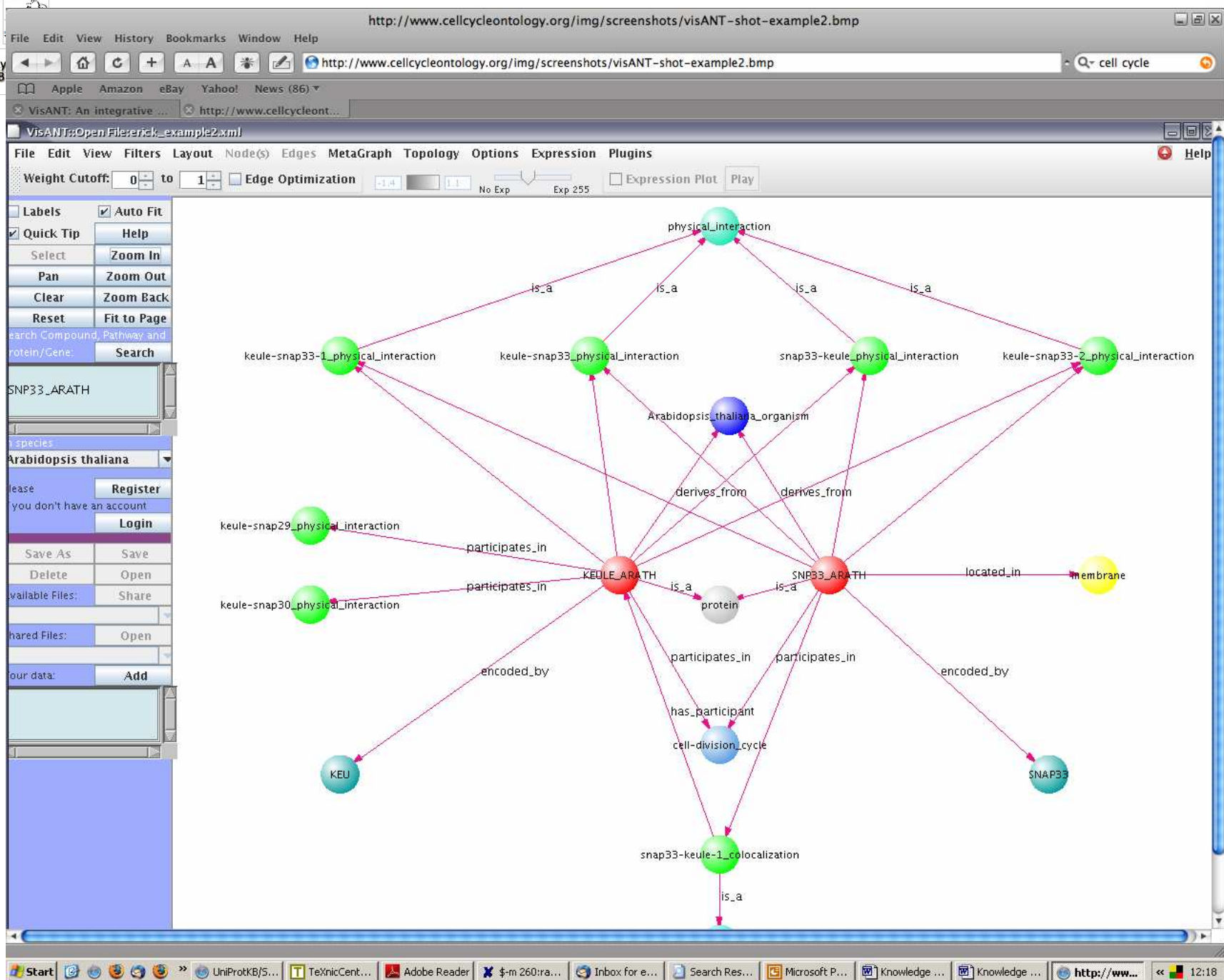
Your data: Add



N:3190,SN:0 | E:13, SE:0

Edges parsed:13

Right-click + drag to PAN, Mouse-wheel to ZOOM





MAIN MENU

- [Home](#)
- [Updates](#)
- [Download](#)
- [Query](#)
 - [SPARQL](#)
 - [OWL-DL](#)
 - [OLS](#)
 - [BioPortal](#)
- [Documentation](#)
- [Tools](#)
- [About](#)

[Home](#) > [Query](#) > [SPARQL](#)

SPARQL



SPARQL stands for **SPARQL Protocol and RDF Query Language**. It is standardized by the *RDF Data Access Working Group* (DAWG) of the W3C. It allows for a query to consist of triple patterns, conjunctions, disjunctions, and optional patterns.

Querying CCO

The following form lets you query the Cell Cycle Ontology through a [SPARQL endpoint](#) hosted at [Plant Systems Biology](#) department of the [Flanders Institute for Biotechnology](#). The underlying triplestore contains over 1 million RDF triples of cell cycle information. This information ranges from processes, interactions, proteins, genes, cellular compartments, and so forth, which were collected from diverse sources (like GO, UniProt, IntAct, etc.).

Query:

```
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
SELECT ?term_id ?term_name
FROM <http://www.cellcycleontology.org/ontology/rdf/CCO>
WHERE {
  ?term_id rdfs:label ?term_name.
  filter regex(?term_name, "cell cycle")
}
```

Run Query

SPARQL queries against CCO are run on [Virtuoso \(OpenLink\)](#). This system provides an infrastructure for storing and querying CCO.

Example 1

The following SPARQL query returns all the proteins (IDs) participating in any known process:

```
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX sp:<http://www.cellcycleontology.org/ontology/rdf/Sp#>
SELECT ?prot ?prot_label ?biological_process
```


Cell Cycle Ontology (A. thaliana)

Tree View

Tree view constructed based on *is_a* hierarchy

- entity
 - continuant
 - cell cycle continuant
 - cellular_component
 - gene
 - gene product
 - protein
 - cell cycle modified protein
 - cell cycle protein
 - core cell cycle protein
 - APC10_ARATH
 - ARTE_ARATH
 - ATK1_ARATH
 - ATK3_ARATH
 - ATM_ARATH
 - ATR_ARATH
 - BRE1A_ARATH
 - BSH_ARATH
 - CCA11_ARATH
 - CCA12_ARATH
 - CCA21_ARATH

Class/Type Details

General

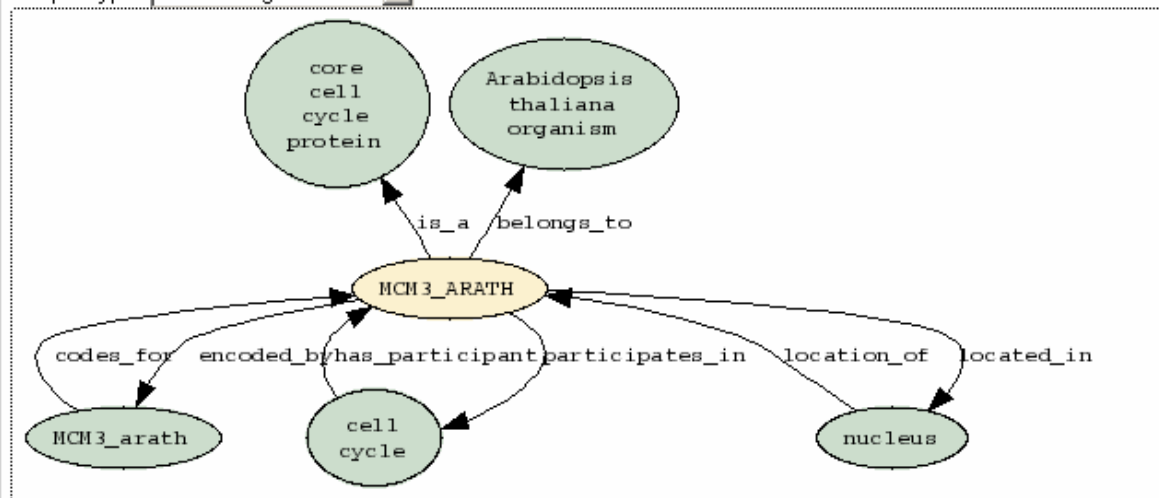
Class/Type Name **MCM3_ARATH**
 Id **CCO:B0002385**

Attributes

Definition **DNA replication licensing factor MCM3 homolog**
 EXACT SYNONYM **"Minichromosome maintenance protein 3 homolog"**
 Database_References **UniProt:Q9FL33**

Graph View

Graph Type



http://www.bioontology.org/ncbo/faces/pages/ontology_list.xhtml

http://www.ebi.ac.uk/ontology-lookup/browse.do?ontName=CCO

ftp://ftp.gen...oPF_noXP.obo Park 79 Hotel Reservations BioPortal SPARQL

EMBL-EBI EB-eye Search All Databases Enter Text Here Go Reset Advanced Search Give us feedback

Databases Tools EBI Groups Training Industry About Us Help Site Index

OLS - Ontology Lookup Service

CCO Ontology Browser

entity

- continuant
 - cellular_component
 - reference
 - organism
 - gene product
 - protein
 - core cell cycle protein
 - SEM1_YEAST
 - CDC28_YEAST
 - TBB_YEAST
 - TOP1_YEAST
 - DNLI_YEAST
 - CDC25_YEAST
 - NOT2_YEAST
 - CDC37_YEAST
 - SC160_YEAST
 - PIS_YEAST
 - KIN28_YEAST
 - CDC7_YEAST
 - KAPA_YEAST
 - CDC31_YEAST
 - RAD1_YEAST
 - RAD52_YEAST
 - TOP2_YEAST
 - CALM_YEAST
 - RAD10_YEAST
 - IF4E_YEAST

Help (hide)

Double-click a term to see its children. The ontology browser is populated dynamically. If there are many children for a given term, there may be a small delay while the browser fetches it. **Click** to highlight a term to see any information associated with it. **Hover** over a term to see its relation with its immediate parent. Root terms will not display any relational information.

Relations

SEM1_YEAST is_a core cell cycle protein

Term Information

ID: Zoom

Name:

Associated information

Highlight a term to view its associated information.

Term Hierarchy

Paths to Root: ☒ Child relationships: ☐

Legend:

is a

part of

OLS Home

Documentation

- Project
- Publications

Developer Resources

- Download
- Implementation Overview
- Javadoc
- Webservice documentation

Contact Us

- Acknowledgements

News

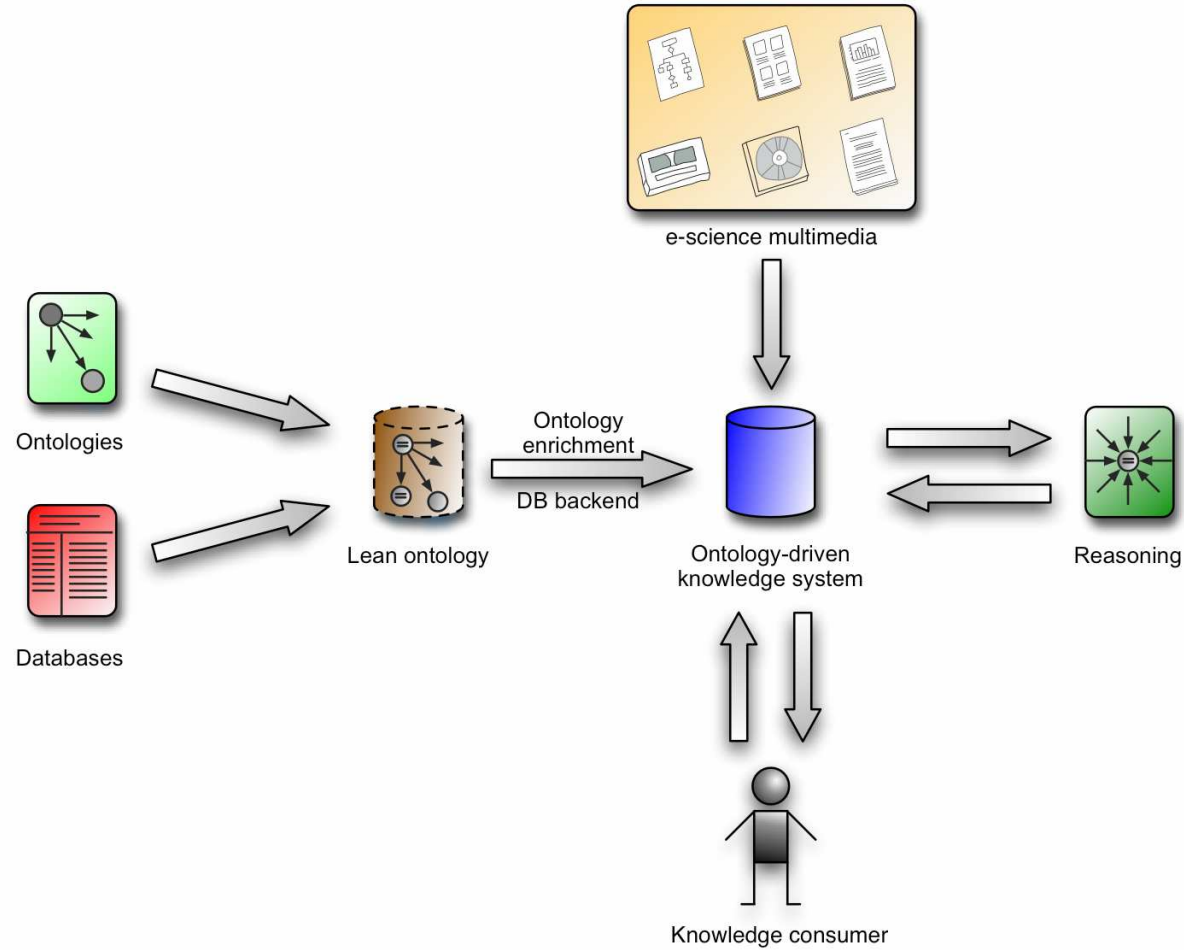
June 2007: Maintenance Release

Implementation documentation has been updated to include more recent dependencies. Please note that unless stated otherwise in specific instances, newer versions of given dependencies should work without issue.

April 2007: Maintenance Release

There is a new release of the OLS available, which includes mostly maintenance fixes and

Outlook into the future

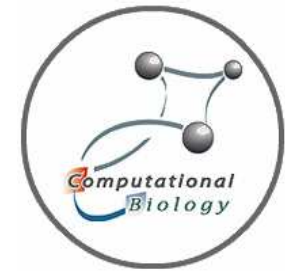


Current issues

- Temporal & spatial representation
 - OBOF not enough...
- Performance (reasoners)
 - Huge ontologies
- Weighed knowledge (*often , sometimes*)

Acknowledgements

- Martin Kuiper (U Ghent/VIB)
- Vladimir Mironov (U Ghent/VIB)
- Mikel Egaña (U Manchester)
- Robert Stevens (U Manchester)
- Ward Blonde (U Ghent)
- Bernard De Baets (U Ghent)
- CCO Users



<http://www.CellCycleOntology.org>