



# A cell-cycle knowledge integration framework

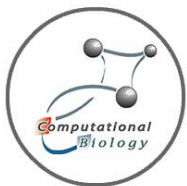
Erick Antezana

Dept. of Plant Systems Biology. Flanders Interuniversity Institute for Biotechnology/Ghent University.

Ghent BELGIUM.

[erant@psb.ugent.be](mailto:erant@psb.ugent.be)

<http://www.psb.ugent.be/cbd/>



# Overview



- Introduction
- Aim
- Data integration pipeline
- CCO engineering
- Exploiting reasoning services
- Conclusions
- Future work

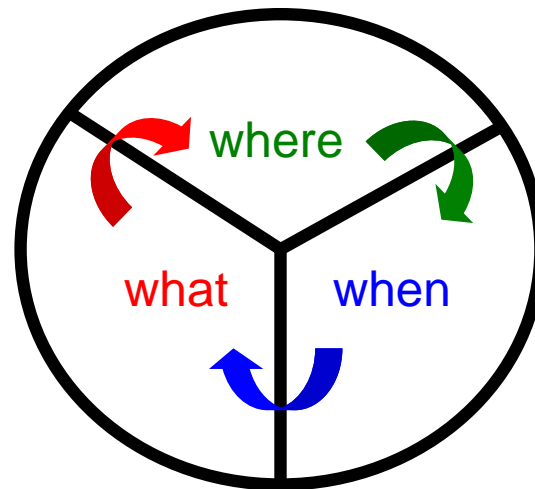
# Introduction



- Amount of data generated in biological experiments continues to grow exponentially
- Shortage of proper approaches or tools for analyzing this data has created a **gap** between **raw data** and **knowledge**
- Lack of a structured documentation of knowledge leaves much of the data extracted from these **raw data unused**
- Differences in the technical languages used (**synonymy** and **polysemy**) have complicated the analysis and interpretation of the data

# Aim

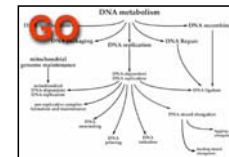
- Capture the knowledge of the CC process especially the dynamic aspects of the terms and their interrelations, and to promote sharing, reuse and enable better computational integration with existing resources
- Sample: “Cyclin B (*what*) is located in Cytoplasm (*where*) during Interphase (*when*)”



- This will allow biologists to ask questions to the KB
- Four model organisms: At, Sc, Sp, Hs

# Method

- CCO should capture the semantics of the temporal aspects and dynamics of the cell-cycle process
- CCO forms the knowledge base core
- Knowledge representation: OBO and OWL-DL
- Existing relationships have been extended
- Data sources:
  - Association files (GO)
  - PPI data: IntAct, BIND, DIP
  - Reactome
  - Cell-cycle functional data
  - Data obtained using bioinformatics



# OBO and OWL

- Open Biomedical Ontologies: OBO



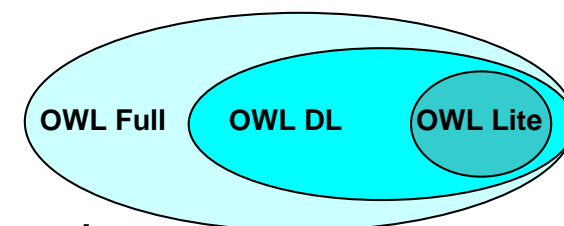
- Standard

- “Human readable”

- Tools (e.g. OBOEdit)

- <http://obo.sourceforge.net>

- Web Ontology Language: OWL (Full, DL, Lite)



- Reasoning capabilities vs. computational cost ratio

- “Computer readable”

- Formal foundation (Description Logics: <http://dl.kr.org/>)



- <http://www.w3c.org/TR/2004/REC-owl-features-20040210>

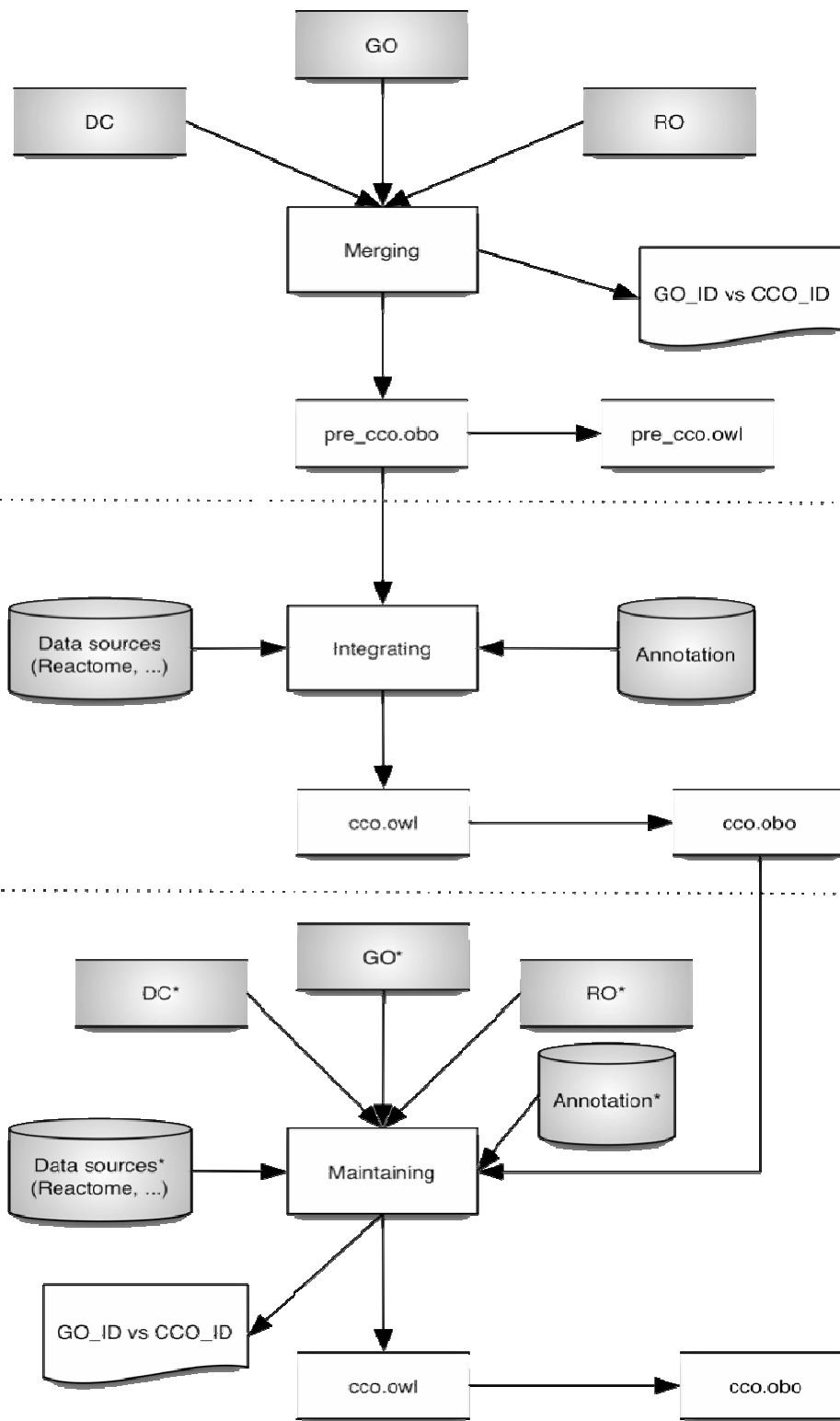


# Pipeline

Set-up

Data inclusion

Life-cycle



- ontology integration
- format mapping
- data integration
- data annotation
- consistency checking
- maintenance
- data annotation
- semantic improvement

# Reusing ontologies



- GO only considers subsumption ([is\\_a](#)) and partonomic inclusion ([part\\_of](#)).
- Maintainability issues in GO.
- GO and the RO: core ontologies in CCO
- All the processes from GO under the cell-cycle (GO:0007049) term were taken into account, while RO was completely imported.
- 304 terms adopted from GO
- 15 relationships from RO.
- The CCO is updated daily and checked using data from GO.



# Motivating scenarios



- **Molecular biologist:** interacting components, events, roles that each component play. Hypothesis evaluation.
- **Bioinformatician:** data integration, annotation, modeling and simulation.
- **General audience:** educational purposes.

# Competency questions



- What is a X-type **CDK**?
- What is Y-type **cyclin**?
- In what events is **CDK Z** involved?
- In what events does **Rb** participate?
- Which **CDKs** are involved in the endoreduplication process?
- Which proteins are phosphorylated by **kinase X**?
- Which **CDK** pertains to [G1 | S | G2 | M] phase?

# Formats mapping: OBO $\Leftrightarrow$ OWL



- Mapping not totally **biunivocal**; however, all the data has been preserved.
- Missing properties in OWL relations:
  - reflexivity,
  - asymmetry,
  - intransitivity and
  - partonomic relationships.
- Existential and universal restrictions cannot be explicitly represented in OBO => **Consider all as existential.**

# Mapping: obo2owl terms

OBO keyword	OWL keyword	OWL element type
[Term]	owl:Class	Class description
id	rdf:ID	Class description
name	rdfs:label	rdf:Property
is_anonymous	NDY	NDY
alt_id	NDY	NDY
def	rdfs:comment	rdf:Property
comment	NDY	NDY
subset	NDY	NDY
synonym	synonym	owl:DataTypeProperty, owl:AnnotationProperty
xref	xref	owl:DataTypeProperty, owl:AnnotationProperty
is_a	rdfs:subClassOf	owl:ObjectProperty
intersection_of	owl:intersectionOf	Class description
union_of	owl:unionOf	Class description
disjoint_of	owl:disjointWith	Class axiom
relationship	NDY	NDY
is_obsolete	owl:DeprecatedClass	Version information
replaced_by	NDY	NDY
consider	owl:equivalentClass	Class axiom

# Mapping: obo2owl relationships

OBO keyword	OWL keyword
[Typedef]	owl:ObjectProperty
builtin	NDY
comment	NDY
def	rdfs:comment
exact_synonym	synonym (workaround)
id	rdf:id
inverse_of	owl:inverseOf
is_a	rdfs:subClassOf
is_anti_symmetric	is_anti_symmetric (workaround)
is_reflexive	is_reflexive (workaround)
is_transitive	rdf:type (TransitiveProperty)
NDY	rdf:type (SymmetricProperty)
name	rdfs:label (string)
xref_analog	NDY

# CCO accession number

CCO : [ C P F R T I B ] n n n n n n n n

namespace                      sub-namespace                      7 digits

C: cellular component  
P: biological process  
F: molecular function  
R: reference  
T: taxon  
I: interaction  
B: biomolecule

Examples in CCO:

CCO: P0000056 (*“cell cycle”*)

CCO: B0001314 (*“p53\_human”*)

In other ontologies:

OBO\_REL: has\_participant

GO:0007049 (*“cell cycle”*)

# CCO entry CCO:P0000016

```
[Term]
id: CCO:P0000016
name: M phase of mitotic cell cycle
def: "Progression through M phase, the part of the mitotic cell cycle during which mitosis and cytokinesis take place." [GOC:mah, ISBN:0815316194]
xref: GO:0000087
xref: Reactome:68886
relationship: part_of CCO:P0000037
is_a: CCO:P0000038
synonym: "M-phase of mitotic cell cycle" [] {scope="exact"}
```

```
<owl:Class rdf:ID="CCO_P0000016">
  <rdfs:label xml:lang="en">M phase of mitotic cell cycle</rdfs:label>
  <xref rdf:datatype="http://www.w3.org/2001/XMLSchema#string">GO:0000087</xref>
  <xref rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Reactome:68886</xref>
  <rdfs:comment rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Progression through M phase, the part of the mitotic cell cycle during which mitosis and cytokinesis take place.</rdfs:comment>
  <synonym rdf:datatype="http://www.w3.org/2001/XMLSchema#string">M-phase of mitotic cell cycle</synonym>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#part_of"/>
      <owl:someValuesFrom rdf:resource="#CCO_P0000037"/>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf rdf:resource="#CCO_P0000038"/>
  <owl:disjointWith rdf:resource="#CCO_P0000270"/>
</owl:Class>
```

# CCO entry CCO:P0000016

```
[Term]
id: CCO:P0000016
name: M phase of mitotic cell cycle
def: "Progression through M phase, the part of the mitotic cell cycle during which mitosis and cytokinesis take place." [GOC:mah, ISBN:0815316194]
xref: GO:0000087
xref: Reactome:68886
relationship: part_of CCO:P0000037
is_a: CCO:P0000038
synonym: "M-phase of mitotic cell cycle" [] {scope="exact"}
```

```
<owl:Class rdf:ID="CCO_P0000016">
  <rdfs:label xml:lang="en">M phase of mitotic cell cycle</rdfs:label>
  <xref rdf:datatype="http://www.w3.org/2001/XMLSchema#string">GO:0000087</xref>
  <xref rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Reactome:68886</xref>
  <rdfs:comment rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Progression through M phase, the part of the mitotic cell cycle during which mitosis and cytokinesis take place.</rdfs:comment>
  <synonym rdf:datatype="http://www.w3.org/2001/XMLSchema#string">M-phase of mitotic cell cycle</synonym>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#part_of"/>
      <owl:someValuesFrom rdf:resource="#CCO_P0000037"/>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf rdf:resource="#CCO_P0000038"/>
  <owl:disjointWith rdf:resource="#CCO_P0000270"/>
</owl:Class>
```



# CCO entry CCO:P0000016

```
[Term]
id: CCO:P0000016
name: M phase of mitotic cell cycle
def: "Progression through M phase, the part of the mitotic cell cycle during which mitosis and cytokinesis take place." [GOC:mah, ISBN:0815316194]
xref: GO:0000087
xref: Reactome:68886
relationship: part_of CCO:P0000037
is_a: CCO:P0000038
synonym: "M-phase of mitotic cell cycle" [] {scope="exact"}
```

```
<owl:Class rdf:ID="CCO_P0000016">
  <rdfs:label xml:lang="en">M phase of mitotic cell cycle</rdfs:label>
  <xref rdf:datatype="http://www.w3.org/2001/XMLSchema#string">GO:0000087</xref>
  <xref rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Reactome:68886</xref>
  <rdfs:comment rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Progression through M phase, the part of the mitotic cell cycle during which mitosis and cytokinesis take place.</rdfs:comment>
  <synonym rdf:datatype="http://www.w3.org/2001/XMLSchema#string">M-phase of mitotic cell cycle</synonym>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#part_of"/>
      <owl:someValuesFrom rdf:resource="#CCO_P0000037"/>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf rdf:resource="#CCO_P0000038"/>
  <owl:disjointWith rdf:resource="#CCO_P0000270"/>
</owl:Class>
```

# CCO entry CCO:P0000016

```
[Term]
id: CCO:P0000016
name: M phase of mitotic cell cycle
def: "Progression through M phase, the part of the mitotic cell cycle during which mitosis and cytokinesis
take place." [GOC:mah, ISBN:0815316194]
xref: GO:0000087
xref: Reactome:68886
relationship: part of CCO:P0000037
is_a: CCO:P0000038
synonym: "M-phase of mitotic cell cycle" [] {scope="exact"}
```

```
<owl:Class rdf:ID="CCO_P0000016">
  <rdfs:label xml:lang="en">M phase of mitotic cell cycle</rdfs:label>
  <xref rdf:datatype="http://www.w3.org/2001/XMLSchema#string">GO:0000087</xref>
  <xref rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Reactome:68886</xref>
  <rdfs:comment rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Progression through M phase, the
part of the mitotic cell cycle during which mitosis and cytokinesis take place.</rdfs:comment>
  <synonym rdf:datatype="http://www.w3.org/2001/XMLSchema#string">M-phase of mitotic cell cycle</synonym>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#part_of"/>
      <owl:someValuesFrom rdf:resource="#CCO_P0000037"/>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf rdf:resource="#CCO_P0000038"/>
  <owl:disjointWith rdf:resource="#CCO_P0000270"/>
</owl:Class>
```

# Reasoning capabilities



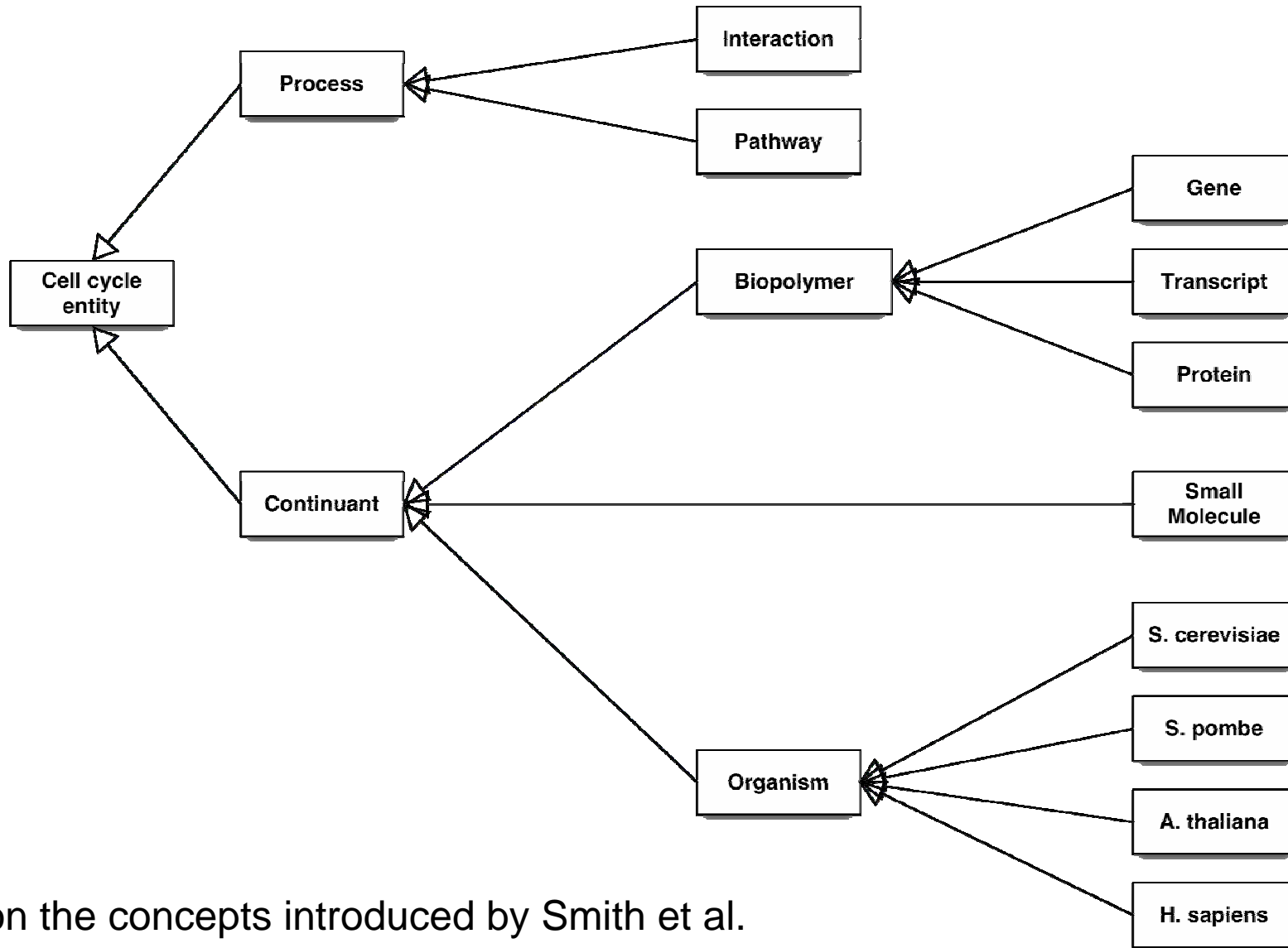
- OWL-DL: mathematical foundation (description logics)
- Automatic detection and handling of inconsistencies and misclassifications
- Reasoners (e.g. RACER, Pellet)
- Protégé (DIG interface)

# Single inheritance principle



- **Principle:** *“No class in a classification should have more than one is\_a parent on the immediate higher level” (Smith B. et al.)*
- Detecting the relationships which violate that rule using a reasoner
- Solution: disjoint among the terms at the same level of the structure
- 32 problems found:
  - 4: “**part\_of**” instead of “**is\_a**”
  - 18: should stay without any change (FP)
  - 10: not consistent (used terminology)

# Upper Level Ontology



Based on the concepts introduced by Smith et al.

# CCO status



- #relationships = #RO + #CCO = 15 + 5 = 20
- #terms = 15 (ULO) + 304 (process branch) + 20 (xref, ref, etc)
- #interactions = 124 (IntAct)
- #genes/proteins/transcripts = 1961
  - TAIR: 228
  - GeneDB\_Spombe: 1032
  - GOA Human: 1292
  - SGD: 798

# CCO in OBO Edit

OBO-Edit version 1.000-beta17: pre\_cco.obo

File Edit Plugins Help

← involved in NEK1\_HUMAN  
 ← involved in NEK1\_HUMAN  
 ← involved in NEK2\_HUMAN  
 ← involved in NEK3\_HUMAN  
 ← involved in NEK4\_HUMAN  
 ← involved in NEK9\_HUMAN  
 ← involved in NF1\_HUMAN  
 ← involved in NIPA\_HUMAN  
 ← involved in NIPBL\_HUMAN  
 ← involved in nnt1  
 ← involved in NOLC1\_HUMAN  
 ← involved in NSMA2\_HUMAN  
 ← involved in NUDC\_HUMAN  
 ← involved in O14777\_HUMAN  
 ← involved in O14835\_HUMAN  
 ← involved in O95082\_HUMAN  
 ← involved in O95607\_HUMAN  
 ← involved in P53\_HUMAN  
 ← involved in P73\_HUMAN  
 ← involved in PAPD5\_HUMAN  
 ← involved in PAR3\_HUMAN  
 ← involved in PAR6A\_HUMAN  
 ← involved in PAR6B\_HUMAN  
 ← involved in PAR6G\_HUMAN  
 ← involved in PARC\_HUMAN  
 ← involved in PARD3\_HUMAN  
 ← involved in PCAF\_HUMAN  
 ← involved in PCL1  
 ← involved in PCL2  
 ← involved in PCL5  
 ← involved in PCL9  
 ← involved in PCNP\_HUMAN  
 ← involved in PFD1\_HUMAN  
 ← involved in PHO85  
 ← involved in PIN1\_HUMAN  
 ← involved in PININ\_HUMAN

Term filter Advanced Options 7 results

ID	Name
CCO:80001314	P53_HUMAN
CCO:80001474	PLK1_HUMAN
CCO:80001473	PPP5_HUMAN

Autoselect Select terms Results label 7 results

ID: CCO:B0001314  
 Namespace: cco  
 Term name: P53\_HUMAN

Definition Comment \* Cross Products

TAS

Synonyms Dbxrefs \*

General DbXrefs  
 UniProt:P04637 Select a dbxref from the list to edit it, or press add to create a new dbxref

Add Del

Commit

DAG Viewer

- cell cycle
  - involved in P53\_HUMAN
  - regulation of cell cycle
    - regulation of progression through cell cycle
      - negative regulation of progression through cell cycle
        - cell cycle arrest
          - involved in P53\_HUMAN
        - cell cycle checkpoint
          - involved in P53\_HUMAN

- biomolecular entity
- protein
  - P53\_HUMAN

4 paths loaded.  Multi-select  Collapse paths



# CCO in Protégé\*

pre\_cco Protégé 3.0 (file:/Users/erick/Desktop/marseille/pre\_cco.ppr, OWL Files (.owl or .rdf))

File Edit Project OWL Wizards Code Window Help

OWLClasses Properties Forms Individuals Metadata

SUBCLASS RELATIONSHIP For Project: pre\_cco

Asserted Hierarchy

- JniProt:Q13564
- JniProt:P41271
- JniProt:Q9BTV7
- JniProt:Q7L8A9
- JniProt:Q9BUP3
- JniProt:Q7FTZ3
- JniProt:Q9LIL4
- JniProt:Q9I-307
- JniProt:Q92888
- JniProt:Q92097
- JniProt:Q95150
- JniProt:Q95229
- JniProt:P04337
- JniProt:P04554
- JniProt:P01111
- JniProt:P01106
- JniProt:P57359
- JniProt:P56524
- JniProt:Q43236
- JniProt:Q43196
- JniProt:P02775
- JniProt:P01589
- JniProt:Q9LFR5
- JniProt:Q9LFKE
- ENSEMBL:ENSP00000348457
- ENSEMBL:ENSP00000350273
- JniProt:Q9BYG4
- JniProt:Q9BYG6
- JniProt:Q14203
- JniProt:Q14204
- JniProt:P51587
- JniProt:Q60934

CLASS EDITOR For Class: UniProt:P04637 (instance of owl:Class, internal name is CCO\_B0J01314)

Name SameAs DifferentFrom

CCO\_B000314

rdfs:comment

Annotations

Property	Value	Lang
rdfs:label	P53_HUMAN	en
source	UniProt:P04637	

Asserted Inferrec

Asserted Conditions

- involved\_in GO:0007049
- involved\_in Reactome:69620
- involved\_in GO:0007050

involved\_in

- GO:0007049
- Reactome:69620
- GO:0007050

Disjoints

Logic View Properties View

\* <http://protege.stanford.edu>



# CCO API



- Set of PERL modules influenced by go-perl
- Features:
  - OBO parsing
  - Ontology handling
  - obo2owl, owl2obo
  - XSL transformations

# CCO availability



- <http://www.sbcellcycle.org/cco/html/index.html>
- OBO, OWL, XML and API (Perl)
- Very soon: advanced queries
- Very soon: <http://www.CellCycleOntology.org>
- “A cell-cycle knowledge integration framework”. Data Integration in Life Sciences, DILS 2006, LNBI 4075, pp. 19-34, 2006.

# CCO online




Cell Cycle Ontology Browser - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://www.sbcellcycle.org/cco/html/cco-lookup.html

W Knowledge represent... Diamonds Onto DL OWL LaTeX Perso poster Bio KW Tools journals perl meet CC iso tmp tmp2 tmp3 >>



## Cell Cycle Ontology Explorer

### Look up terms in the Cell Cycle Ontology

Start typing a term name in the text box below. A list of prospective matches will pop-up, either choose the one you are looking from that list or continue typing the entire name. The pointed out term will be displayed on the right pane. Type for instance **mitotic cell cycle** and then click on the button named 'Look up' to retrieve the entry that has as name **mitotic cell cycle**.

P53\_HUMAN ?

Look up Reset

id:	CCO:B0001314		
name:	P53_HUMAN		
is_a:	<a href="#">protein</a>		
relationship:			
type:	participates_in	target:	<a href="#">cell cycle</a>
type:	participates_in	target:	<a href="#">cell cycle arrest</a>

id:	CCO:B0001315		
name:	IFNW1_HUMAN		
is_a:	<a href="#">protein</a>		
relationship:			
type:	participates_in	target:	<a href="#">cell cycle arrest</a>

id:	CCO:B0001316		
name:	INHA_HUMAN		
is_a:	<a href="#">protein</a>		
relationship:			
type:	participates_in	target:	<a href="#">cell cycle arrest</a>

id:	CCO:B0001317		
-----	--------------	--	--

Done

# Conclusions



- A data integration pipeline prototype covering the entire life cycle of the knowledge base.
- Concrete problems and initial results related to the implementation of automatic format mappings between ontologies and inconsistency checking issues are shown.
- Existing integration obstacles due to the diversity of data formats and lack of formalization approaches as well as the trade-offs that are common in biological sciences.

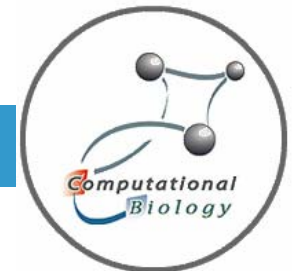
# Future work



- The knowledge will be weighted or scored according to some defined evidence codes expressing the support media similar to those implemented in GO (experimental, electronically inferred, and so forth).
- A [query system](#) and a web user interface are also foreseen. The ultimate aim of the project is to support [hypothesis evaluation](#) about cell-cycle regulation issues.

# Acknowledgments

- Martin Kuiper (UGent/VIB)
- Vladimir Mironov (UGent/VIB)
- Elena Tshiporkova (UGent/VIB)
- Mikel Egaña (Manchester University, Robert Stevens' group)



# A cell-cycle knowledge integration framework

Erick Antezana

Dept. of Plant Systems Biology. Flanders Interuniversity Institute for Biotechnology/Ghent  
University. Ghent BELGIUM.

[erant@psb.ugent.be](mailto:erant@psb.ugent.be)

