The Cell Cycle Ontology: an application ontology for data integration



^ ^ * * ***

NTNU

Erick Antezana¹, Vladimir Mironov^{1,4}, Mikel Egaña², Robert Stevens², Ward Blondé³, Bernard De Baets³, and Martin Kuiper^{1,4}

1 VIB Dept. of Plant Systems Biology, Ghent University. Technologiepark 927, B-9052 Ghent, Belgium. E-mail: erick.antezana@gmail.com

2 The University of Manchester. School of Computer Science, Oxford Road, M13 9PL Manchester, UK. E-mail: {eganaarm, stevensr}@cs.man.ac.uk

3 Ghent University, Department of Applied Mathematics, Biometrics and Process Control, Coupure links 653, 9000 Ghent, Belgium. E-mail: {ward.blonde, bernard.debaets}@ugent.be

4 Norwegian University of Science and Technology, Department of Biology, Høgskoleringen 5, 7491 Trondheim, Norway. E-mail: {mironov, kuiper}@nt.ntnu.no

http://www.CellCycleOntology.org



Objective

To capture the knowledge about the **cell cycle** process (particularly its dynamic facets) and to promote sharing, reuse and enable better computational integration with existing resources (semantic web). The ultimate aim is to support evaluation and generation of hypotheses via reasoning services about cell-cycle regulation. **Target organisms:** *S. cerevisiae, S. pombe, A. thaliana and H. sapiens*.



<pre>Fig 3. CCO in OBO-Edit.</pre>							
cco_A_thaliana Protégé 3.3 beta (filer/home/erant/protege/or ile Edit Project OWL Code Tools Window Help	nologies/cco_A_thaliana.pprj, OWL / RDF Files)						
SUBCLASS EXPLORER	SUBCLASS EXPLORER CLASS EDITOR $+ - F T$						
or Project: 🜒 cco_A_thaliana	For Project Co_A_thaliana For Class: Cass: Cass:						
sserted Hierarchy 😵 🕻 p1:CCO_80004559	Interred Hierarchy Image: Color Image:						
p1:CCO_B0004560 Proteins interacting	p1:CCO_B0002240 Proteins interacting in a known place						
 Proteins_interacting_in_a_known_place 	● p1:CCO_B0002158						
Proteins_interacting_in_the_cytoplasm Proteins_interacting_in_the_cytoplasm_and_nucle	eus p1:CCO_80002296						
Proteins_interacting_in_the_nucleus	● p1:CCO_B0004509						
Proteins_interacting_in_the_cytoplasm_and_nucle Forteins_located_in_a_known_place	2us p1:CCO_B0004510 C C C B0004510 C C C C B0004510 C C C C C C C C C C C C C C C C C C C						
For the second secon	Proteins_interacting_in_the_nucleus Proteins_located_in_a_known_place						
Proteins_interacting_in_the_cytoplasm Proteins_interacting_in_the_cytoplasm_and_nucle	eus p1:CCO_B0004511 (p1:located_in some p1:CCO_C0000635) or (p1:part_of some (p1:located_in some p1:CCO_ eus						
V SProteins_interacting_in_the_nucleus	p1:CCO_B0004519 p1:CCO_B0004519 p1:CCO_B00045175 p1:CCO_B00045755 p1:CCO_B00045755 p1:CCO_B00045755 p1:CCO_B00045755 p1:CCO_B00045755						
Proteins_interacting_in_the_cytoplasm_and_nucle Proteins_located_in_the_chlroroplast	eus pr:CCO_B0004525						
Proteins_located_in_the_cytoplasm Onetring_interpreting_in_the_sytoplasm	Proteins_interacting_in_the_cytop Destring logated in the chlorenplast						
 Proteins_interacting_in_the_cytoplasm Proteins_interacting_in_the_cytoplasm_and_nucle 	eus p1:CCO_B0002190						
Proteins_located_in_the_mitochondrion	p1:CCO_B0002195						
Proteins_located_in_the_membrane Proteins_located_in_the_mitochondrion	● p1:CCO_80002245						
Proteins_located_in_the_nucleus Order interpretation in the nucleus	p1:CCO_B0004526 Image: Constraint of the system is the system						
 Proteins_interacting_in_the_nucleus Proteins_interacting_in_the_cytoplasm_and_nucle 	eus p1:CCO_B0002156						
p1:CCO_U0000006	p1:CCO_B0002161						
:CCO_U0000010	p1:CCO_B0002195						
:CCO_Y0000001	p1:CCO_B0002232						
0_0000002							
· · · · · · · · · · · · · · · · · · ·	R De Logic View Properties View						
Class	Changed direct superclasses						
p1:CCO_80004558	Moved from p1:CCO_U0000005 to Proteins_interacting_in_the_cytoplasm_and_nucleus						
p1:CCO_B0004559	Moved from p1:CCO_U0000005 to Proteins_interacting						
p1:CCO_C0000831 Proteins_located_in_the_mitochondrion	Removed p1:CCO_C0001677 Removed Proteins_located_in_a_known_place						
Classification Results							
	Fig 4. CCO in Protégé.						
Home Updates Do	wwnload Query Documentation Tools About © search Search						
SPARQL OWL-DL OLS BioPortal							
MAIN MENU							
	ery > sparql						
Home SPARQ							
CDADOL -to							

Name	SWI4_YEAST			transcriptio
Subnamespace	protein			Saccharom ros annuisiae organism
Definition	Regulatory protein S	=		Saccharoniyees enevisiae organism
references	P25302			
synonym	Cell-cycle box factor			
synonym	Protein ART1		-	

mputational

Biology

- Within the EU FP6 project DIAMONDS (LSHG-CT-2004-512143) one of the objectives was to build a data integration platform dedicated to cell cycle biology.
- The Cell Cycle Ontology was chosen as data integration paradigm.
- NorayBio, a software company in Bilbao, Spain (people involved: Iñaki Bilbao, Aitzol Illarramendi, Marta Acilu), developed a software platform that allows querying and visualization of the CCO knowledge, and it provides links to Expression Profiler and ArrayExpress data.

(http://www.semantic-systems-biology)

Reasoning results

- There are a number of relationships in GO (core source of CCO) that might have been better annotated as *part_of* instead of *is_a*.
- The results inspired the GO team to make some amendments to the process part of the GO (e.g regulation of cell cycle).
- Inconsistencies found in the data about the cellular

relationships (*Fig 1*) of cell-cycle components (proteins, genes, cellular locations, phases, ...). The data sources are:

 GO (Cellular Location branch, and branches for 'cell cycle' (GO:0007049), 'cell division' (GO:0051301), 'cell proliferation' (GO:0008283), 'DNA replication' (GO:0006260) and 'chromosome segregation)

nt Act

S NCBI

UniProt

VIB

- RO
- MI (IntAct ontology)
- GOA files
- PPI: IntAct
- NCBI taxonomy
- UniProt
- Cell cycle functional data
- Data obtained with bio-tools (e.g.OrthoMCL)

OBO and OWL-DL formats have been chosen for representing the knowledge. RACER is mainly used for checking the data consistency and for doing classifications.





localizations and protein-protein interactions.

Conclusions and Results

- A fully automated data integration pipeline (nightly launched) was developed (*Fig 2*).
- Concrete problems and results related to the implementation of automatic format mappings (OWL, XML, DOT, GML) between ontologies and inconsistency checking issues have been identified.
- Several exports in commonly used formats have been developed (*Fig 3-7*).
- 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.



 Knowledge will be weighted (e.g. evidence codes) expressing the support media similar to those implemented in GO (experimental, electronically inferred, and so forth).

• Ontolome analysis (e.g. hypothesis generation by

Fig 2. Data integration pipeline.

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

Fig 6. CCO in the BioPortal.



ontology alignments).

 An advanced query system will be developed (DLbased).

• More data to be integrated (upon feedback).

