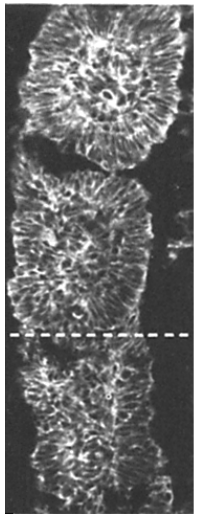
The Cell Behavior Ontology: describing the biological behaviors of real and simulated cells seen as spatially active agents

James P. Sluka*, Sherry G. Clendenon, Maciej Swat and James A. Glazier

Biocomplexity Institute, Indiana University, Bloomington, Indiana, USA

Long Term Vision: Common annotation across multiple data sources (Somitogenesis Example)

Image Data

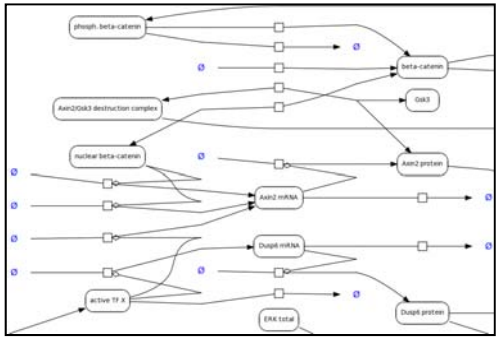


uArray Data

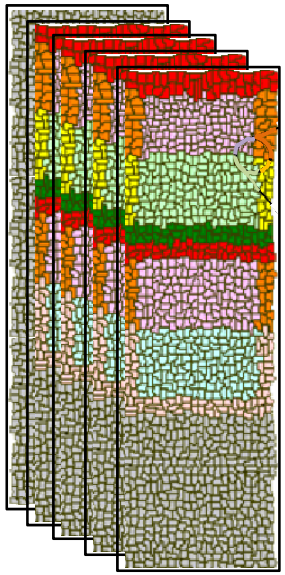
Table 1
Genes differentially expressed between psm and somite I–V identified by microarray and independently confirmed.

PSM[a]	
Similar to KIAA0314	unknown
Unknown clone H3121H10	unknown
Troy	Signal Transduction
Msg1	Transcription
Sin3B	Transcription
Fibronectin[b]	ECM/Cytoskeletal
SOM I–V[c]	
RIKEN cDNA 1010001J06	unknown
p8	unknown function
Mest	unknown function
Clim2	Transcription
Nab1	Transcription
Ezh1	Transcription
Zfp217	Transcription
Ran BP8	Nuclear Entry/Export
Arl4	Nuclear Entry/Export
Gfpt1	ECM/Cytoskeletal
Caveolin1	ECM/Cytoskeletal
Selp	ECM/Cytoskeletal
Collagen 2a[d]	ECM/Cytoskeletal
Collagen 9a[e]	ECM/Cytoskeletal

SBML Model

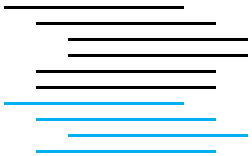


CC3D Model



Semantic Markup

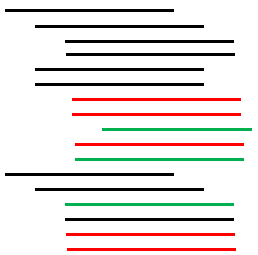
Species: chicken
Process: embryogenesis
Sub process: somitogenesis



Species: chicken
Process: embryogenesis
Sub process: somitogenesis

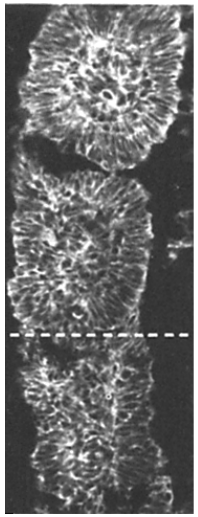
Species: chicken
Process: embryogenesis
Sub process: somitogenesis

Species: chicken
Process: embryogenesis
Sub process: somitogenesis



Long Term Vision: Common annotation across multiple data sources (Somitogenesis Example)

Image Data

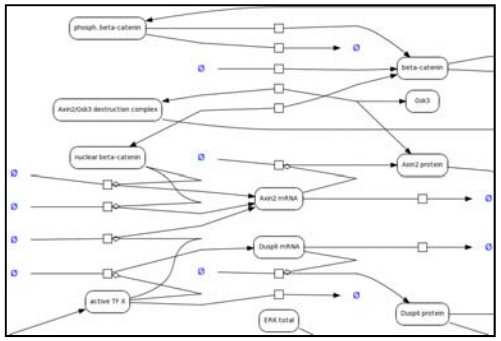


uArray Data

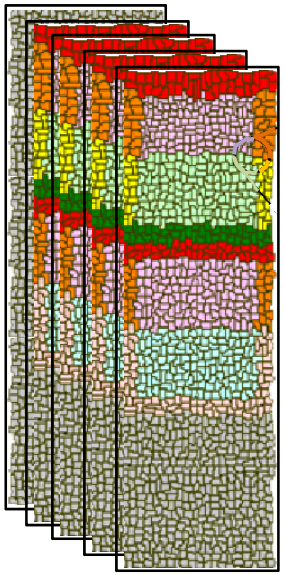
Table 1
Genes differentially expressed between psm and somite I–V identified by microarray and independently confirmed.

PSM[a]	
Similar to KIAA0314	unknown
Unknown clone H3121H10	unknown
Troy	Signal Transduction
Msg1	Transcription
Sin3B	Transcription
Fibronectin[b]	ECM/Cytoskeletal
SOM I–V[c]	
RIKEN cDNA 1010001J06	unknown
p8	unknown function
Mest	unknown function
Clim2	Transcription
Nab1	Transcription
Ezh1	Transcription
Zfp217	Transcription
Ran BP8	Nuclear Entry/Export
Arl4	Nuclear Entry/Export
Gfpt1	ECM/Cytoskeletal
Caveolin1	ECM/Cytoskeletal
Selp	ECM/Cytoskeletal
Collagen 2a[d]	ECM/Cytoskeletal
Collagen 9a[e]	ECM/Cytoskeletal

SBML Model



CC3D Model



Semantic Markup

Species: chicken
Process: embryogenesis
Sub process: somitogenesis

Species: chicken
Process: embryogenesis
Sub process: somitogenesis

Species: chicken
Process: embryogenesis
Sub process: somitogenesis

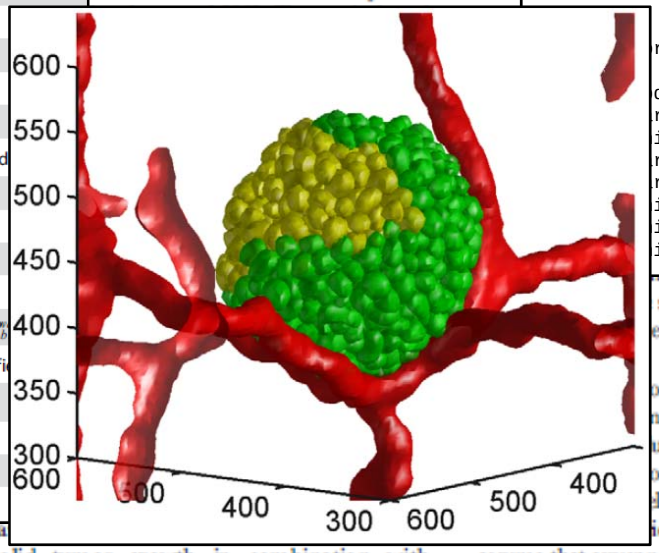
Species: chicken
Process: embryogenesis
Sub process: somitogenesis

Common markup describing the ***biological system***

Typical *ad hoc* biomodel publication modality

Cells	Behaviors
Tumor cells	
Normal	-proliferate
	-consume oxygen
	-change to hypoxic
	-change to necrotic
Hypoxic	-proliferate
	-consume oxygen field
	-change to normal
	-change to necrotic
	-secrete long-diffusing proangiogenic field
Necrotic	-shrink
	-disappear
Endothelial cells	
Vascular	-consume oxygen field
	-supply oxygen field at partial pressure P_b^*
	-secrete short-diffusing chemoattractant fi
	-chemotax up gradients of field $C(\vec{l})$
	-elastically connect to neighboring
	vascular and inactive neovascular cells
	-lose elastic connections, when $l > l_{max}$

```
class MitosisSteppable(MitosisSteppableBase):
    def __init__(self,_simulator,_frequency=1):
        MitosisSteppableBase.__init__(self,_simulator, _frequency)
    def step(self,mcs):
        cells_to_divide=[]
        for cell in self.cellList:
            if cell.type == 1 and cell.volume>64:
                cells_to_divide.append(cell)
            if cell.type== 4 and cell.volume>128:
                cells_to_divide.append(cell)
r cell in cells_to_divide:
  self.divideCellRandomOrientation(cell)
dateAttributes(self):
rentCell=self.mitosisSteppable.parentCell
ildCell=self.mitosisSteppable.childCell
rentCell.targetVolume=parentCell.targetVolume/2
rentCell.lambdaVolume=parentCell.lambdaVolume
ildCell.type=parentCell.type
ildCell.targetVolume=parentCell.targetVolume
ildCell.lambdaVolume=parentCell.lambdaVolume
```



- Paper prose
- Paper figure
- Paper math
- Code
- Results

Often don't agree

The view from 10,000m

Fundamental physical processes

- Movement
  - translation
  - rotation
  - diffusion
- Shape change
- Volume change
  - expansion
  - contraction
- Creation
- Deletion
- Phenotypic change
- Barrier crossing

What is an ontology *really*?

An ontology is a particular *view of reality, not necessarily complete nor comprehensive, that encompasses a defined set of objects, processes and relationships within that reality.*

Controlled Vocabulary	**Hierarchy of Terms (isA)**	**Full Ontology**
Cell Hepatocyte Leukocyte Organ Heart Liver	1. Cell a. Hepatocyte b. Leukocyte 2. Organ a. Heart b. Liver	1. Cell a. Hepatocyte b. Leukocyte 2. Organ a. Heart b. Liver ***partOf***

CBO

- CBO is an OWL Ontology in Protégé 4 (Reasoners!)
- Based on Basic Formal Ontology (BFO)
- CBO currently contains 210 terms and 42 properties
  - Physical objects (cell, basement membrane, ..)
  - Qualities of physical objects (location, volume, …)
  - Processes (cell growth, apoptosis, …)
  - Qualities of processes (equations, parameters, …)

The cell behavior ontology: describing the intrinsic biological behaviors of real and model cells seen as active agents. Sluka JP, Shirinifard A, Swat M, Cosmanescu A, Heiland RW, Glazier JA. *Bioinformatics*. 2014 Aug 15;30(16):2367-74.

http://cbo.biocomplexity.indiana.edu/cbo/

https://**bioportal**.bioontology.org/ontologies/CBO

High-level classes in the CBO and corresponding BFO classes

CBO Class	Subclass of BFO Class
Thing	
CBO_Object	**BFO:snap (continuant)**
BioEntityType	object
▶ ExtracellularMatrix	
ExtraCellularFluid	
Molecule	
Cell	
▶ System	object
▶ Field	object
▶ PhysicalObjectQuality	quality
▶ RateFunction	quality
▶ SystemQuality	quality
▶ Energy	quality
▶ EnergyQuality	quality
▶ FieldQuality	quality
▶ CellState	quality
PhysicalEntityType	quality
CorpuscularEntity	
ContinuousEntity	
DiffuseEntity	
▶ GeometricalEntity	independent_continuant
▶ CellPart	specifically_dependent_continuant
▶ CellMembraneLocalStructure	specifically_dependent_cont.
CBO_Process	**BFO:span (occurrent)**
▶ FundamentalPhysicalProcess	process
▶ CellProcess	process
▶ ExtracellularMatrixProcess	process
FieldProcess	process
StructuralProcess	process
▶ MoleculeProcess	process
▶ TransportProcess	process
▶ TemporalEntity	connected_temporal_region

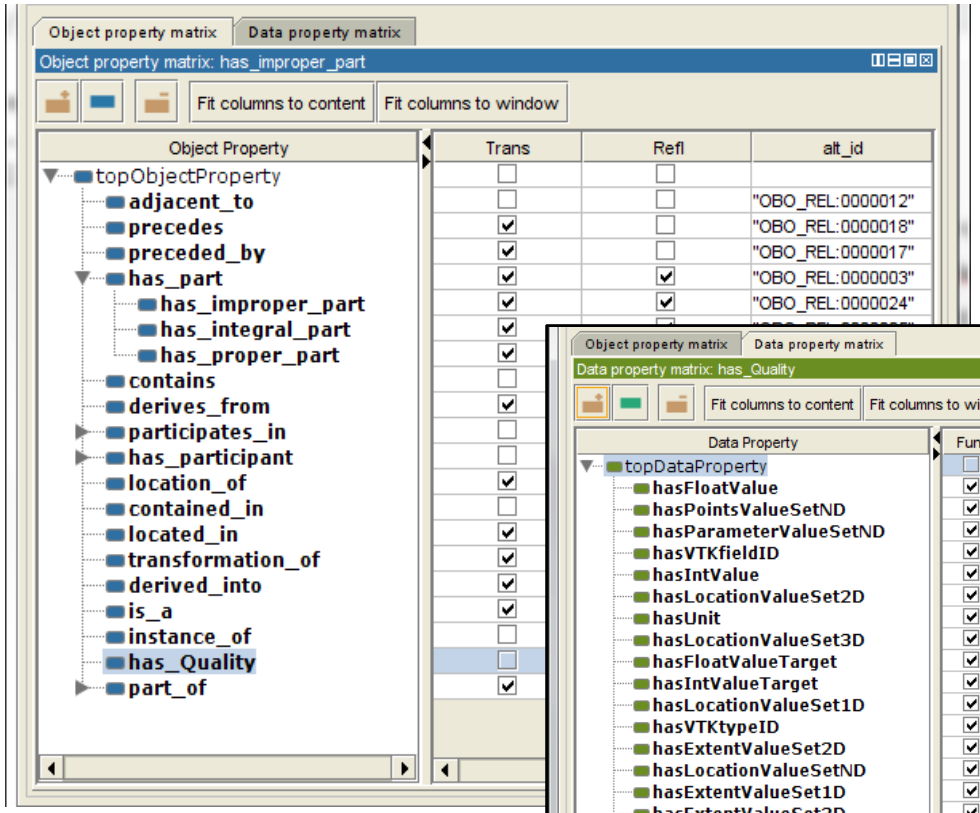
CBO_Object (BFO:snap) describes the physical entities and entity qualities of a biological model.

CBO_Process (BFO:span) describes the processes in which CBO_Object(s) participate.

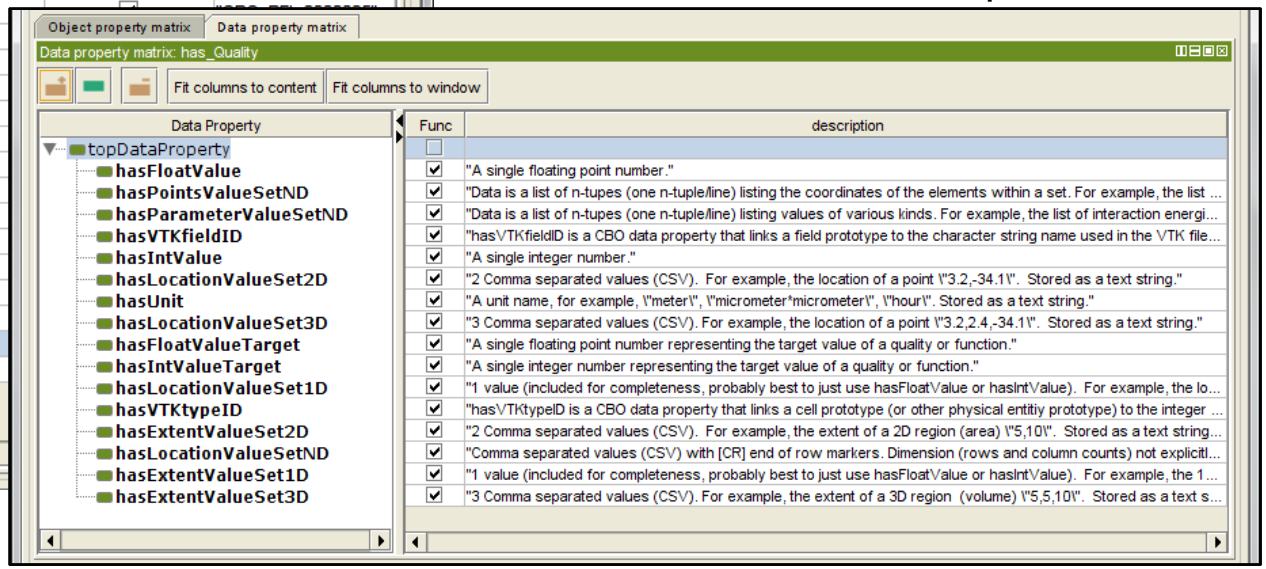
Classes marked with ▶ have additional subclasses that are not shown

CBO Object Properties and Data Properties

Object Properties in CBO.



Data Properties in CBO



Life on a lattice (or not)

Multicell models are generally either:

- Center based (cell is a point + radius)
- Lattice based (cell is a collection of pixels)

Multicell models generally include:

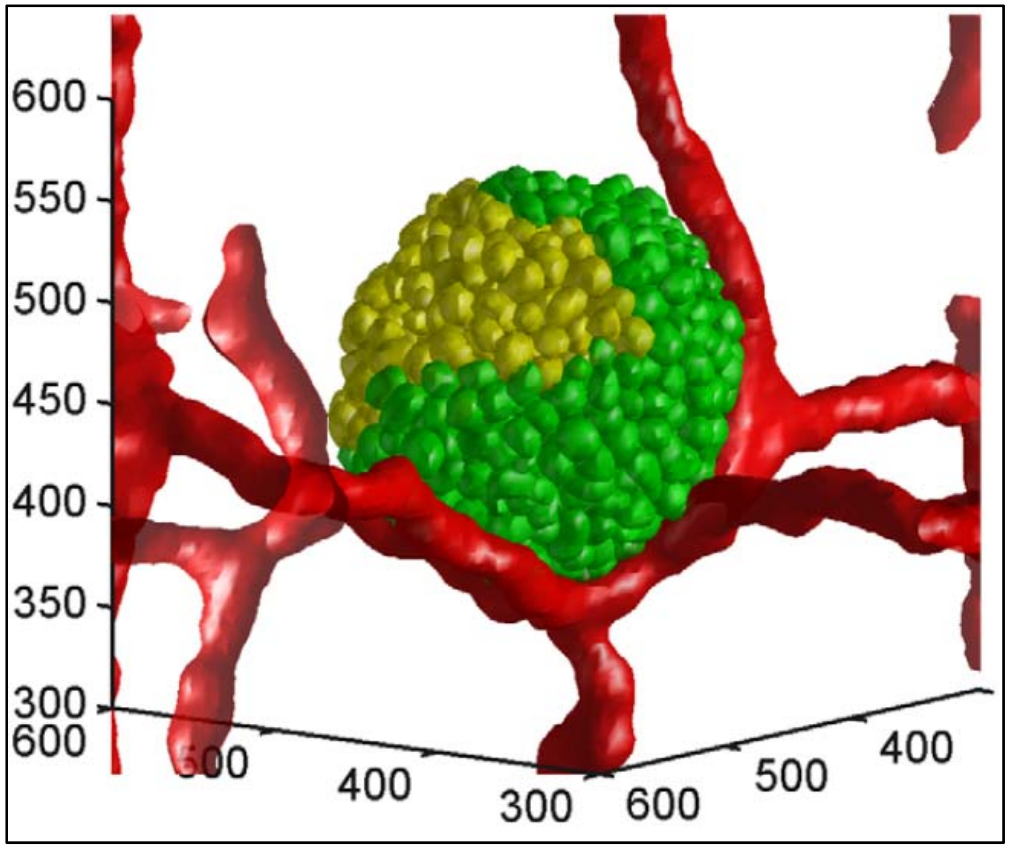
- Cells
  - Location, volume, surface, …
  - Proliferation (creation)
  - Death (deletion)
  - Movement
  - Adhesion
- Non-cell materials
  - E.g. extracellular matrix
  - Chemical (diffusing) fields

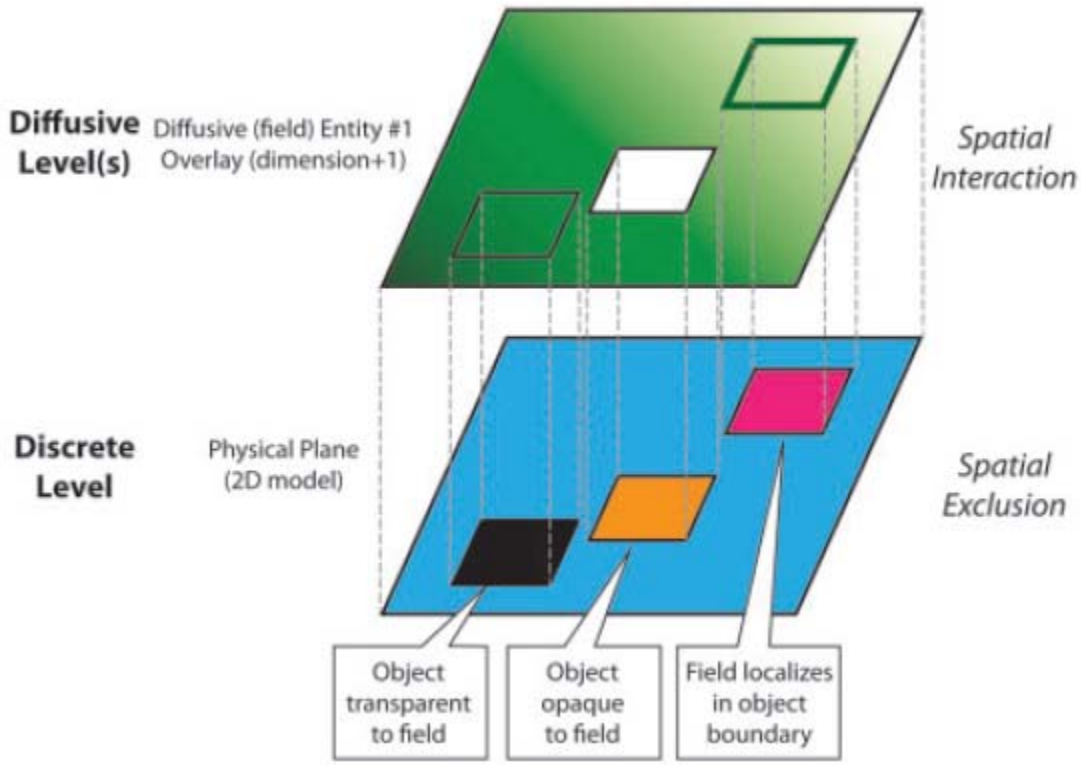
Life doesn't exist on a lattice but relevant biological data is lattice based (e.g., microscope images)

Need to describe the *spatiality* of the system

- 2D and 3D cells and other objects
- Fields
- Time evolution of both

Very large output data sets



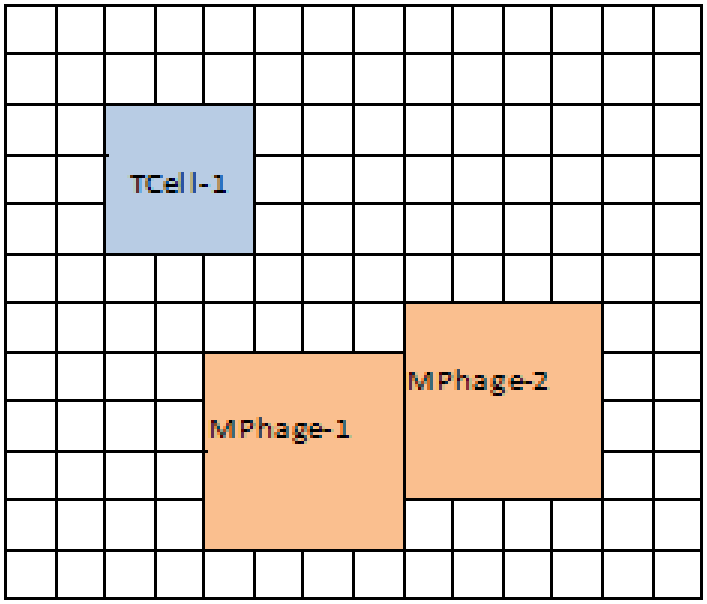


2D Example: The lower panel shows three CBO discrete physical objects (CBO:CorpuscularEntity) embedded in a blue field of medium (CBO:ContinuousEntity with the additional quality of a CBO:PhysicalObjectQuality:Fill). These objects are spatially exclusive. No two objects of these types can occupy the same region of space (pixel) at the same time.

The upper panel shows a CBO: Field, which is not spatially exclusive and by default can overlay any other object. A CBO:Field can interact with the discrete level.

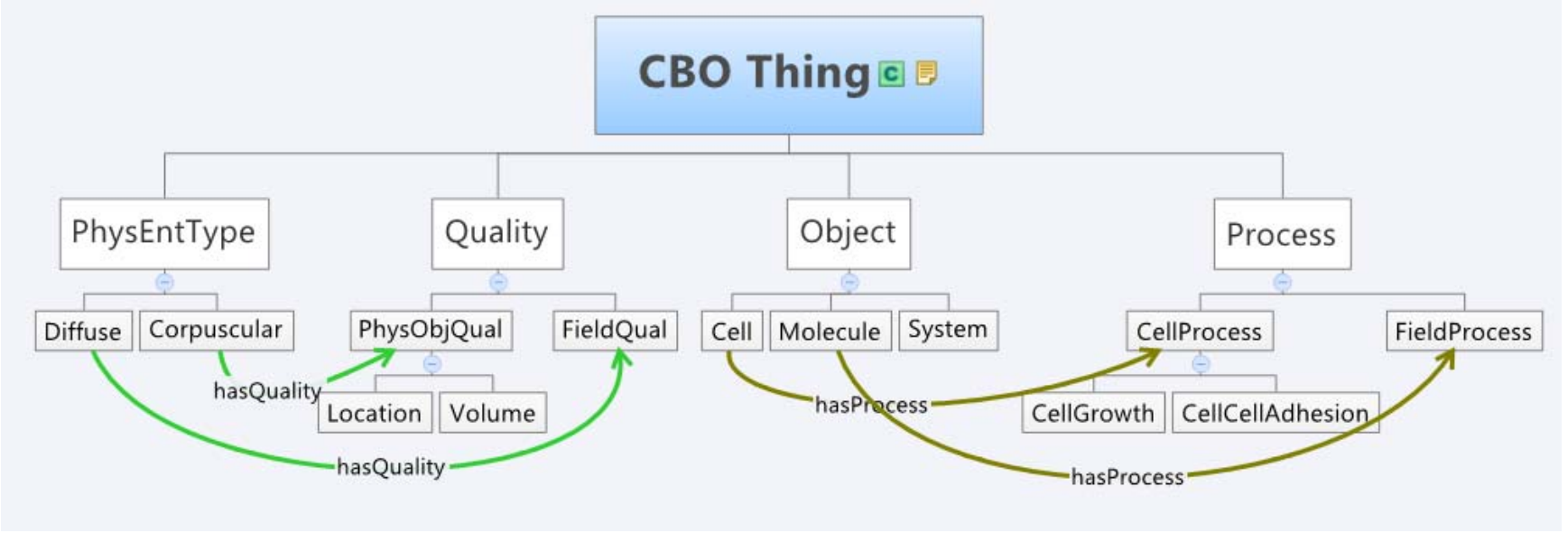
Instantiate a simple model in CBO

Describe a model with two cell types and three cell instances including cell phenotypes, locations, volumes and adhesion.

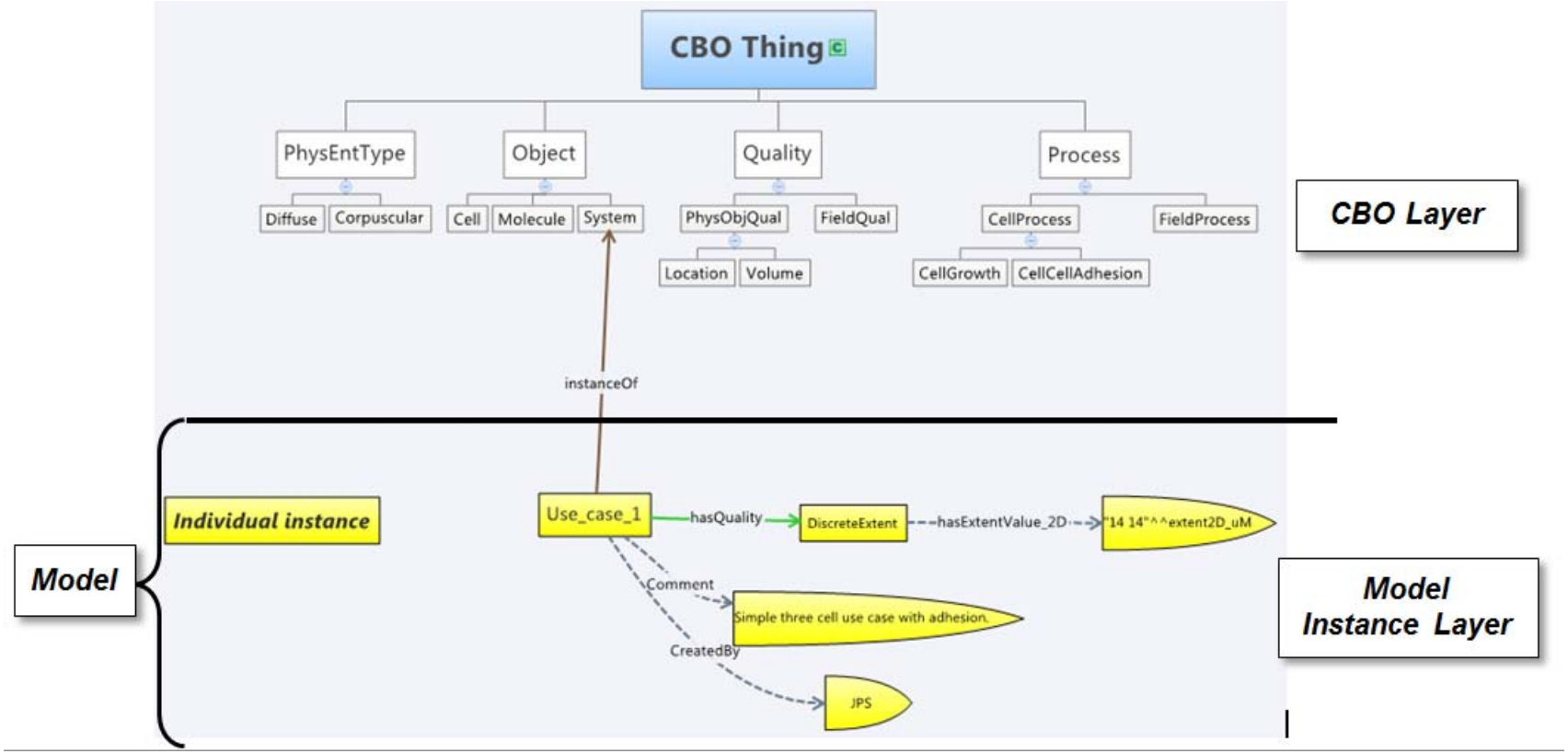


Use Case:

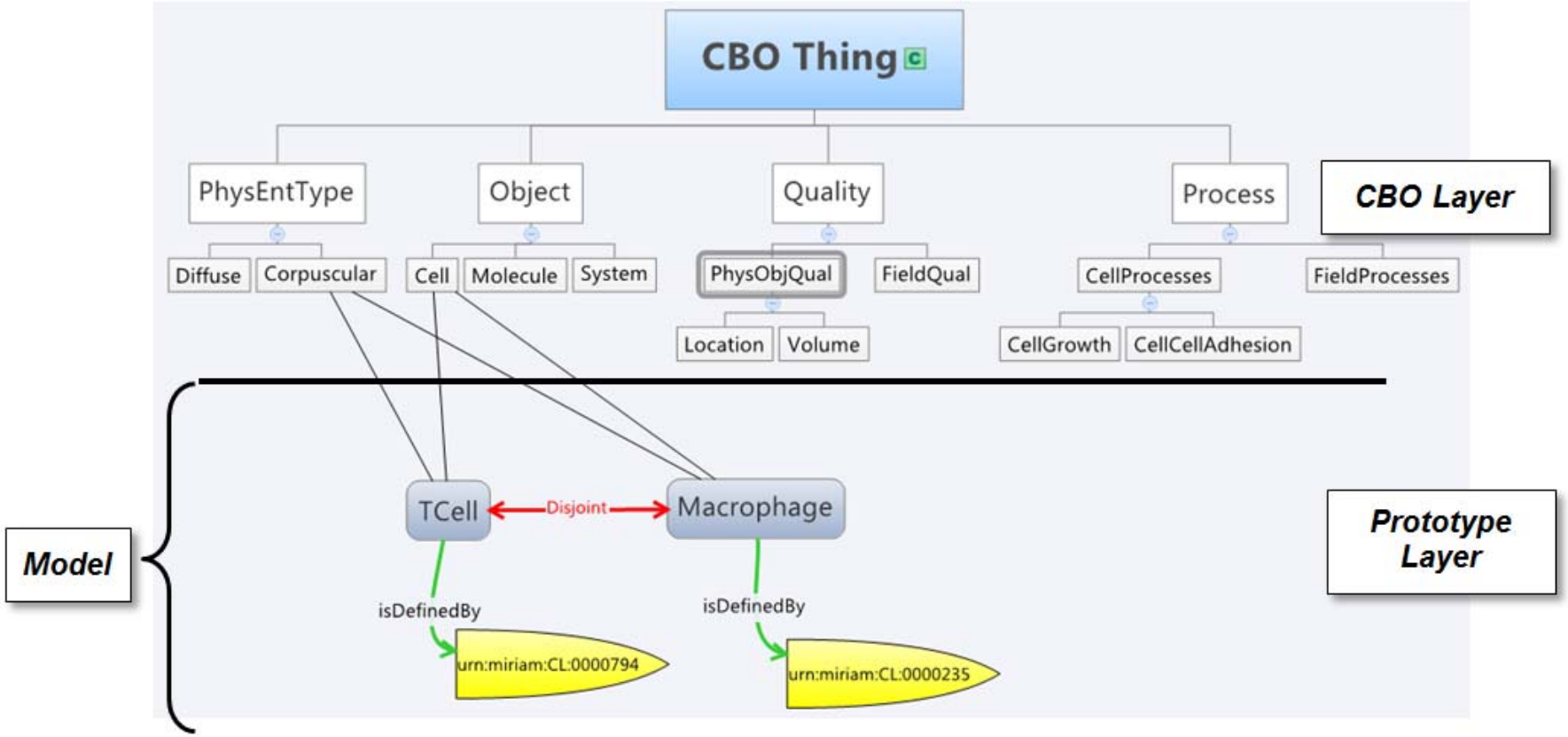
- Simplified view of the CBO
  (17 of 194 classes shown)
- Unlabeled relationships are "isA".
- Indicated relationships may or may not be "hard coded" into the CBO.



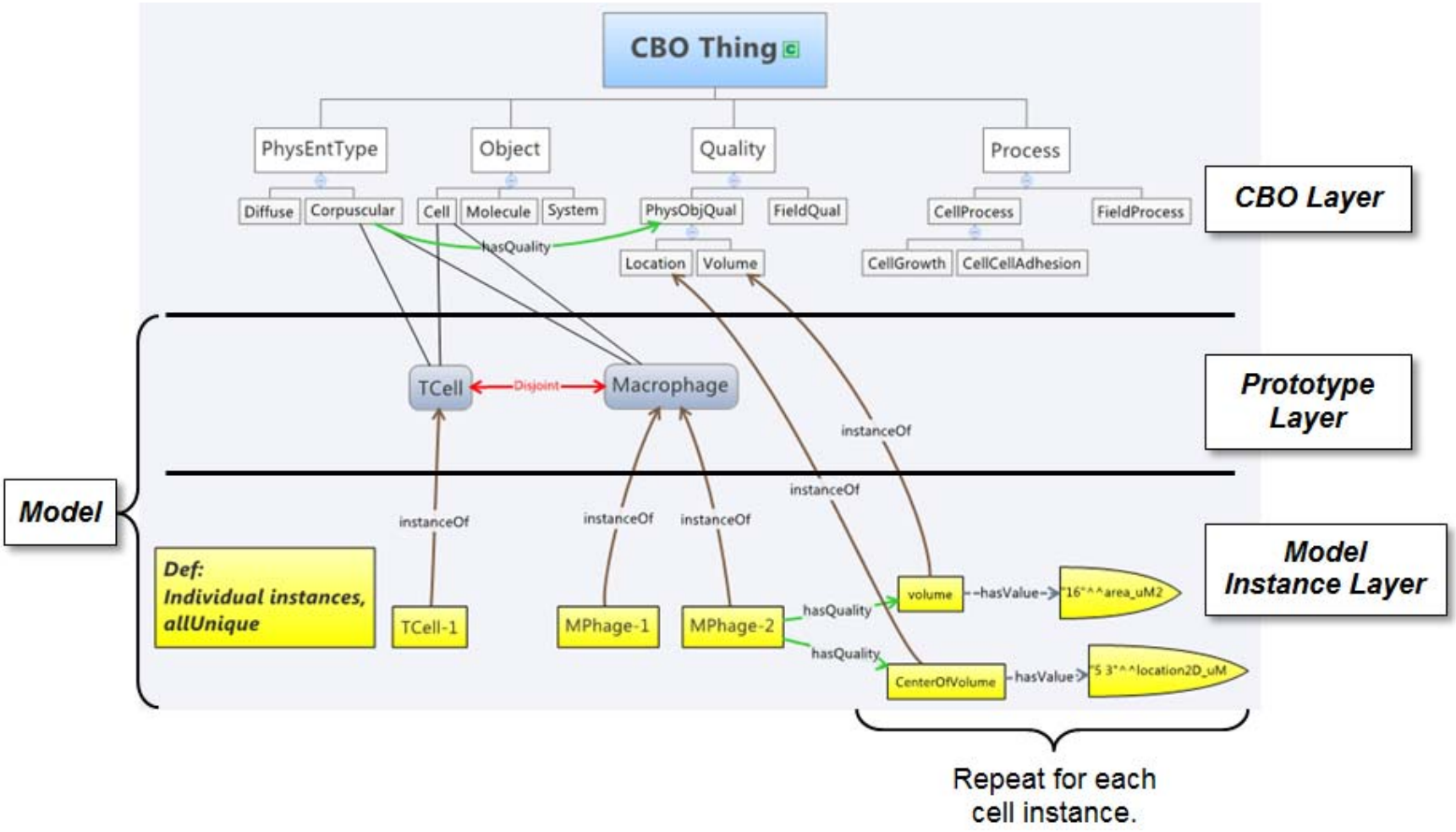
Define the "system"



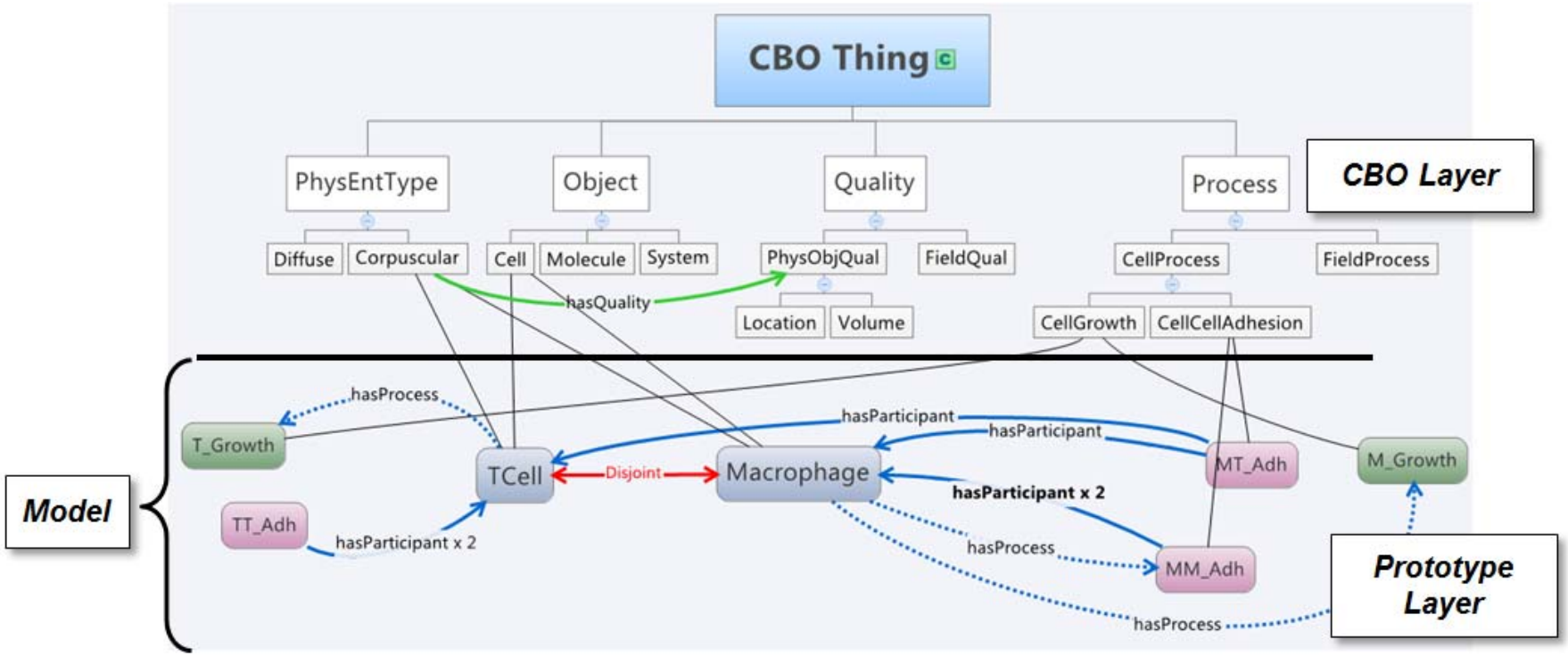
Define prototype cell classes



Define individual cell instances



Processes are defined for prototype classes



CBO annotation of the tumor growth with angiogenesis model

Cells	Behaviors	CBO Term
Tumor cells		isA Cell isA CorpuscularEntity
Normal	-proliferate	participates_in CellGrowth participates_in SymmetricCellDivisionRandom-CleavagePlane
	-consume oxygen field	participates_in MoleculeDeletion
	-change to hypoxic	participates_in PhenotypicChange
	-change to necrotic	participates_in PhenotypicChange
Hypoxic	-proliferate	participates_in CellGrowth participates_in SymmetricCellDivisionRandom-CleavagePlane
	-consume oxygen field	participates_in MoleculeDeletion
	-change to normal	participates_in PhenotypicChange
	-change to necrotic	participates_in PhenotypicChange
	-secrete long-diffusing proangiogenic field V	participates_in MoleculeDeletion
Necrotic	-shrink	participates_in CellVolumeChange
	-disappear	participates_in Necrosis
Endothelial cells		isA Cell isA CorpuscularEntity
Vascular	-consume oxygen field	participates_in MoleculeDeletion
	-supply oxygen field at partial pressure P	participates_in MoleculeCreation
	-secrete short-diffusing chemoattractant field C	participates_in MoleculeCreation
	-chemotax up gradients of field C	participates_in Chemotaxis
	-elastically connect to neighboring vascular and inactive neovascular cells	participates_in CellCellAdhesion has_part CellPart:CellMembranePart:TightJunction
Fields		
Oxygen		isA Molecule, isA DiffuseEntity, isA MolecularField participates_in Diffusion, participates_in Creation participates_in Deletion

Describe a molecular field

Class: VEGF1
  Annotations:
    dc:description "Endothelial-derived short diffusing VEGF isoform.",
    rdfs:isDefinedBy "http://purl.bioontology.org/ontology/MSH/D042461",
    rdfs:isDefinedBy "http://purl.obolibrary.org/obo/PR_000017284"
  SubClassOf:
    MolecularField,
    Molecule,
    derives_from some VEGF1_creation,
    derives_from some VEGF1_deletion,
    DiffuseEntity

vascular endothelial growth factor A

Class: VEGF1_creation
  Annotations:
    dc:description "Process that describes the creation of the VEGF1 field.",
    rdfs:isDefinedBy "http://purl.obolibrary.org/obo/GO_0010573",
  SubClassOf:
    MoleculeCreation,
    has_participant some (NeoVascular or Vascular)

vascular endothelial growth factor production

Class: VEGF1_deletion
  Annotations:
    rdfs:isDefinedBy "http://purl.obolibrary.org/obo/GO_0010573",
    dc:description "Process that describes the deletion (consumption, degradation, clearance)
                of the VEGF1 field."
  SubClassOf:
    has_participant some ((not (NeoVascular)) and (not (Vascular))),
    MoleculeDeletion

Storing Lattice Data

For efficiency, data storage can be separate from the CBO-OWL meta-model.

Need a file format that:

- Can store arbitrary lattice data (many "color" channels, multidimensional, time)
- 2D and 3D with Time
- Would also like to be able to store shape primitives (line, plane, polygon)
- Store center models
- Has existing tools for visualization and manipulation
- Has a database option for storage

One solution has been implemented and one in the works:

- Visualization Tool Kit (VTK)
- Open Microscopy Environment (OME, OMERO & OMEtiff)

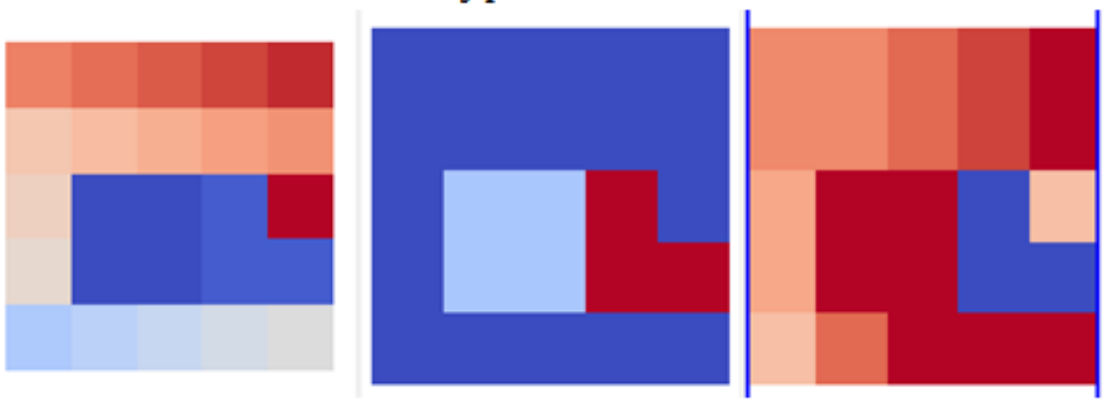
http://www.vtk.org/
http://www.openmicroscopy.org

Storing Lattice Data in VTK

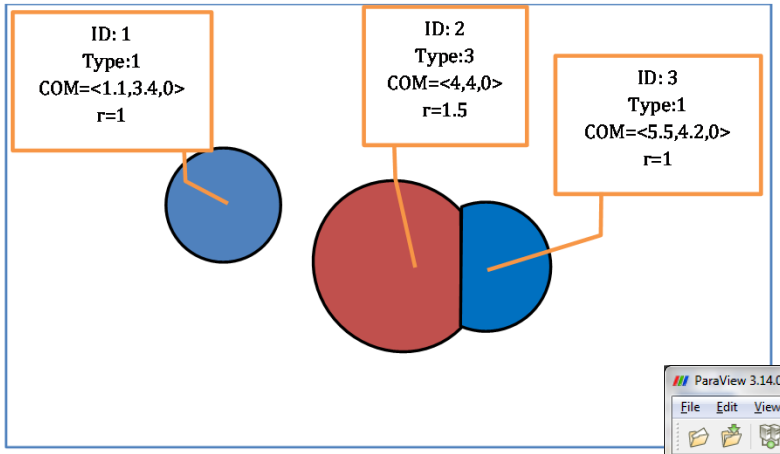
Cell ID Field "CellId"	Cell Type Field "CellType"	Field "FieldA"
10 11 12 13 14 15 1 1 2 2 16 1 1 2 27 17 18 19 20 21 22 23 24 25 26 *NOTE:* <0,0> is bottom left	0 0 0 0 0 0 4 4 12 12 0 4 4 12 0 0 0 0 0 0 0 0 0 0 0	1 1.3 1.5 1.5 1.5 1.1 1.5 1.5 0.1 0.1 1.1 1.5 1.5 0.1 1 1.2 1.2 1.3 1.4 1.5 1.2 1.2 1.3 1.4 1.5

If we rendered these three fields in a tool such as ParaView (http://www.paraview.org/) top left corner in ParaView) we get graphical representations of the data in each field:

CellId CellType FieldA

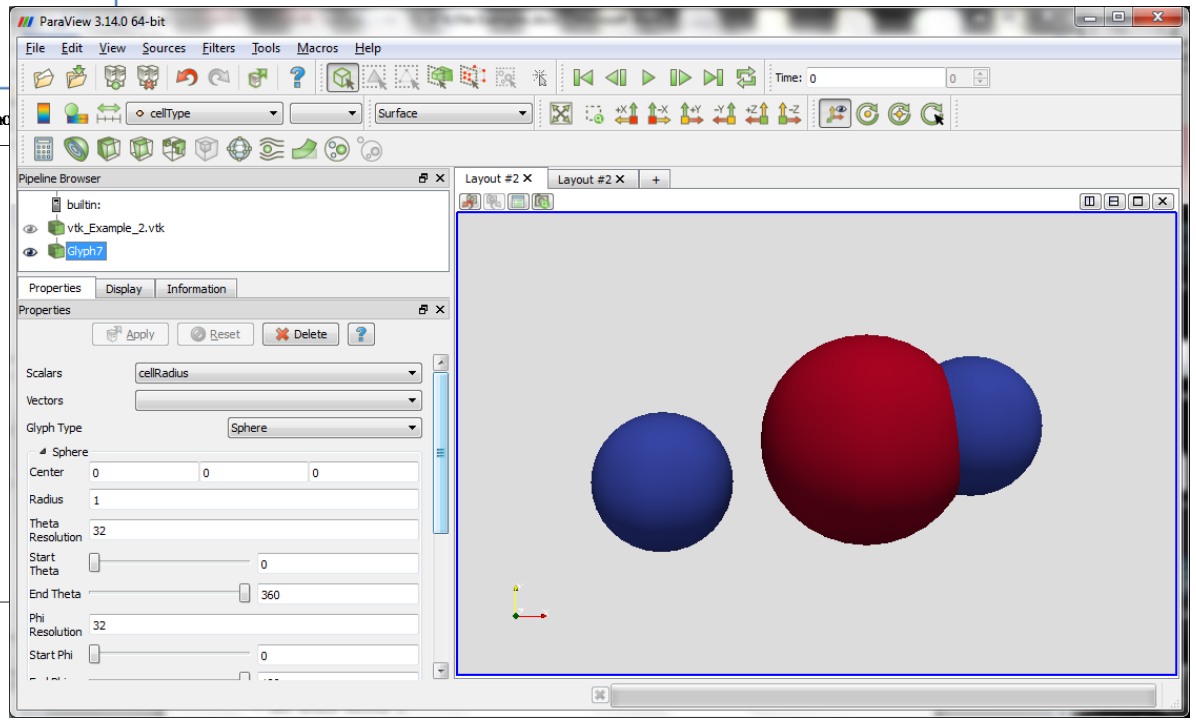


VTK can also handle “center” data

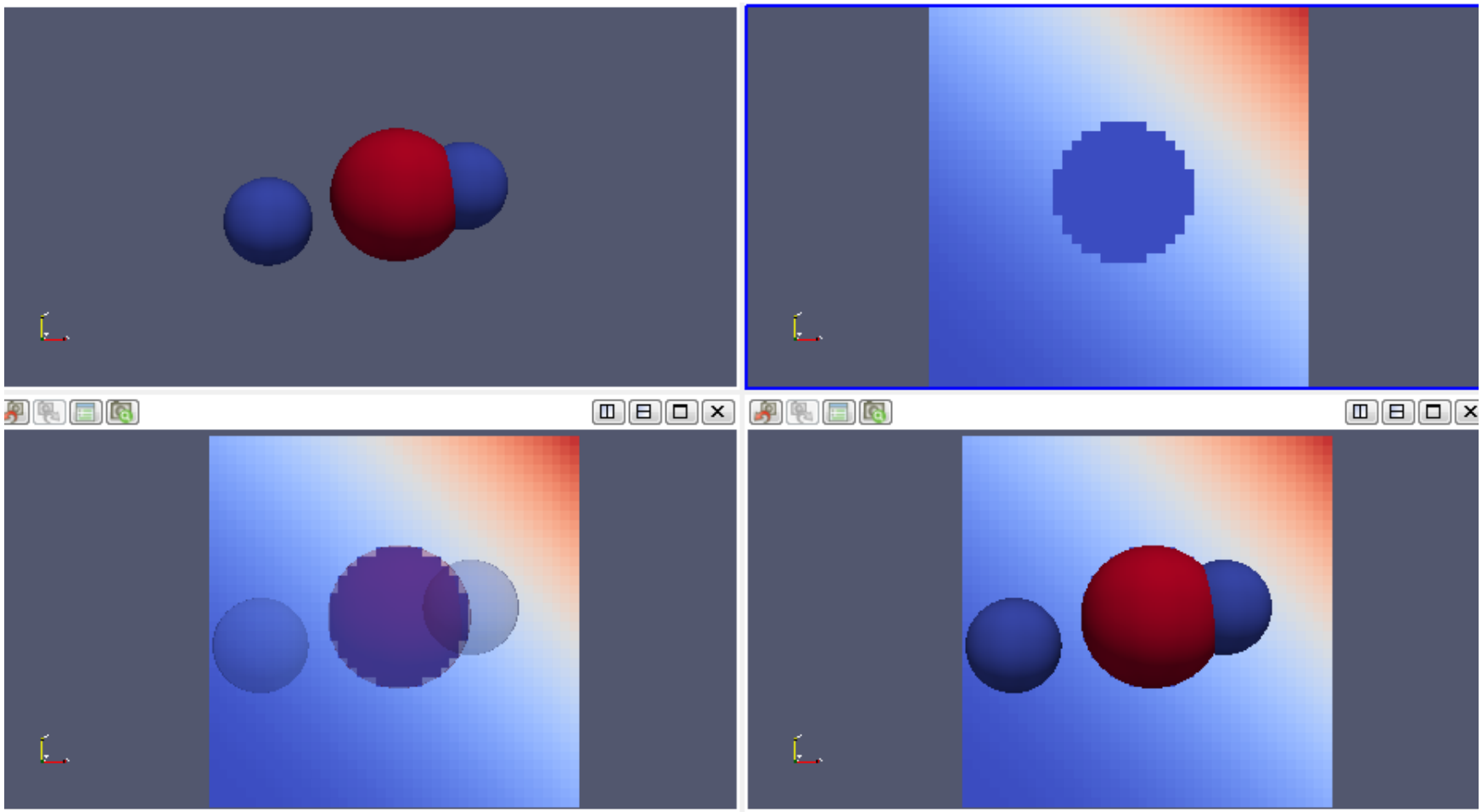


A VTK file for this model uses the "DATASET POLYDATA" file type and is sho

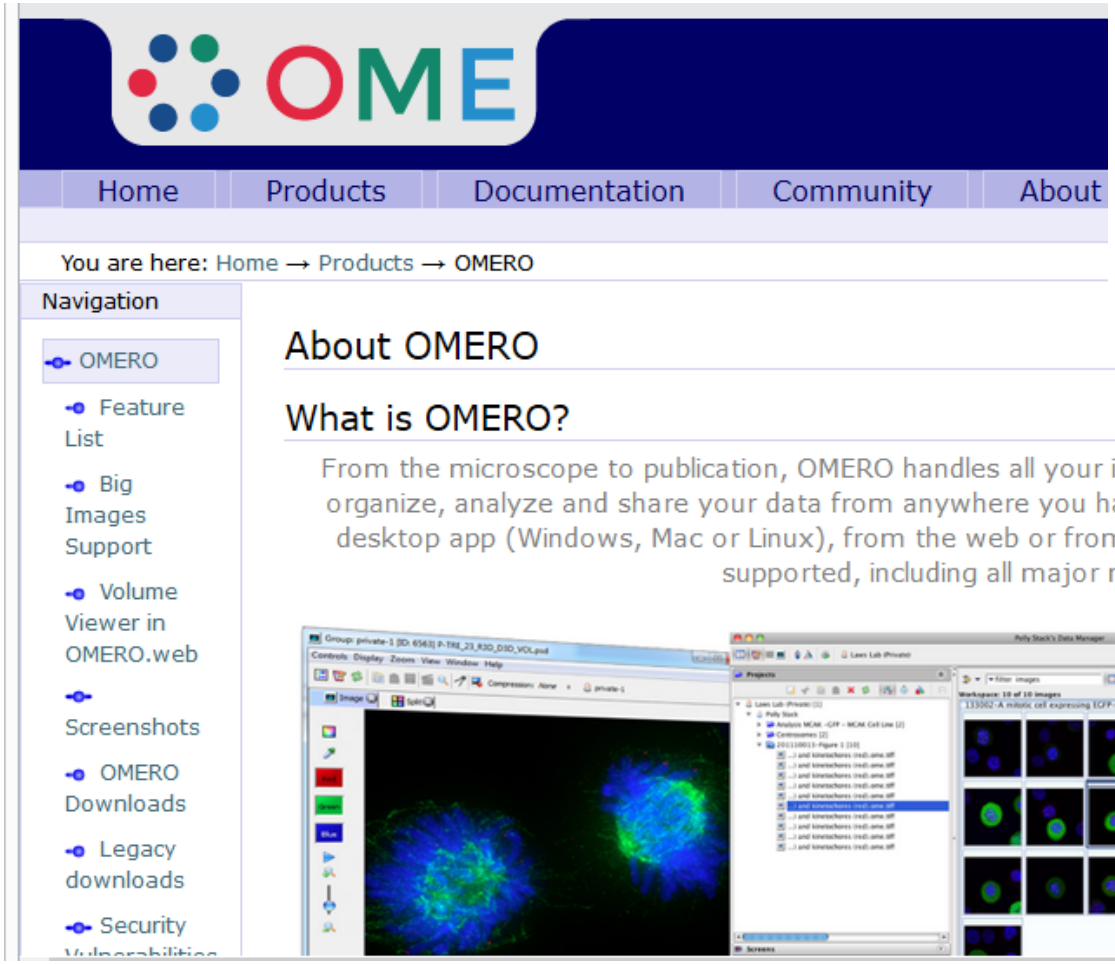
```
# vtk DataFile Version 2.0
Example 1: Lattice model with cell types and a field
ASCII
DATASET POLYDATA
POINTS 3 float
1.1 3.4 0.0
4.0 4.0 0.0
5.5 4.2 0.0
VERTICES 3 6
1 0
1 1
1 2
POINT_DATA 3
SCALARS cellType int
LOOKUP_TABLE default
1 2 1
SCALARS cellID int
LOOKUP_TABLE default
1 2 3
SCALARS cellRadius float
LOOKUP_TABLE default
1.0 1.5 1.0
```



VTK can also handle “center” data and pixel based field data



Open Microscopy Environment's OMERO



- Store:
  - Very large "image" datasets
  - Meta-data
- Rendering
- Analysis
- Access control and publication

CBO OWL Meta-models can be queried with SPARQL

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
SELECT ?s ?o
  WHERE {   ?s rdfs:isDefinedBy ?o .    }
  ORDER BY ?s
```

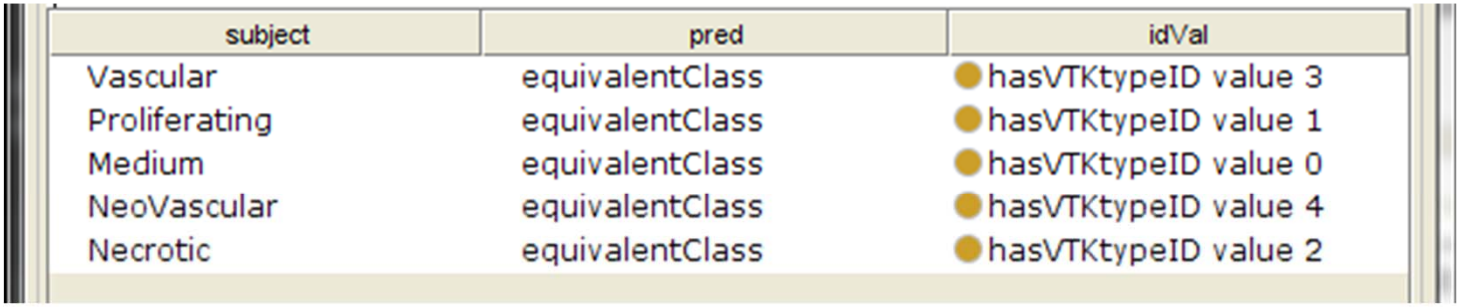
s	
AdhesionMolecule	"urn:miriam:obo.fma:67214"@
Apoptosis	"urn:miriam:obo.go:0006915"@
Apoptosis	"PATO:0000638"@
AutophagicDeath	"urn:miriam:obo.go:0048102"@
BasalPart	"urn:miriam:obo.go:0045178"@
BasalPart	"urn:miriam:obo.fma:72558"@
BasementMembrane	"urn:miriam:obo.go:0005605"@
BodyShape	"urn:miriam:opb:'bounded Volume Shape'"@
BrushBorder	"urn:miriam:obo.go:0005903"@
BrushBorder	"urn:miriam:obo.fma:70977"@

Extract all of the cells in the meta-model.

```
PREFIX owl:  <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX CBO:  <http://cbo.biocomplexity.indiana.edu/svn/cbo/trunk/CBO_1_0.owl#>

SELECT  ?subject ?pred ?idVal
WHERE {
    ?subject  rdfs:subClassOf      CBO:Cell .
    ?subject  ?pred                ?idVal   .
    ?subject  owl:equivalentClass  ?idVal   .  }
```

Returns the prototype cells defined in the metamodel along with their VTK ids:



subject	pred	idVal
Vascular	equivalentClass	hasVTKtypeID value 3
Proliferating	equivalentClass	hasVTKtypeID value 1
Medium	equivalentClass	hasVTKtypeID value 0
NeoVascular	equivalentClass	hasVTKtypeID value 4
Necrotic	equivalentClass	hasVTKtypeID value 2

“Simple” Python programs can extract information from the CBO-OWL + simulation snapshot VTK files

```
Name Spaces found in the OWL file:
             http://www.semanticweb.org/ontologies/2013/1/7/untitled-ontology-286#
CBO_1_0      http://cbo.biocomplexity.indiana.edu/svn/cbo/trunk/CBO_1_0.owl#
CBO_1_02     http://cbo.biocomplexity.indiana.edu/svn/cbo/trunk/CBO_1_0#
dc           http://purl.org/dc/elements/1.1/
oboInOwl     http://www.geneontology.org/formats/oboInOwl#
owl          http://www.w3.org/2002/07/owl#
rdf          http://www.w3.org/1999/02/22-rdf-syntax-ns#
rdfs         http://www.w3.org/2000/01/rdf-schema#
xml          http://www.w3.org/XML/1998/namespace
xsd          http://www.w3.org/2001/XMLSchema#

Searching for system parameters:
Looking up unit names:
http://purl.obolibrary.org/obo/UO_0000017
  http://www.w3.org/2000/01/rdf-schema#label = micrometer
http://purl.obolibrary.org/obo/UO_0000031
  http://www.w3.org/2000/01/rdf-schema#label = minute

In Main():
Cell Type ID numbers
      ID    Name
      0     Medium
      1     Proliferating
      2     Necrotic
      3     Vascular
      4     NeoVascular
Field Names:
      Glucose
      VEGF1
      VEGF2
System Parameters:
      SystemPixelDistanceScale = 4.0
      SystemPixelDistanceScale unit = micrometer
      VTK_FileTimeStep = 1.0
      VTK_FileTimeStep unit = minute

Finished parsing OWL file: ... p/VascularTumor/AngioTumor.o
```

```
Processing VTK files, found  10  VTK files
Processing VTK File = Step_0000.vtk   MCS= 0    Processing fields...

VTK 'fields' in the VTK files:
      CellId
      CellType
      CellVolume
      ClusterId
      Glucose
      VEGF1
      VEGF2

Total Cell Volumes by cell type (micrometer^3):
  MCS  minute(s)       Medium    Proliferating         Necrotic         Vascular      NeoVascular
    0       0.0     12645120            41600                0            96576            16704
  600     600.0     12537024           154048                0            91776            17152
 1200    1200.0     12494912           187648              768            92352            24320
 1800    1800.0     12448512           229888                0            92800            28800
 2400    2400.0     12374784           297792                0            93120            34304
 3000    3000.0     12288256           372160                0            92864            46720

Average field pixel values:
  MCS  minute(s)         Glucose           VEGF1           VEGF2
    0       0.0         4.79804         0.00008         0.00003
  600     600.0         8.81348         0.00717         0.00651
 1200    1200.0         9.51673         0.00753         0.00710
 1800    1800.0         9.26159         0.00803         0.00795
 2400    2400.0         8.48257         0.00838         0.00991
 3000    3000.0         7.49264         0.00908         0.00988

Cell Counts:
  MCS  minute(s)   Proliferating         Necrotic         Vascular      NeoVascular
    0       0.0              64                0               24                4
  600     600.0              64                0               24                4
 1200    1200.0              63                1               24                5
 1800    1800.0              87                0               24                5
 2400    2400.0             102                0               24                6
 3000    3000.0             111                0               24               10
```

Thank You!

Support:

- US EPA
- NIH/NIGMS
- Indiana University
- NSF