Populous:
A tool for Populating OWL Ontologies from Templates

Simon Jupp
Bio Health Informatics Group,
School of Computer Science,
University of Manchester, UK

10.12.2010 SWAT4LS 2010, Berlin.

Ontology development barriers

- The underlying representation
- The art of modeling
- Community engagement
- The tool support

MANCHESTER 1824

The University of Manchester

Templates approach

- Shield the underlying technology
- Collect repetitive information consistently
- Put constraints on the input data
- Abstraction from any complex modeling

MANCHESTER 1824

The University of Manchester

Capturing the differentia

Knowledge

All Eukarytoic Cells are either nucleated or anucleate, some cells are multinucleate

Ontologically

'Eukarytoic Cells' *has_nucleation* some 'Nucleation'
'Nucleation' *subClassOf* {mononucleate , binucleate , polynucleate , anucleate}

Differentia

'Eukarytoic Cells' *has_nucleation some* 'Nucleation'
'Nucleation' *subClassOf* {mononucleate , binucleate , polynucleate , anucleate}

Real Examples

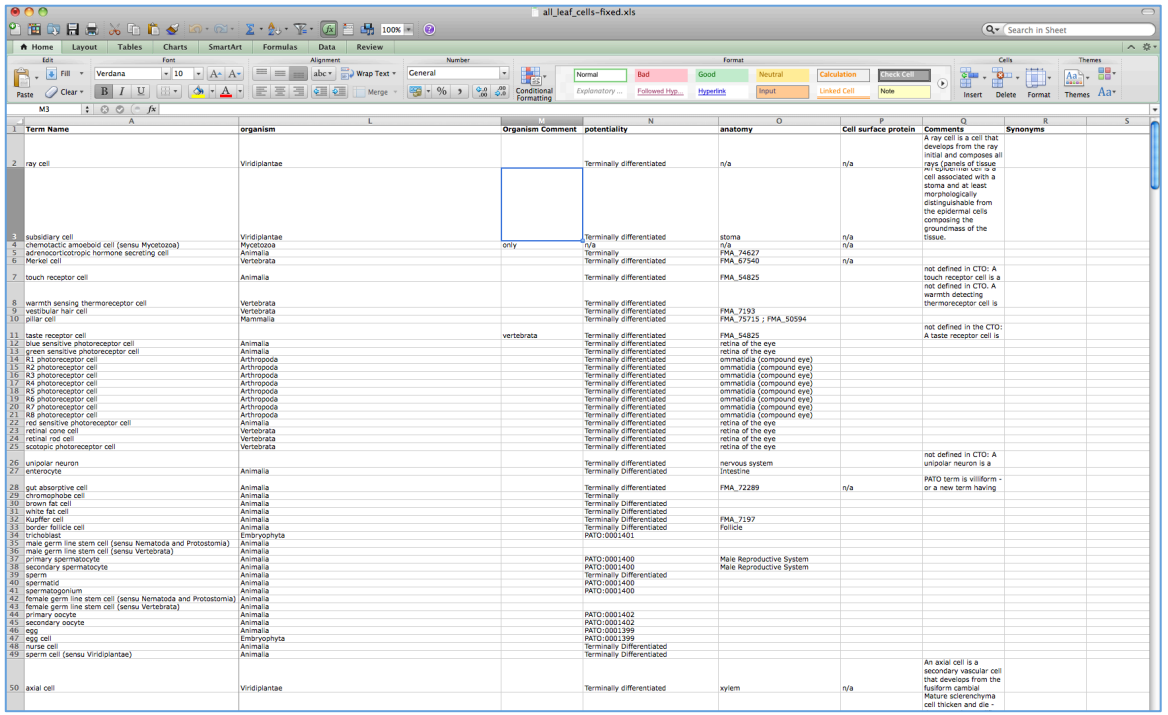
'Eukarytoic Cells'	'Nucleation'
Mononculear phagocyte	mononcuelate
Flight Muscle cell	multinucleate
Red Blood cell	anucleate

MANCHESTER 1824

The University of Manchester

BHIG

Spreadsheets

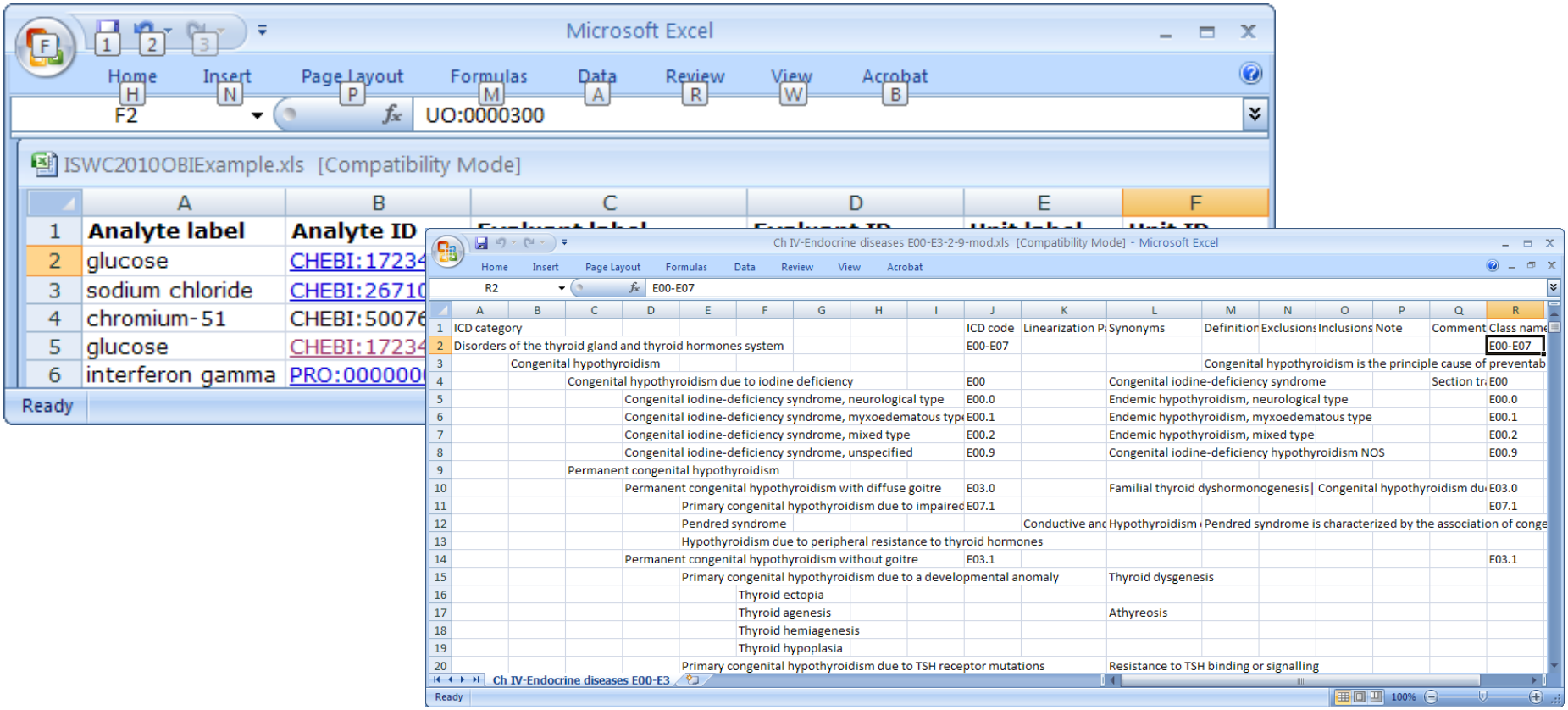


	Term Name	organism	Organism Comment	potentiality	anatomy	Cell surface protein	Comments	Synonyms
2	ray cell	Viridiplantae		Terminally differentiated	n/a	n/a	A ray cell is a cell that develops from the ray initial and composes all rays (panels of tissue	
3	subsidiary cell	Viridiplantae		Terminally differentiated	stoma	n/a	cell associated with a stoma and at least morphologically distinguishable from the epidermal cells composing the groundmass of the tissue.	
4	chemotactic amoeboid cell (sensu Mycetozoa)	Mycetozoa	only	n/a	n/a	n/a		
5	adrenocorticotropic hormone secreting cell	Animalia		Terminally	FMA_74627			
6	Merkel cell	Vertebrata		Terminally differentiated	FMA_67540	n/a		
7	touch receptor cell	Animalia		Terminally differentiated	FMA_54825		not defined in CTO: A touch receptor cell is a	
8	warmth sensing thermoreceptor cell	Vertebrata		Terminally differentiated			not defined in CTO. A warmth detecting thermoreceptor cell is	
9	vestibular hair cell	Vertebrata		Terminally differentiated	FMA_7193			
10	pillar cell	Mammalia		Terminally differentiated	FMA_75715 ; FMA_50594			
11	taste receptor cell		vertebrata	Terminally differentiated	FMA_54825		not defined in the CTO: A taste receptor cell is	
12	blue sensitive photoreceptor cell	Animalia		Terminally differentiated	retina of the eye			
13	green sensitive photoreceptor cell	Animalia		Terminally differentiated	retina of the eye			
14	R1 photoreceptor cell	Arthropoda		Terminally differentiated	ommatidia (compound eye)			
15	R2 photoreceptor cell	Arthropoda		Terminally differentiated	ommatidia (compound eye)			
16	R3 photoreceptor cell	Arthropoda		Terminally differentiated	ommatidia (compound eye)			
17	R4 photoreceptor cell	Arthropoda		Terminally differentiated	ommatidia (compound eye)			
18	R5 photoreceptor cell	Arthropoda		Terminally differentiated	ommatidia (compound eye)			
19	R6 photoreceptor cell	Arthropoda		Terminally differentiated	ommatidia (compound eye)			
20	R7 photoreceptor cell	Arthropoda		Terminally differentiated	ommatidia (compound eye)			
21	R8 photoreceptor cell	Arthropoda		Terminally differentiated	ommatidia (compound eye)			
22	red sensitive photoreceptor cell	Animalia		Terminally differentiated	retina of the eye			
23	retinal cone cell	Vertebrata		Terminally differentiated	retina of the eye			
24	retinal rod cell	Vertebrata		Terminally differentiated	retina of the eye			
25	scotopic photoreceptor cell	Vertebrata		Terminally differentiated	retina of the eye			
26	unipolar neuron			Terminally differentiated	nervous system		not defined in CTO: A unipolar neuron is a	
27	enterocyte	Animalia		Terminally Differentiated	intestine			
28	gut absorptive cell	Animalia		Terminally differentiated	FMA_72289	n/a	PATO term is villiform - or a new term having	
29	chromophobe cell	Animalia		Terminally				
30	brown fat cell	Animalia		Terminally Differentiated				
31	white fat cell	Animalia		Terminally Differentiated				
32	Kupffer cell	Animalia		Terminally Differentiated	FMA_7197			
33	border follicle cell	Animalia		Terminally Differentiated	Follicle			
34	trichoblast	Embryophyta		PATO:0001401				
35	male germ line stem cell (sensu Nematoda and Protostomia)	Animalia						
36	male germ line stem cell (sensu Vertebrata)	Animalia						
37	primary spermatocyte	Animalia		PATO:0001400	Male Reproductive System			
38	secondary spermatocyte	Animalia		PATO:0001400	Male Reproductive System			
39	sperm	Animalia		Terminally Differentiated				
40	spermatid	Animalia		PATO:0001400				
41	spermatogonium	Animalia		PATO:0001400				
42	female germ line stem cell (sensu Nematoda and Protostomia)	Animalia						
43	female germ line stem cell (sensu Vertebrata)	Animalia						
44	primary oocyte	Animalia		PATO:0001402				
45	secondary oocyte	Animalia		PATO:0001402				
46	egg	Animalia		PATO:0001399				
47	egg cell	Embryophyta		PATO:0001399				
48	nurse cell	Animalia		Terminally Differentiated				
49	sperm cell (sensu Viridiplantae)	Animalia		Terminally Differentiated				
50	axial cell	Viridiplantae		Terminally differentiated	xylem	n/a	An axial cell is a secondary vascular cell that develops from the fusiform cambial Mature sclerenchyma cell thicken and die -	

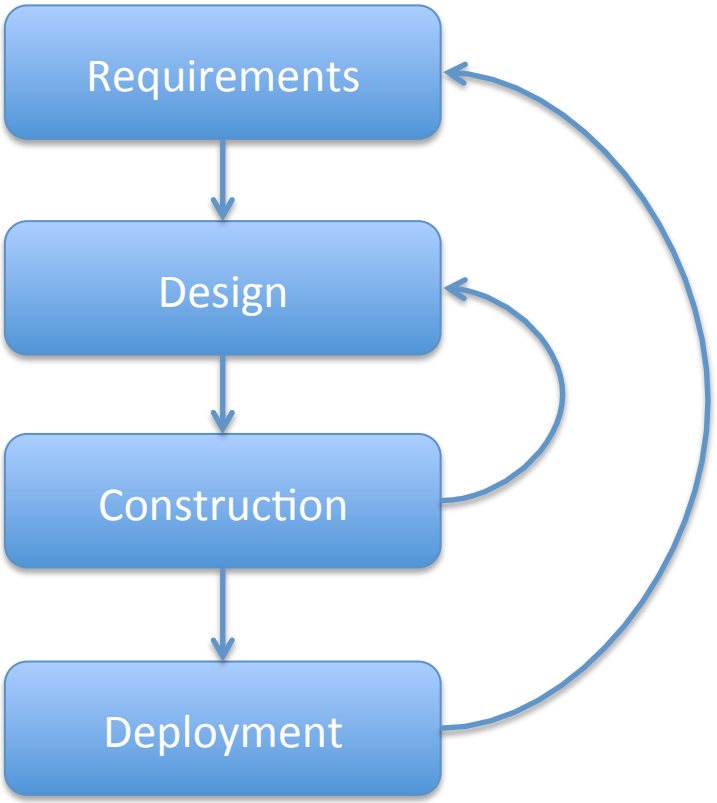
- A popular tool for data management
- Good at collecting regular data
- Users from many domains

Related work

- Cell type ontology normalisation
- OBI and Quick Term Templates
- ICD-11



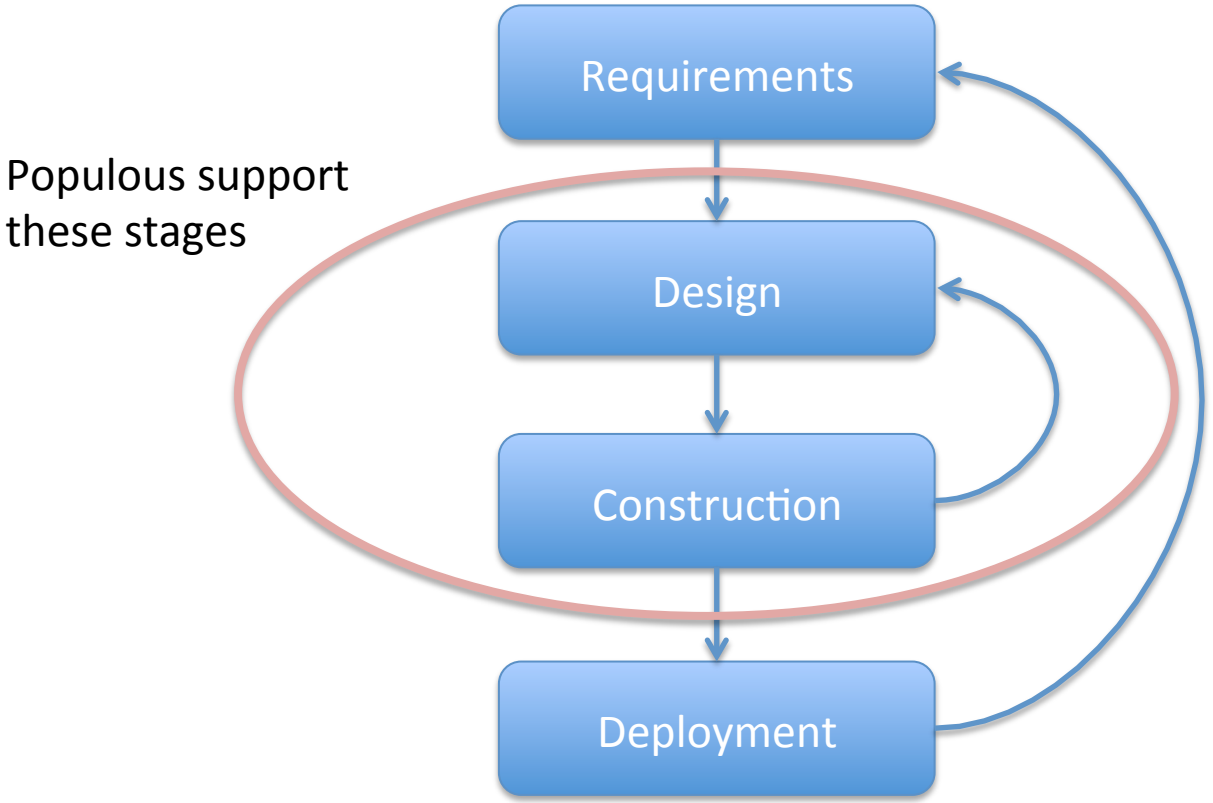
Typical Stages in Ontology development



MANCHESTER 1824

The University of Manchester

Typical Stages in Ontology development



The University of Manchester

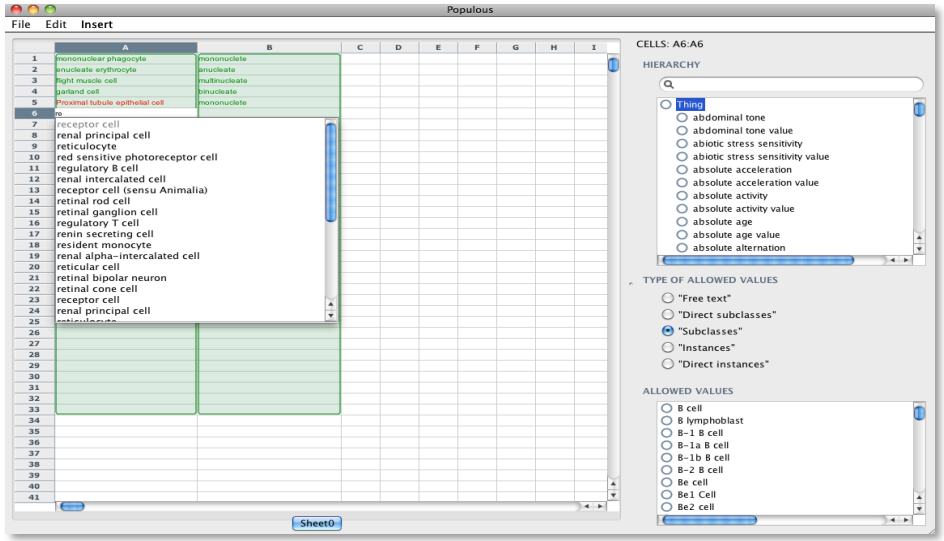
Previous work

- RDF based tools
  - Excel2RDF, Convert2RDF, RDF123
- Protégé 4.0 plugins
  - Excel Importer, Matrix Plugin
- Protégé 3.4
  - Mapping Master (Protégé 3)

MANCHESTER 1824

The University of Manchester

Populous



http://www.e-lico.eu/populous

- Generic tool for populating ontology templates
- Spreadsheet style interface
- Supports validation at the point of data entry
- Expressive Pattern language for OWL Ontology generation

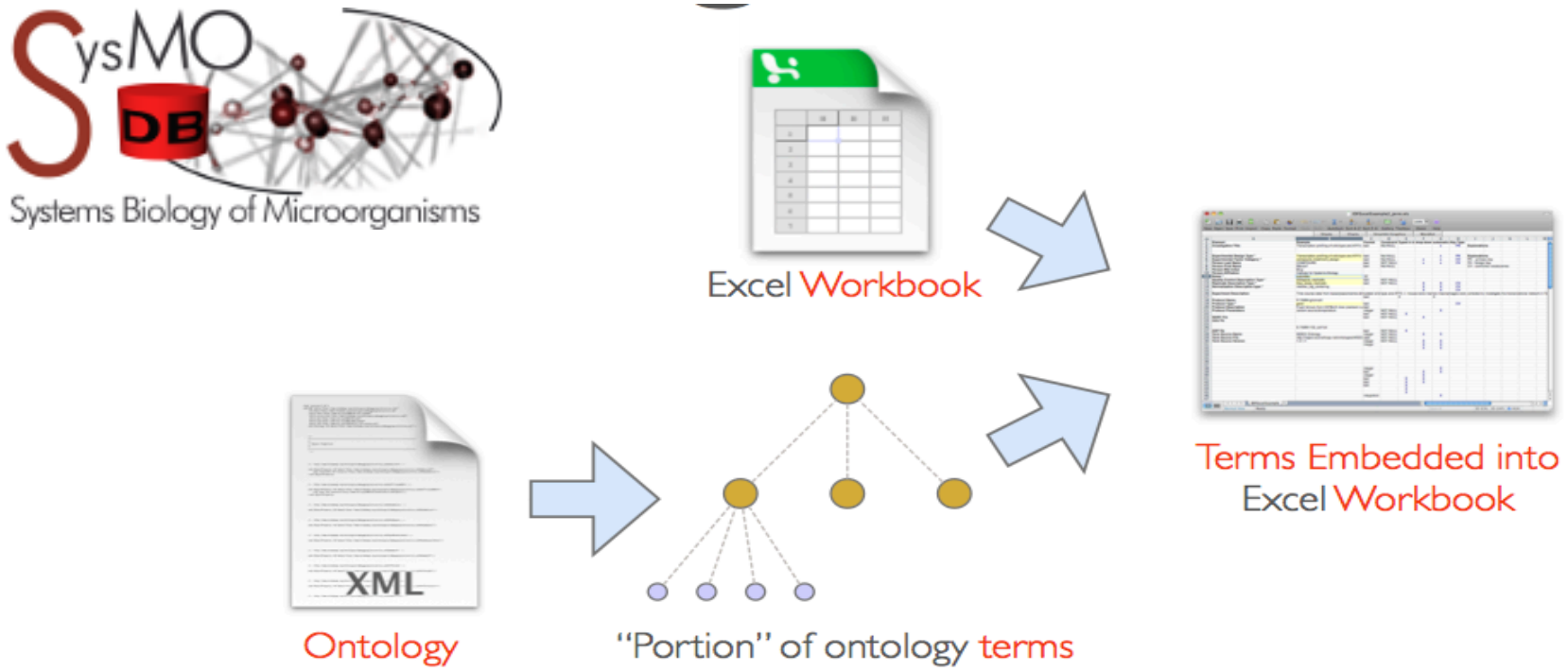
MANCHESTER 1824

The University of Manchester

Built on top of RightField

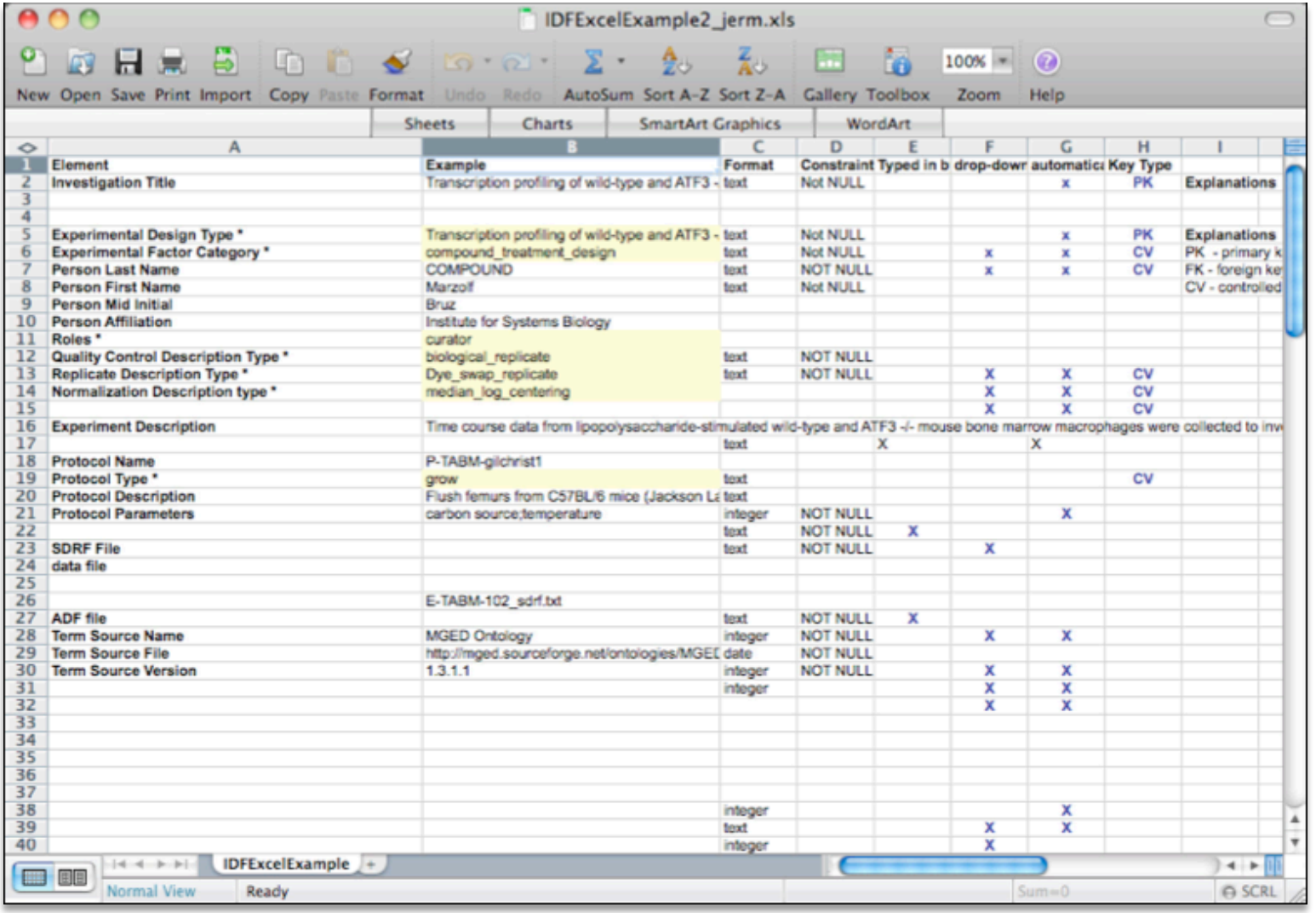
http://www.sysmo-db.org/rightfield

- Semantic Annotation by Stealth



The University of Manchester

Export to Excel

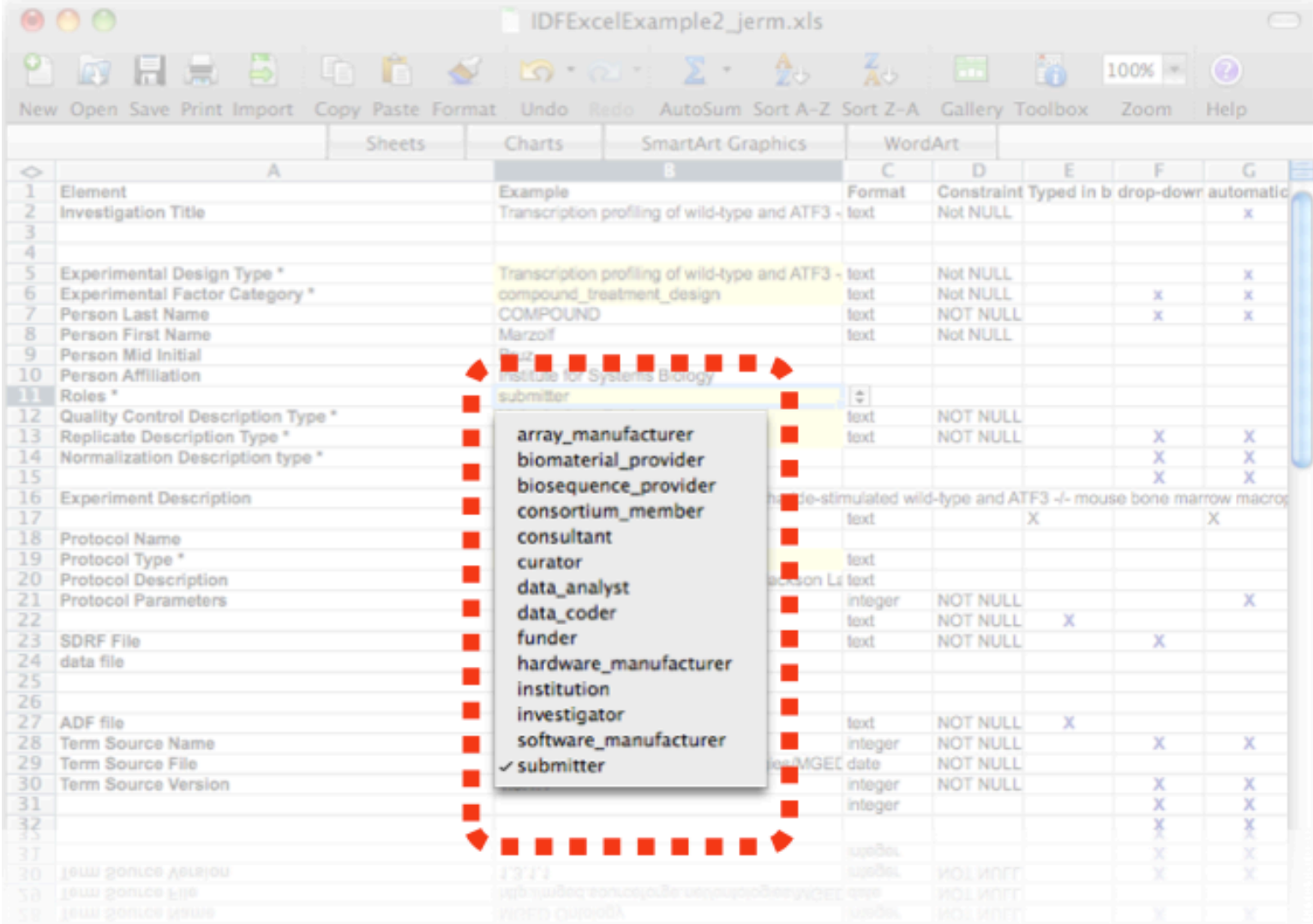


	A	B	C	D	E	F	G	H	I
1	Element	Example	Format	Constraint	Typed in b	drop-dowr	automatic	Key Type	
2	Investigation Title	Transcription profiling of wild-type and ATF3 -	text	Not NULL			x	PK	Explanations
3									
4									
5	Experimental Design Type *	Transcription profiling of wild-type and ATF3 -	text	Not NULL			x	PK	Explanations
6	Experimental Factor Category *	compound_treatment_design	text	Not NULL		x	x	CV	PK - primary k
7	Person Last Name	COMPOUND	text	NOT NULL		x	x	CV	FK - foreign ke
8	Person First Name	Marzolf	text	Not NULL					CV - controlled
9	Person Mid Initial	Bruz							
10	Person Affiliation	Institute for Systems Biology							
11	Roles *	curator							
12	Quality Control Description Type *	biological_replicate	text	NOT NULL					
13	Replicate Description Type *	Dye_swap_replicate	text	NOT NULL		X	X	CV	
14	Normalization Description type *	median_log_centering				X	X	CV	
15						X	X	CV	
16	Experiment Description	Time course data from lipopolysaccharide-stimulated wild-type and ATF3 -/- mouse bone marrow macrophages were collected to inv							
17			text		X		X		
18	Protocol Name	P-TABM-gilchrist1							
19	Protocol Type *	grow	text					CV	
20	Protocol Description	Flush femurs from C57BL/6 mice (Jackson La	text						
21	Protocol Parameters	carbon source;temperature	integer	NOT NULL			X		
22			text	NOT NULL	X				
23	SDRF File		text	NOT NULL		X			
24	data file								
25									
26		E-TABM-102_sdrf.txt							
27	ADF file		text	NOT NULL	X				
28	Term Source Name	MGED Ontology	integer	NOT NULL		X	X		
29	Term Source File	http://mged.sourceforge.net/ontologies/MGEI	date	NOT NULL					
30	Term Source Version	1.3.1.1	integer	NOT NULL		X	X		
31			integer			X	X		
32						X	X		
33									
34									
35									
36									
37									
38			integer				X		
39			text			X	X		
40			integer			X			

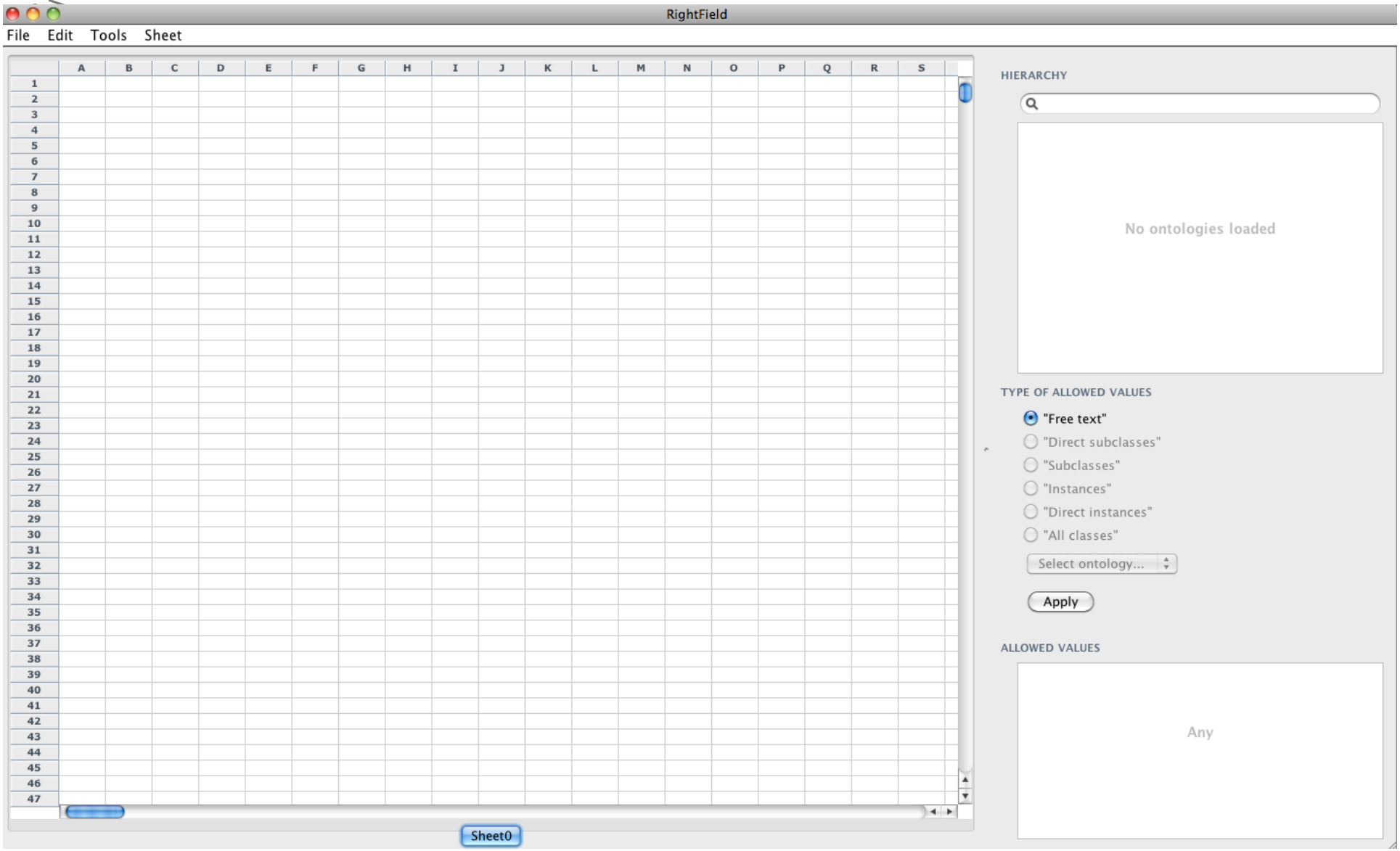
MANCHESTER 1824

The University of Manchester

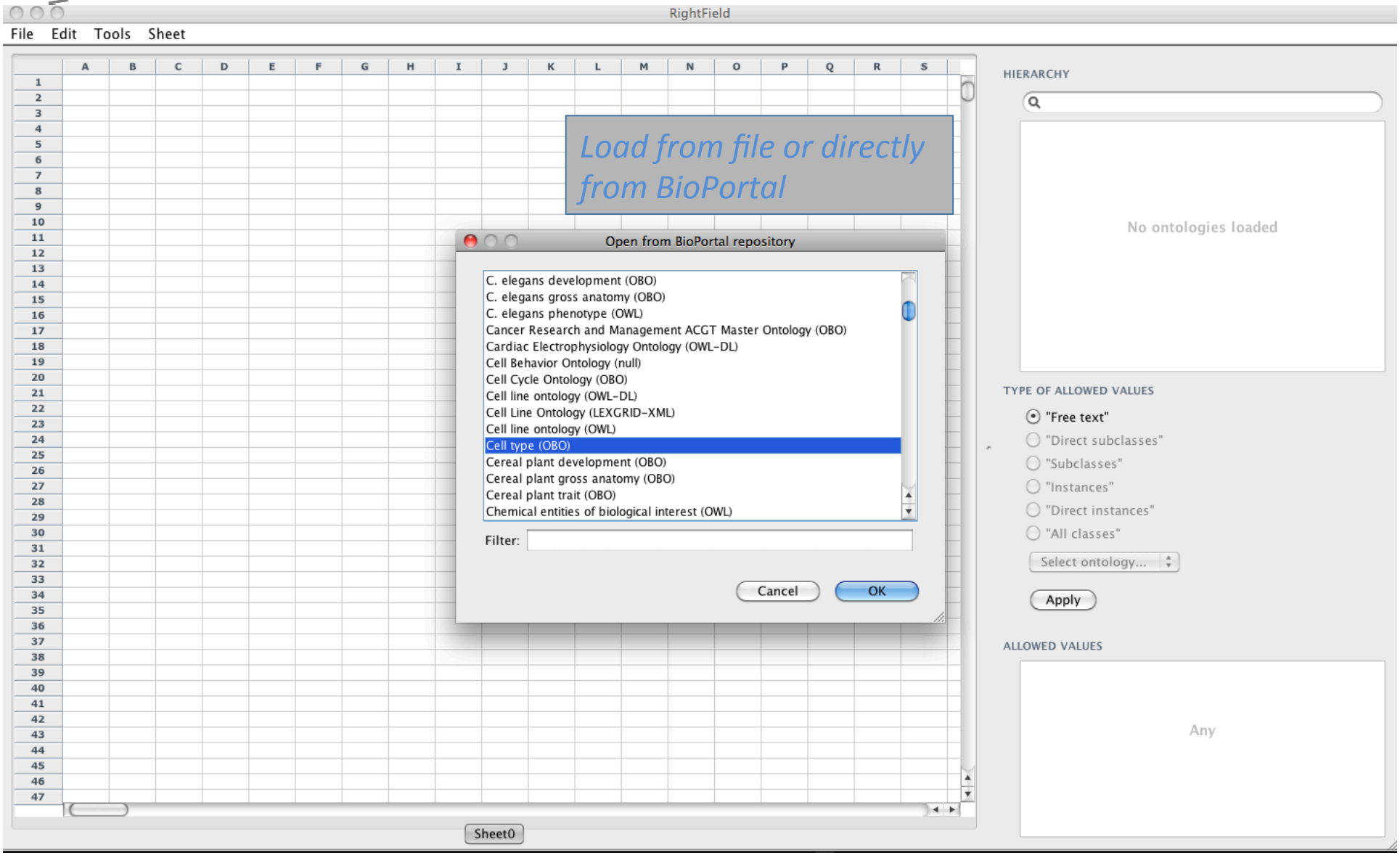
Excel Validations



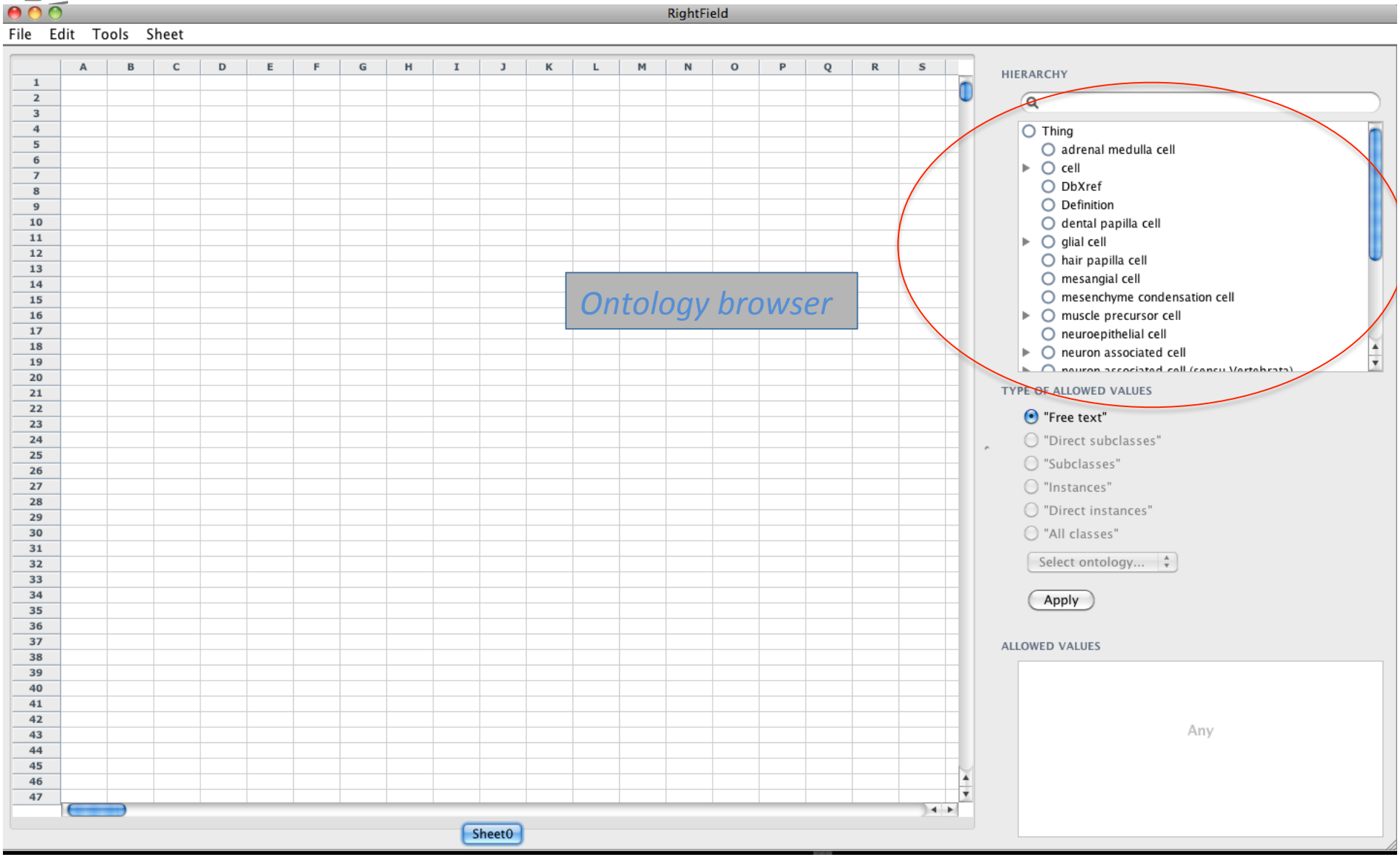
Creating Templates



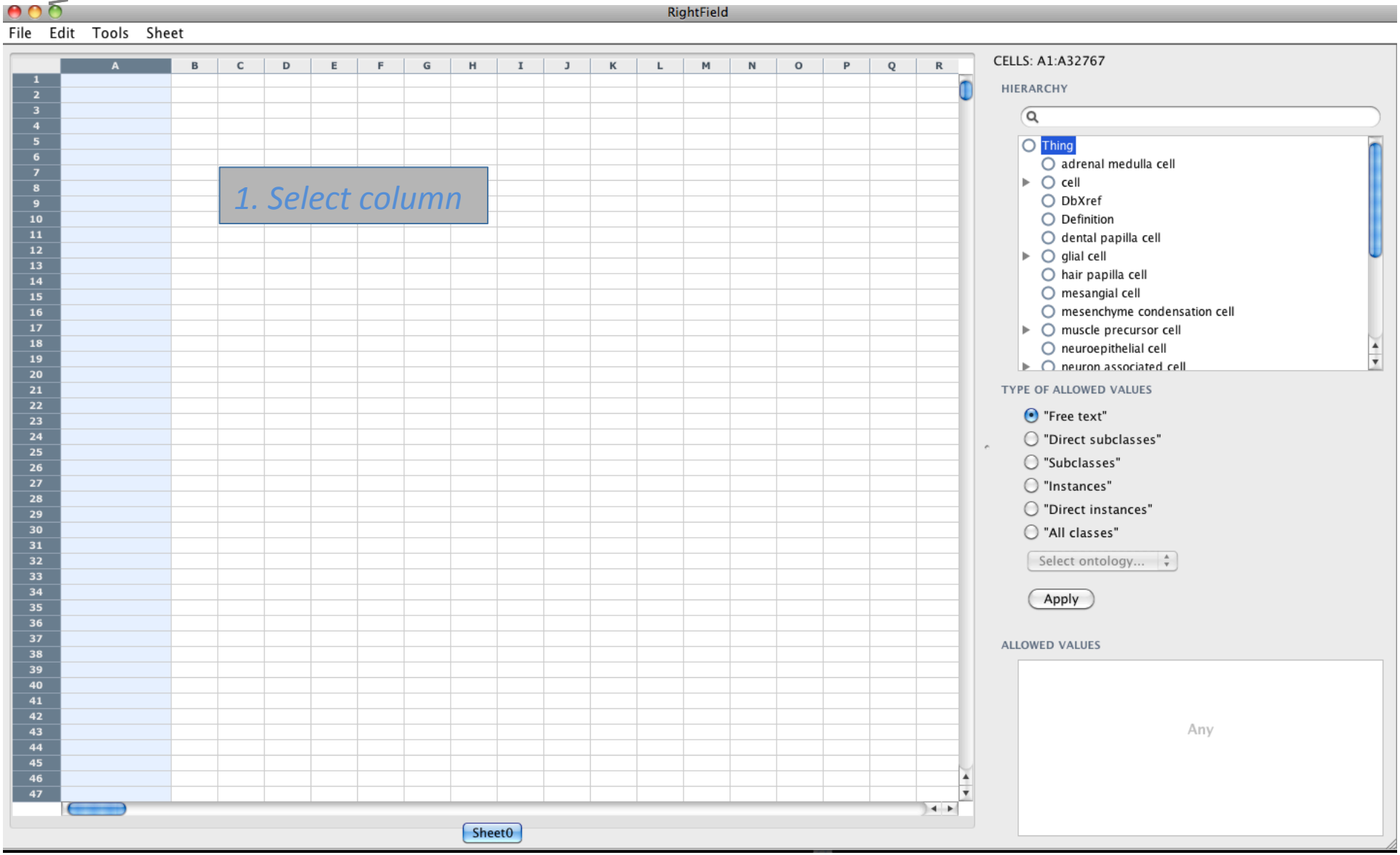
Creating Templates



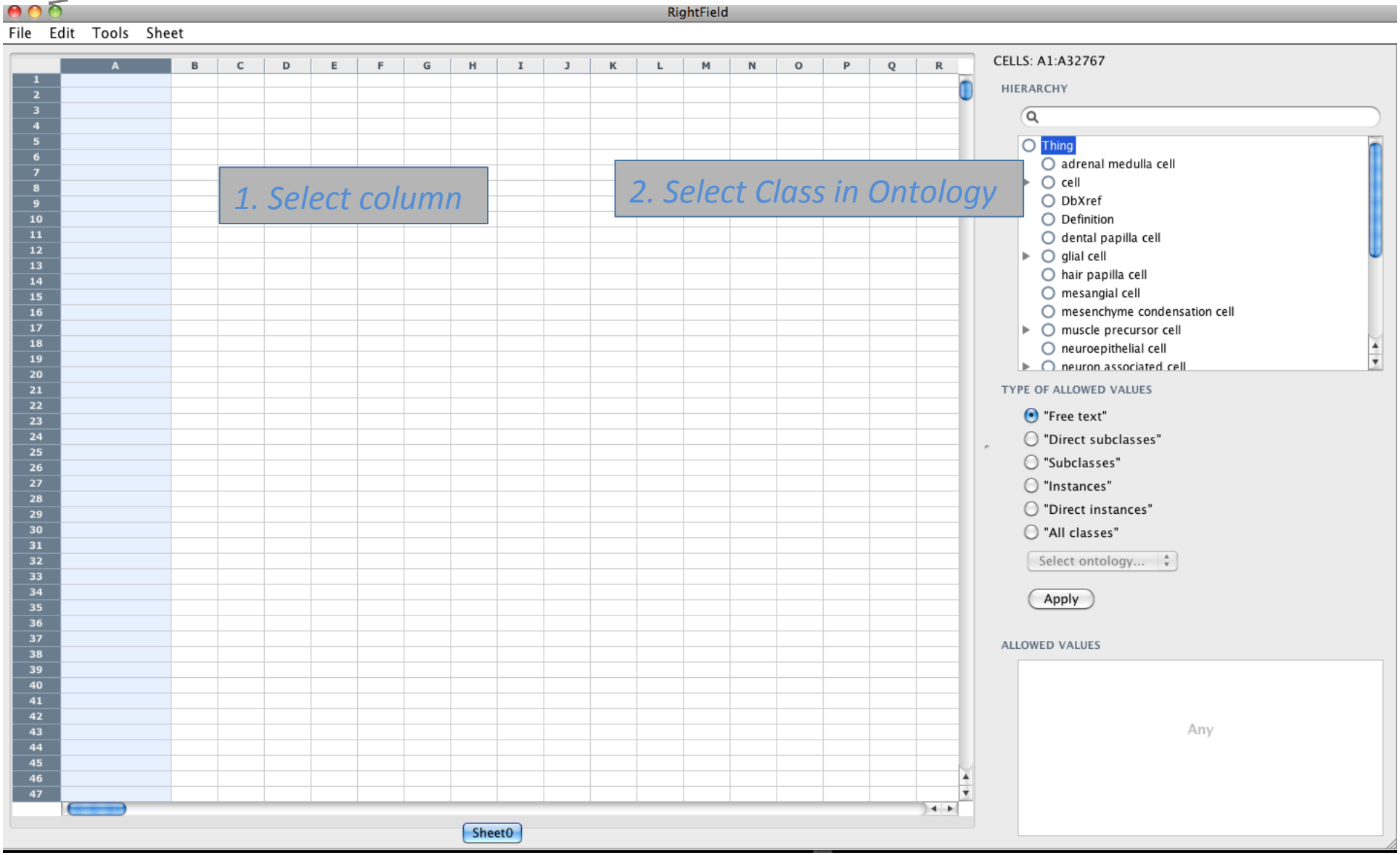
Creating Templates



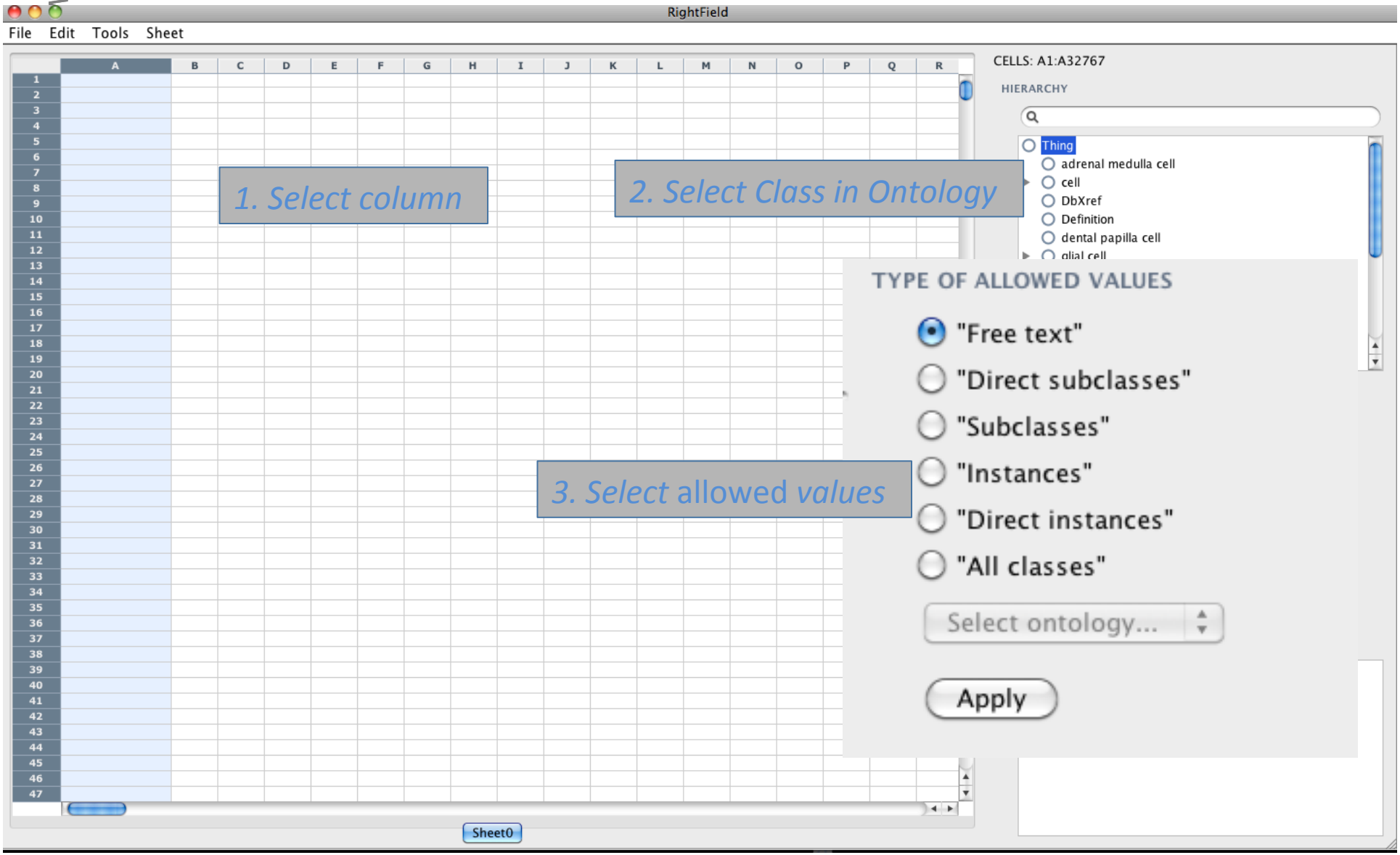
Creating Templates



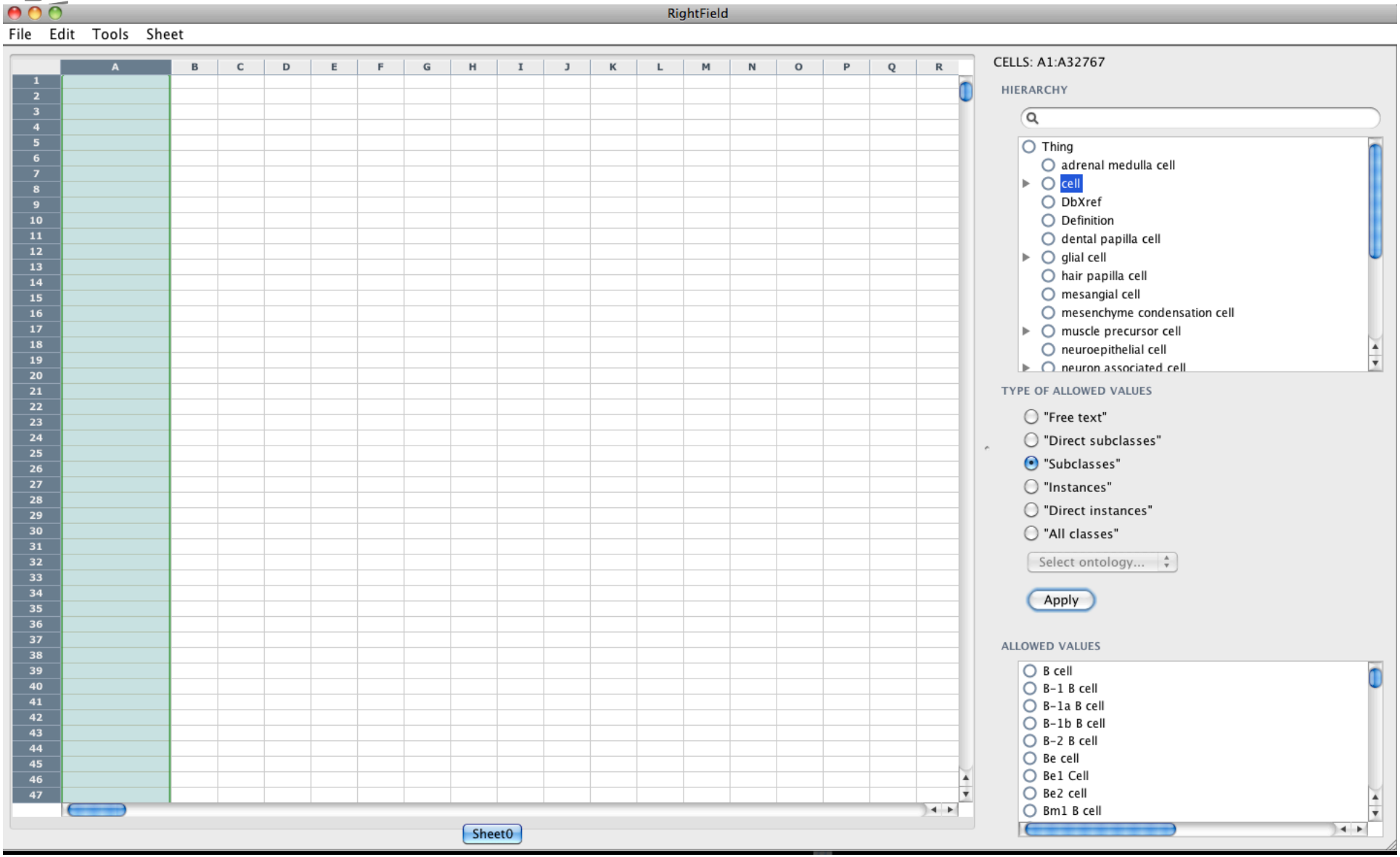
Creating Templates



Creating Templates

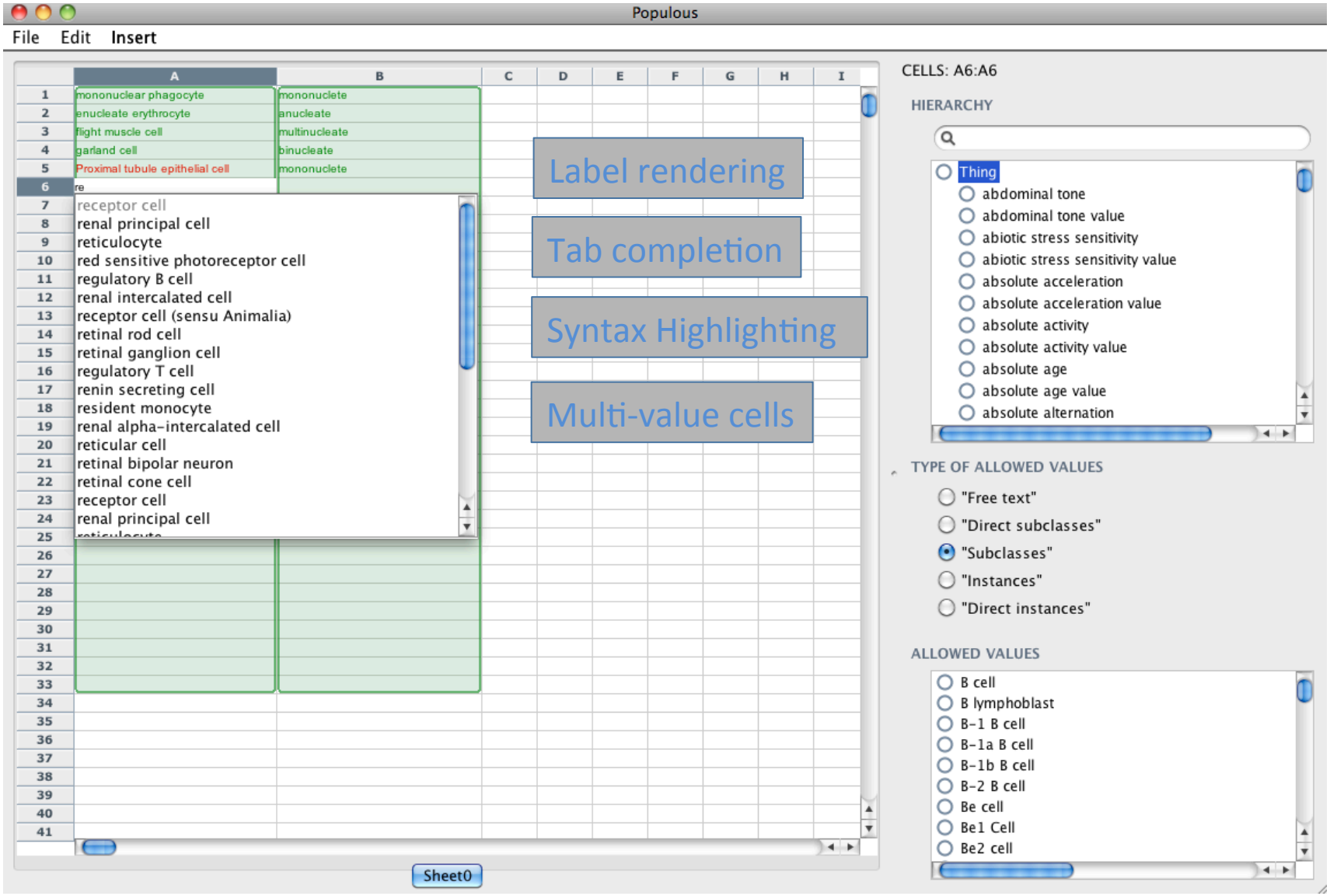


Creating Templates



The University of Manchester

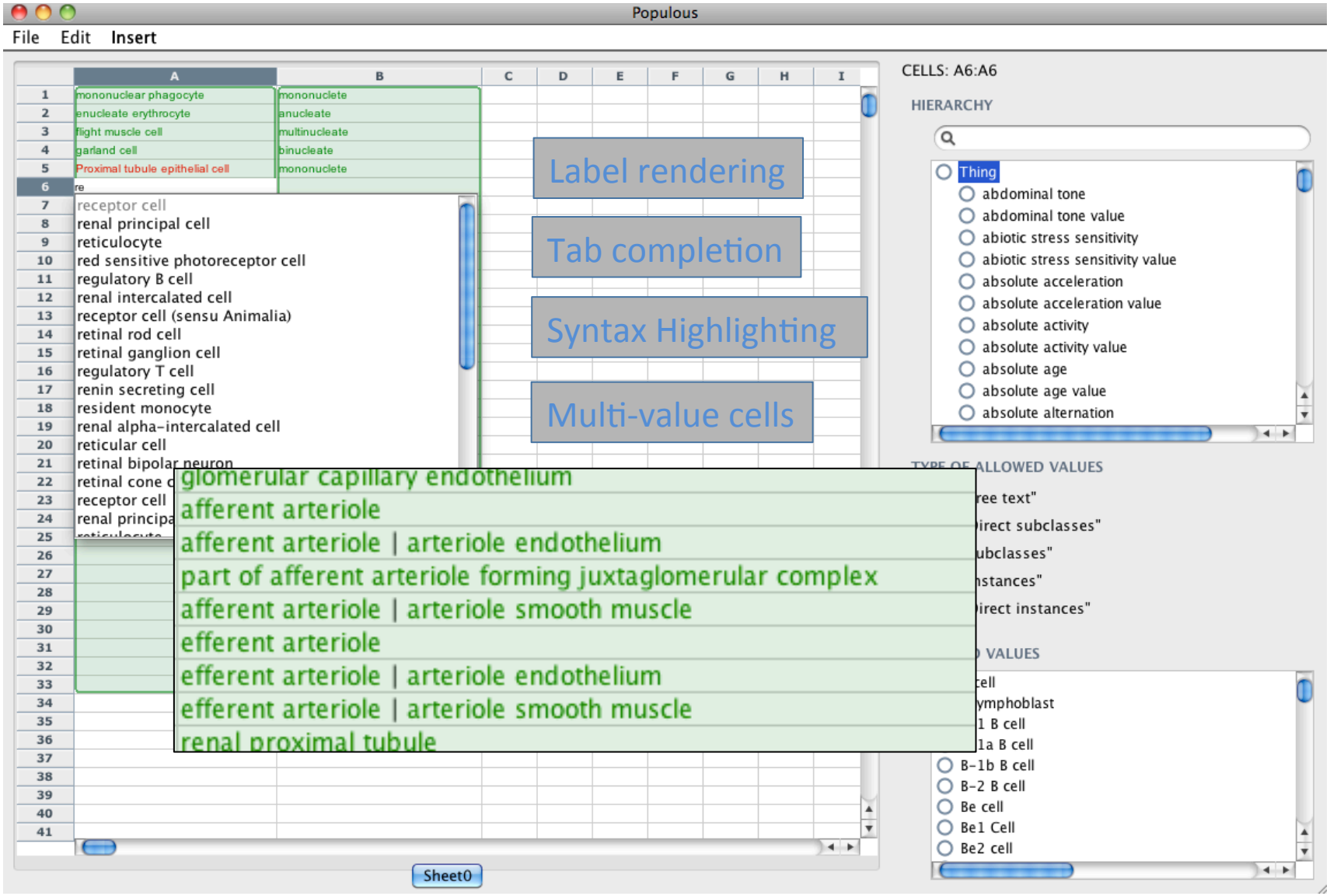
Editing Templates



MANCHESTER 1824

The University of Manchester

Editing Templates



Ontology Pre-Processing Language

Pattern

```
A cell type is equivalent to a cell type
that is part of some anatomy
```

Ontology Pre-Processing Language

Pattern

```
A cell type is equivalent to a cell type
that is part of some anatomy
```

OPPL Script

```
?cell:CLASS,
?anatomyPart:CLASS,
?anatomy:CLASS =
      (CL:0000000 part_of some ?anatomyPart)

BEGIN
ADD ?cell equivalentTo ?anatomy
END;
```

Variables

Create axioms

MANCHESTER 1824

The University of Manchester

Ontology Pre-Processing Language

Pattern

```
A cell type is equivalent to a cell type
that is part of some anatomy
```

Variable mapper

```
?cell -> 'Kidney Cell'[CL:0003523]
?anatomyPart -> 'Kidney'[FMA:629093]
```

OPPL Script

```
?cell:CLASS,
?anatomyPart:CLASS,
?anatomy:CLASS =
      (CL:0000000 part_of some ?anatomyPart)

BEGIN
ADD ?cell equivalentTo ?anatomy
END;
```

MANCHESTER 1824

The University of Manchester

Resulting OWL axioms

Example

```
A ‘Kidney Cell’ is equivalent to a cell
that is part of the ‘Kidney’
```

Generated OWL (Manchester Syntax)

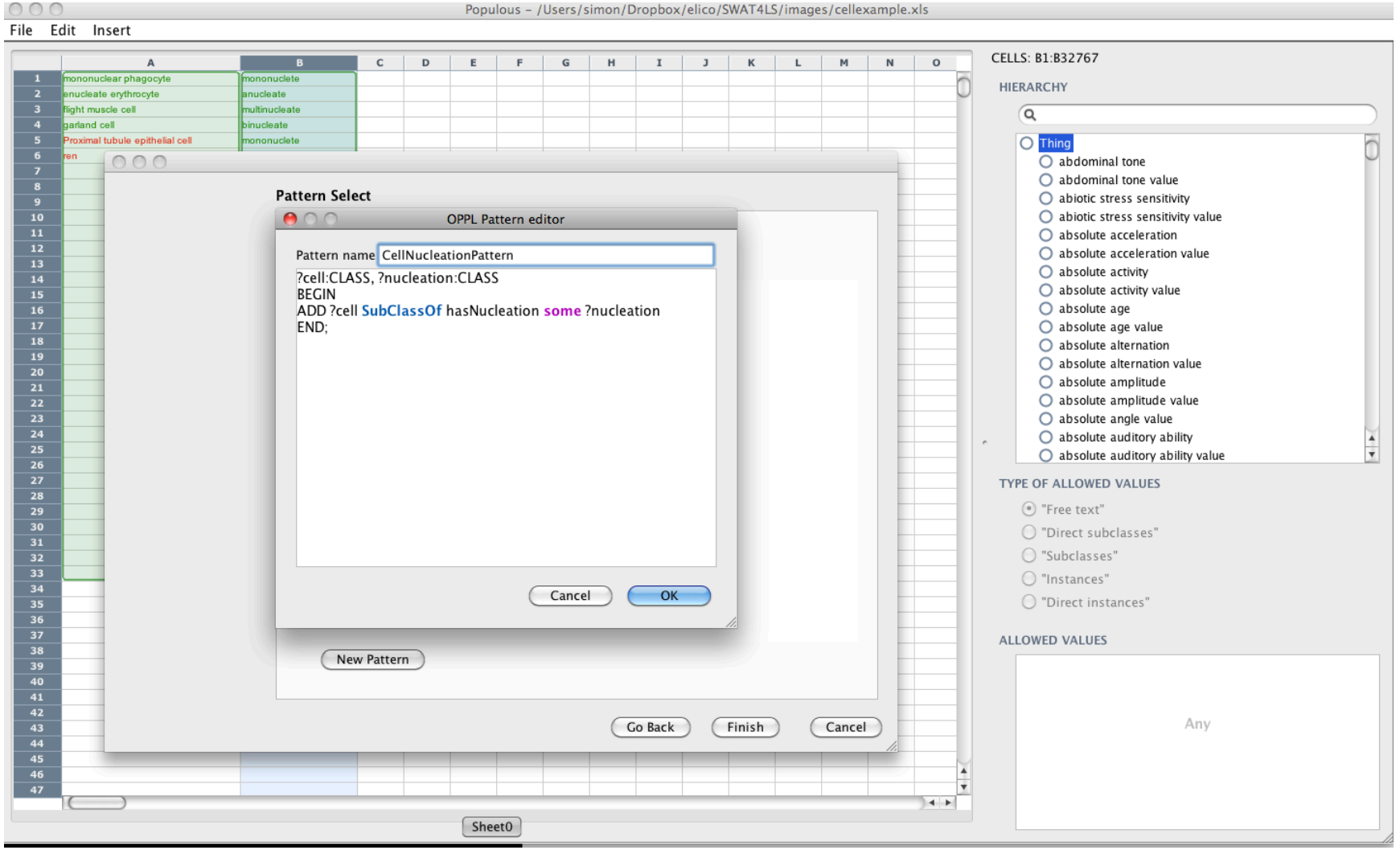
```
Class: CL:0003523

Annotation:
rdfs:label ‘Kidney Cell’

EquivalentTo:
CL:0000000 and OBO_REL:part_of some FMA:629093
```

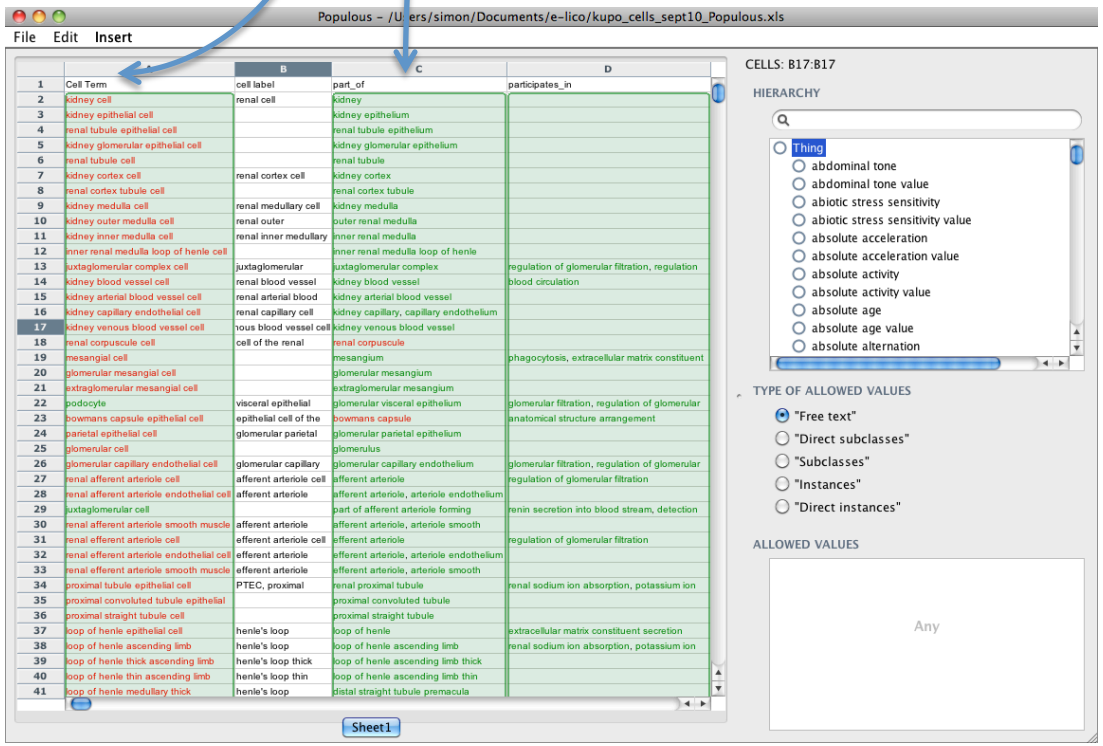
The University of Manchester

OPPL Wizzard

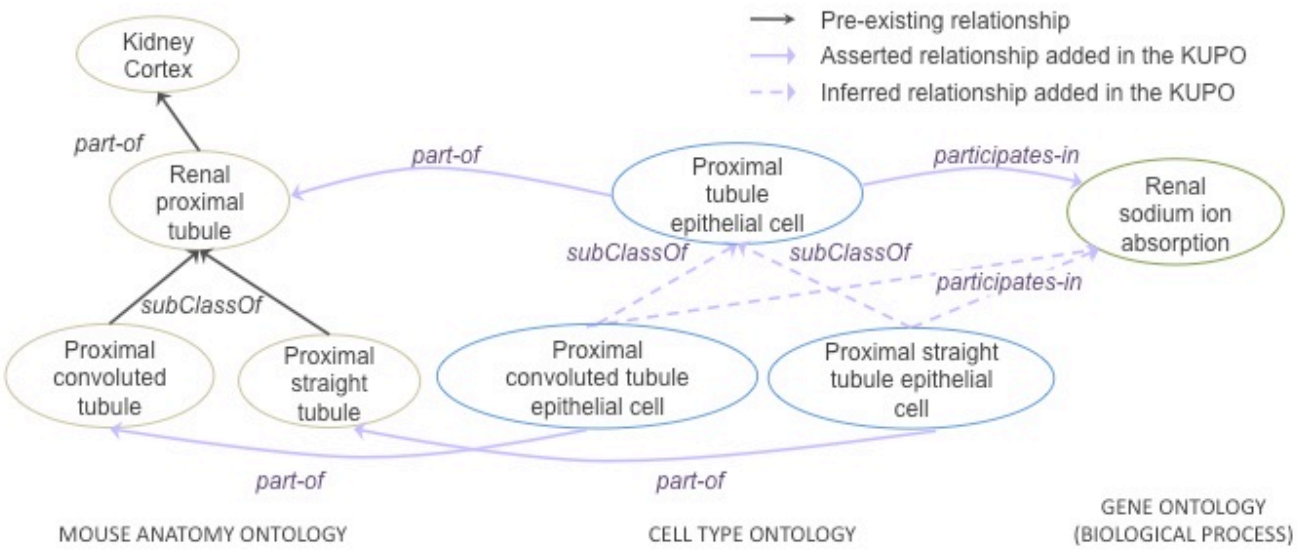


Variable mapping

```
?cell:CLASS,
?anatomyPart:CLASS,
?partOfRestriction:CLASS = CL_0000000 and part_of some ?anatomyPart,
?anatomyIntersection:CLASS = createIntersection(?
partOfRestriction.VALUES)
BEGIN
ADD ?cell equivalentTo ?anatomyIntersection
END;
```



Use case – Kidney and Urinary Pathway Ontology (KUPO)



MANCHESTER 1824

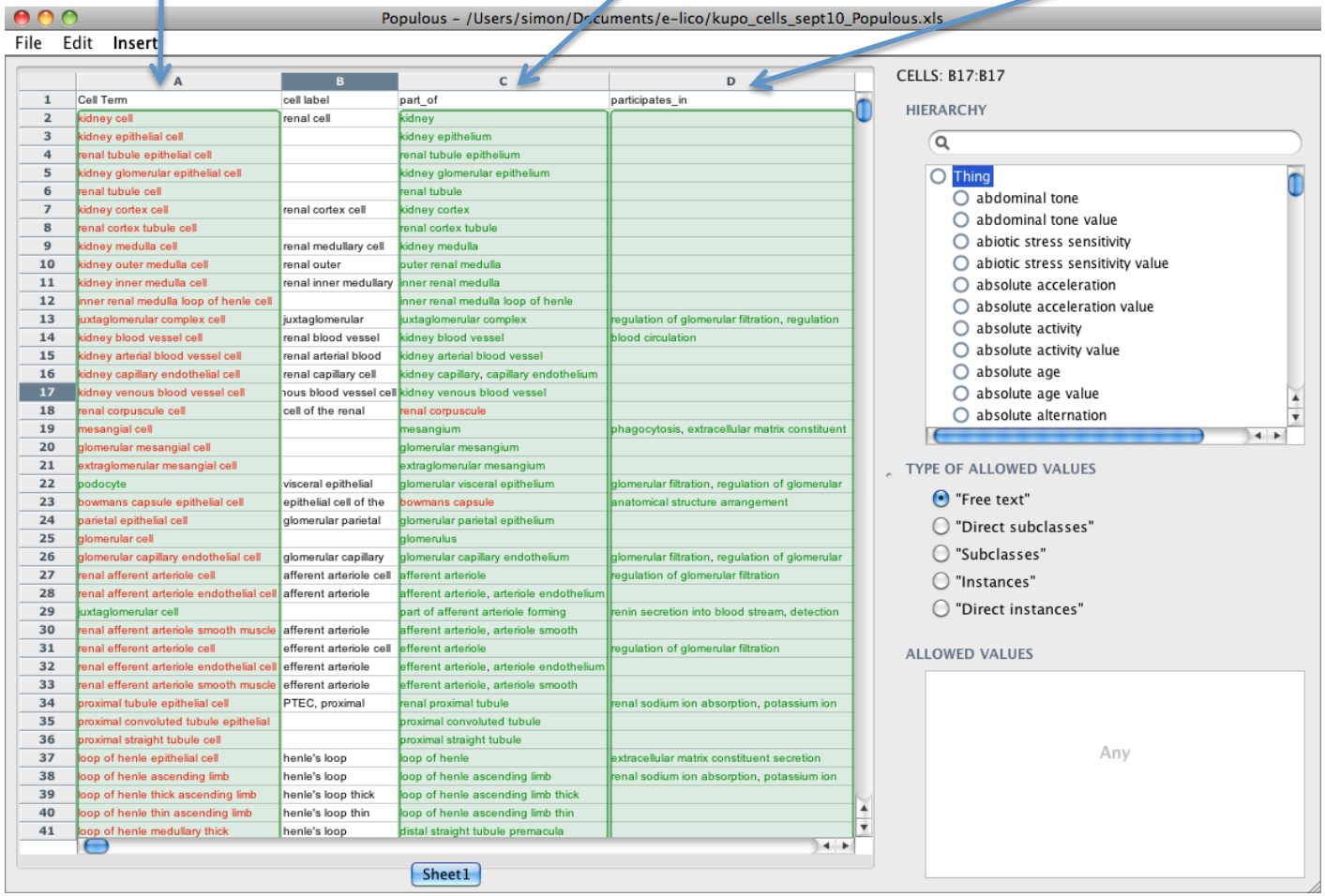
The University of Manchester

KUPO in Populous

Cell type Ontology

Mouse anatomy Ontology

Biological Process (GO)



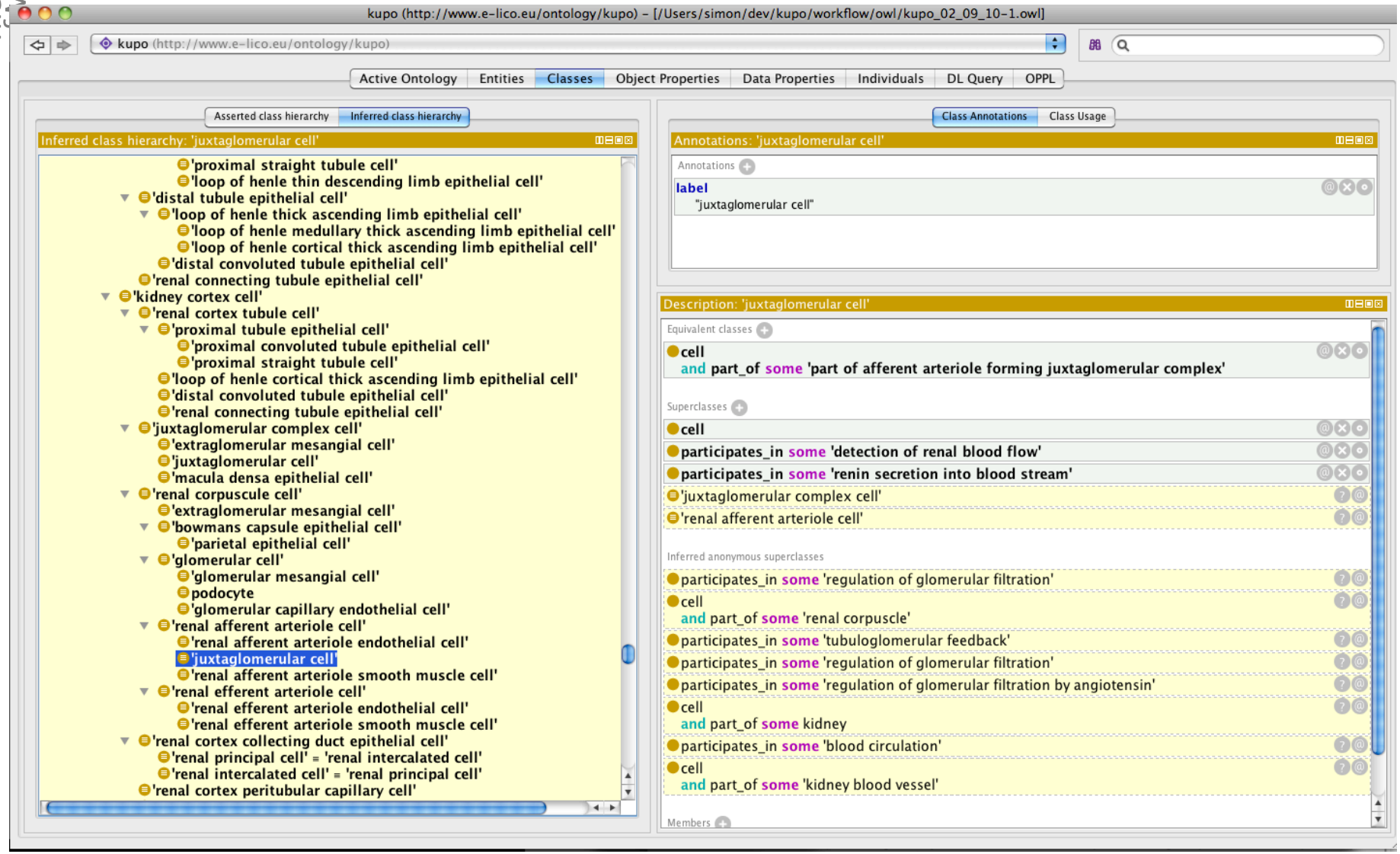
	A	B	C	D
1	Cell Term	cell label	part_of	participates_in
2	kidney cell	renal cell	kidney	
3	kidney epithelial cell		kidney epithelium	
4	renal tubule epithelial cell		renal tubule epithelium	
5	kidney glomerular epithelial cell		kidney glomerular epithelium	
6	renal tubule cell		renal tubule	
7	kidney cortex cell	renal cortex cell	kidney cortex	
8	renal cortex tubule cell		renal cortex tubule	
9	kidney medulla cell	renal medullary cell	kidney medulla	
10	kidney outer medulla cell	renal outer	outer renal medulla	
11	kidney inner medulla cell	renal inner medullary	inner renal medulla	
12	inner renal medulla loop of henle cell		inner renal medulla loop of henle	
13	juxtaglomerular complex cell	juxtaglomerular	juxtaglomerular complex	regulation of glomerular filtration, regulation
14	kidney blood vessel cell	renal blood vessel	kidney blood vessel	blood circulation
15	kidney arterial blood vessel cell	renal arterial blood	kidney arterial blood vessel	
16	kidney capillary endothelial cell	renal capillary cell	kidney capillary, capillary endothelium	
17	kidney venous blood vessel cell	nous blood vessel cell	kidney venous blood vessel	
18	renal corpuscule cell	cell of the renal	renal corpuscule	
19	mesangial cell		mesangium	phagocytosis, extracellular matrix constituent
20	glomerular mesangial cell		glomerular mesangium	
21	extraglomerular mesangial cell		extraglomerular mesangium	
22	podocyte	visceral epithelial	glomerular visceral epithelium	glomerular filtration, regulation of glomerular
23	bowmans capsule epithelial cell	epithelial cell of the	bowmans capsule	anatomical structure arrangement
24	parietal epithelial cell	glomerular parietal	glomerular parietal epithelium	
25	glomerular cell		glomerulus	
26	glomerular capillary endothelial cell	glomerular capillary	glomerular capillary endothelium	glomerular filtration, regulation of glomerular
27	renal afferent arteriole cell	afferent arteriole cell	afferent arteriole	regulation of glomerular filtration
28	renal afferent arteriole endothelial cell	afferent arteriole	afferent arteriole, arteriole endothelium	
29	juxtaglomerular cell		part of afferent arteriole forming	renin secretion into blood stream, detection
30	renal afferent arteriole smooth muscle	afferent arteriole	afferent arteriole, arteriole smooth	
31	renal efferent arteriole cell	efferent arteriole cell	efferent arteriole	regulation of glomerular filtration
32	renal efferent arteriole endothelial cell	efferent arteriole	efferent arteriole, arteriole endothelium	
33	renal efferent arteriole smooth muscle	efferent arteriole	efferent arteriole, arteriole smooth	
34	proximal tubule epithelial cell	PTEC, proximal	renal proximal tubule	renal sodium ion absorption, potassium ion
35	proximal convoluted tubule epithelial		proximal convoluted tubule	
36	proximal straight tubule cell		proximal straight tubule	
37	loop of henle epithelial cell	henle's loop	loop of henle	extracellular matrix constituent secretion
38	loop of henle ascending limb	henle's loop	loop of henle ascending limb	renal sodium ion absorption, potassium ion
39	loop of henle thick ascending limb	henle's loop thick	loop of henle ascending limb thick	
40	loop of henle thin ascending limb	henle's loop thin	loop of henle ascending limb thin	
41	loop of henle medullary thick	henle's loop	distal straight tubule premacula	

MANCHESTER 1824

The University of Manchester

BHIG

KUPO in OWL



MANCHESTER 1824

The University of Manchester

Limitations of approach

- Assumes regular patterns
  - Developing patterns is hard
  - Not always sensible
- Handling exceptions to patterns
  - Need special syntax, but how far do you go…
- Regular data (entity-per-row assumption)

MANCHESTER 1824

The University of Manchester

Future plans

- Mapping Master Support
- Additional syntax
- Ontology Alignment
- Template creation from Protégé OPPL patterns plugin.
- Building KUP knowledge base

MANCHESTER 1824

The University of Manchester

Summary

- Populous is for populating templates
- Focus on supporting domain experts
- Expressive pattern language for modeling OWL
- Abstract knowledge from the modeling
- Engaging the experts by stealth!

MANCHESTER 1824

The University of Manchester

Acknowledgments

- **RightField**

  Matthew Horridge, Katy Wolstencroft, Stuart Owen, Carole Goble

- **OPPL**

  Luigi Iannone, Mikel Aranguren, Alan Rector, Robert Stevens

- **KUPO**

  Robert Stevens, Julie Klein, Joost Schanstra

- e-**LICO**

  EU-FP7 Collaborative Project (2009-2012)
  Theme ICT-4.4: Intelligent Content and Semantics