Harnessing the Semantic Web to Answer Scientific Questions:

A Health Care and Life Sciences Interest Group demo

Alan Ruttenberg, Science Commons

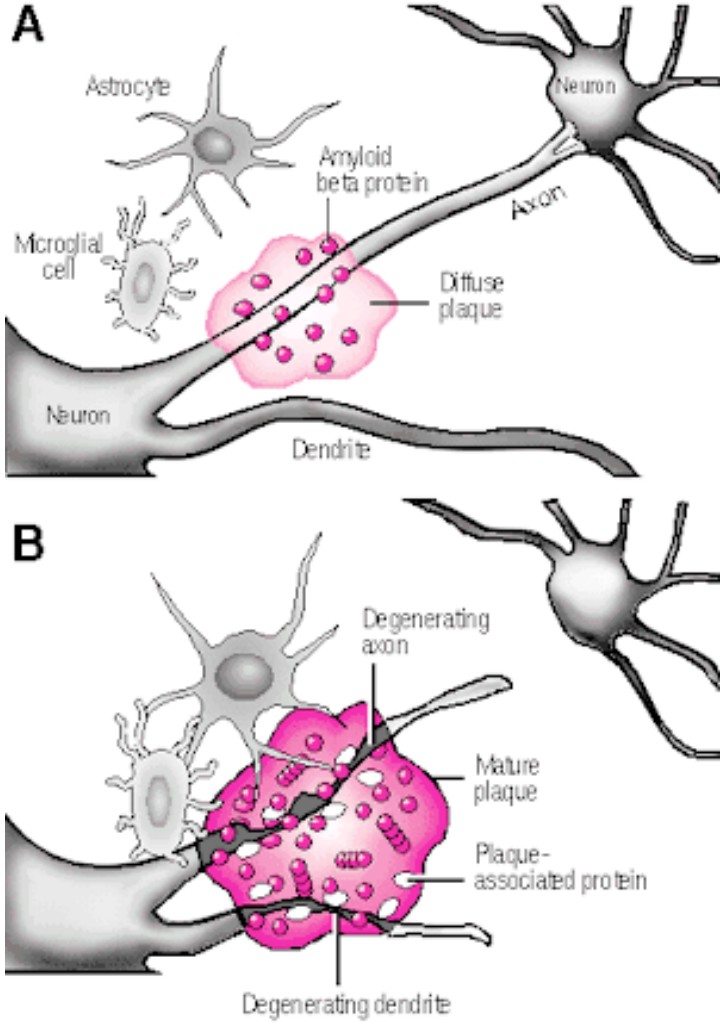
science commons

Accelerating the Scientific Research Cycle

Four examples

- A simple SPARQL query
- Looking for Alzheimer's Disease targets
- Mashup: Google Maps/SPARQL/Allen Brain Atlas
- Exploring ontologies with LSW (Lisp Semantic Web)

A simple query:
Biological processes in dendrites?



Alzheimer's disease is characterized by neural degeneration. Among other things, there is damage to dendrites and axons, parts of nerve cells.

What resources do we have available to learn more about biological processes in dendrites?

The Gene Ontology

The gene ontology names many biological processes and tells us which genes are known to be involved in those processes.

Let’s ask it what processes it knows about involving dendrites

Biological processes naming dendrites

```
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX go: <http://purl.org/obo/owl/GO#>
PREFIX obo: <http://www.geneontology.org/formats/oboInOwl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

select  ?name  ?class ?definition
from <http://purl.org/commons/hcls/20070416>
where
{   graph <http://purl.org/commons/hcls/20070416/classrelations>
     {?class rdfs:subClassOf go:GO_0008150}
    ?class rdfs:label ?name.
    ?class obo:hasDefinition ?def.
    ?def rdfs:label ?definition
    filter(regex(?name,"[Dd]endrite"))
}
```

URI for Biological Process

From the “console”

result | request | response

name	class	definition
dendrite development	http://purl.org/obo/owl/GO#GO_0016358	The process whose specific outcome is the progression of the dendrite over time, from its formation to the mature structure. A dendrite is a freely branching protoplasmic process of a nerve cell.
dendrite regeneration	http://purl.org/obo/owl/GO#GO_0031104	The regrowth of dendrites following their loss or damage.
dendrite morphogenesis	http://purl.org/obo/owl/GO#GO_0048813	The process by which the anatomical structures of dendrite are generated and organized. Morphogenesis pertains to the creation of form. A dendrite is a freely branching protoplasmic process of a nerve cell.
regulation of dendrite morphogenesis	http://purl.org/obo/owl/GO#GO_0048814	Any process that modulates the frequency, rate or extent of dendrite morphogenesis.
regulation of dendrite development	http://purl.org/obo/owl/GO#GO_0050773	Any process that modulates the frequency, rate or extent of dendrite development.
negative regulation of dendrite morphogenesis	http://purl.org/obo/owl/GO#GO_0050774	Any process that stops, prevents or reduces the frequency, rate or extent of dendrite morphogenesis.
positive regulation of dendrite morphogenesis	http://purl.org/obo/owl/GO#GO_0050775	Any process that activates or increases the frequency, rate or extent of dendrite morphogenesis.

But the answers are also available by a “GET”

```
/sparql/?query=PREFIX%20owl%3A%20%3Chttp%3A%2F%2Fwww.w3.org
%2F2002%2F07%2Fowl%23%3E%0APREFIX%20go%3A%20%3Chttp%3A%2F%2
Fpurl.org%2Fobo%2Fowl%2FGO%23%3E%0APREFIX%20obo%3A%20%3Chtt
p%3A%2F%2Fwww.geneontology.org%2Fformats%2FoboInOwl%23%3E%0
APREFIX%20rdfs%3A%20%3Chttp%3A%2F%2Fwww.w3.org%2F2000%2F01%
2Frdf-
schema%23%3E%0A%0Aselect%20%20%3Fname%20%20%3Fclass%20%3Fde
finition%0Afrom%20%3Chttp%3A%2F%2Fpurl.org%2Fcommons%2Fhcls
%2F20070416%3E%0Awhere%0A%7B%20%20%20graph%20%3Chttp%3A%2F%
2Fpurl.org%2Fcommons%2Fhcls%2F20070416%2Fclassrelations%3E%
0A%20%20%20%20%20%7B%3Fclass%20rdfs%3AsubClassOf%20go%3AGO_
0008150%7D%0A%20%20%20%20%3Fclass%20rdfs%3Alabel%20%3Fname.
%0A%20%20%20%20%3Fclass%20obo%3AhasDefinition%20%3Fdef.%0A%
20%20%20%20%3Fdef%20rdfs%3Alabel%20%3Fdefinition%20%0A%20%2
0%20%20filter(regex(%3Fname%2C%22%5BDd%5Dendrite%22))%0A%7D
%0A&format=&maxrows=50
```

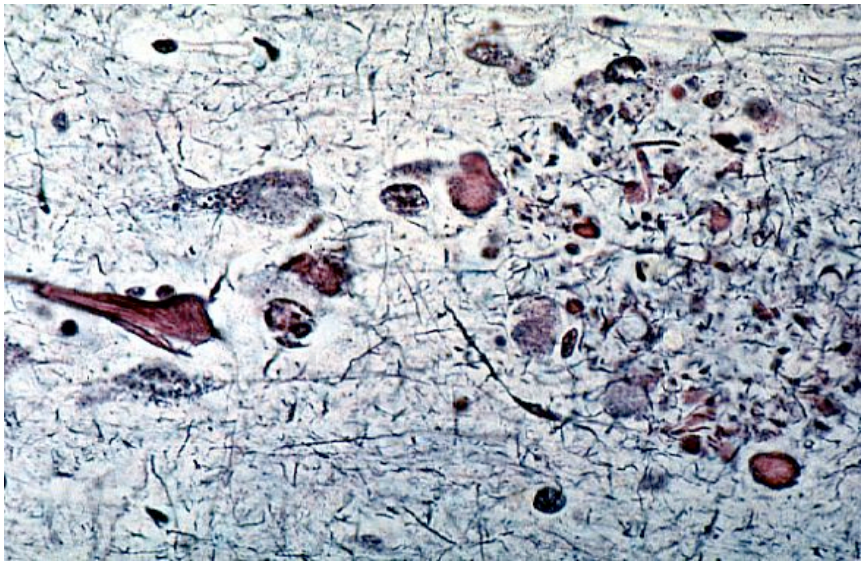
So someone, somewhere else, can build something better

Looking for Alzheimer Disease targets

Signal transduction pathways are considered to be rich in “druggable” targets - proteins that might respond to chemical therapy

CA1 Pyramidal Neurons are known to be particularly damaged in Alzheimer’s disease.

Casting a wide net, can we find candidate genes known to be involved in signal transduction and active in Pyramidal Neurons?



A SPARQL query spanning 4 sources

```
prefix go: <http://purl.org/obo/owl/GO#>
prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
prefix owl: <http://www.w3.org/2002/07/owl#>
prefix mesh: <http://purl.org/commons/record/mesh/>
prefix sc: <http://purl.org/science/owl/sciencecommons/>
prefix ro: <http://www.obofoundry.org/ro/ro.owl#>

select ?genename ?processname
where
{  graph <http://purl.org/commons/hcls/pubmesh>
    { ?paper ?p mesh:D017966 .
      ?article sc:identified_by_pmid ?paper.
      ?gene sc:describes_gene_or_gene_product_mentioned_by ?article.
    }
  graph <http://purl.org/commons/hcls/goa>
    { ?protein rdfs:subClassOf ?res.
      ?res owl:onProperty ro:has_function.
      ?res owl:someValuesFrom ?res2.
      ?res2 owl:onProperty ro:realized_as.
      ?res2 owl:someValuesFrom ?process.
   graph <http://purl.org/commons/hcls/20070416/classrelations>
    {{?process <http://purl.org/obo/owl/obo#part_of> go:GO_0007166}
      union
     {?process rdfs:subClassOf go:GO_0007166 }}
      ?protein rdfs:subClassOf ?parent.
      ?parent owl:equivalentClass ?res3.
      ?res3 owl:hasValue ?gene.
     }
  graph <http://purl.org/commons/hcls/gene>
    { ?gene rdfs:label ?genename }
  graph <http://purl.org/commons/hcls/20070416>
    { ?process rdfs:label ?processname}
}
```

Mesh: Pyramidal Neurons

↓

Pubmed: Journal Articles

↓

Entrez Gene: Genes

↓

GO: Signal Transduction

Inference required

Results

Many of the genes are indeed related to Alzheimer's Disease through gamma secretase (presenilin) activity

DRD1, 1812	adenylate cyclase activation
ADRB2, 154	adenylate cyclase activation
ADRB2, 154	arrestin mediated desensitization of G-protein coupled receptor protein signaling pathway
DRD1IP, 50632	dopamine receptor signaling pathway
DRD1, 1812	dopamine receptor, adenylate cyclase activating pathway
DRD2, 1813	dopamine receptor, adenylate cyclase inhibiting pathway
GRM7, 2917	G-protein coupled receptor protein signaling pathway
GNG3, 2785	G-protein coupled receptor protein signaling pathway
GNG12, 55970	G-protein coupled receptor protein signaling pathway
DRD2, 1813	G-protein coupled receptor protein signaling pathway
ADRB2, 154	G-protein coupled receptor protein signaling pathway
CALM3, 808	G-protein coupled receptor protein signaling pathway
HTR2A, 3356	G-protein coupled receptor protein signaling pathway
DRD1, 1812	G-protein signaling, coupled to cyclic nucleotide second messenger
SSTR5, 6755	G-protein signaling, coupled to cyclic nucleotide second messenger
MTNR1A, 4543	G-protein signaling, coupled to cyclic nucleotide second messenger
CNR2, 1269	G-protein signaling, coupled to cyclic nucleotide second messenger
HTR6, 3362	G-protein signaling, coupled to cyclic nucleotide second messenger
GRIK2, 2898	glutamate signaling pathway
GRIN1, 2902	glutamate signaling pathway
GRIN2A, 2903	glutamate signaling pathway
GRIN2B, 2904	glutamate signaling pathway
ADAM10, 102	integrin-mediated signaling pathway
GRM7, 2917	negative regulation of adenylate cyclase activity
LRP1, 4035	negative regulation of Wnt receptor signaling pathway
ADAM10, 102	Notch receptor processing
ASCL1, 429	Notch signaling pathway
HTR2A, 3356	serotonin receptor signaling pathway
ADRB2, 154	transmembrane receptor protein tyrosine kinase activation (dimerization)
PTPRG, 5793	transmembrane receptor protein tyrosine kinase signaling pathway
EPHA4, 2043	transmembrane receptor protein tyrosine kinase signaling pathway
NRTN, 4902	transmembrane receptor protein tyrosine kinase signaling pathway
CTNND1, 1500	Wnt receptor signaling pathway

The Allen Brain Atlas

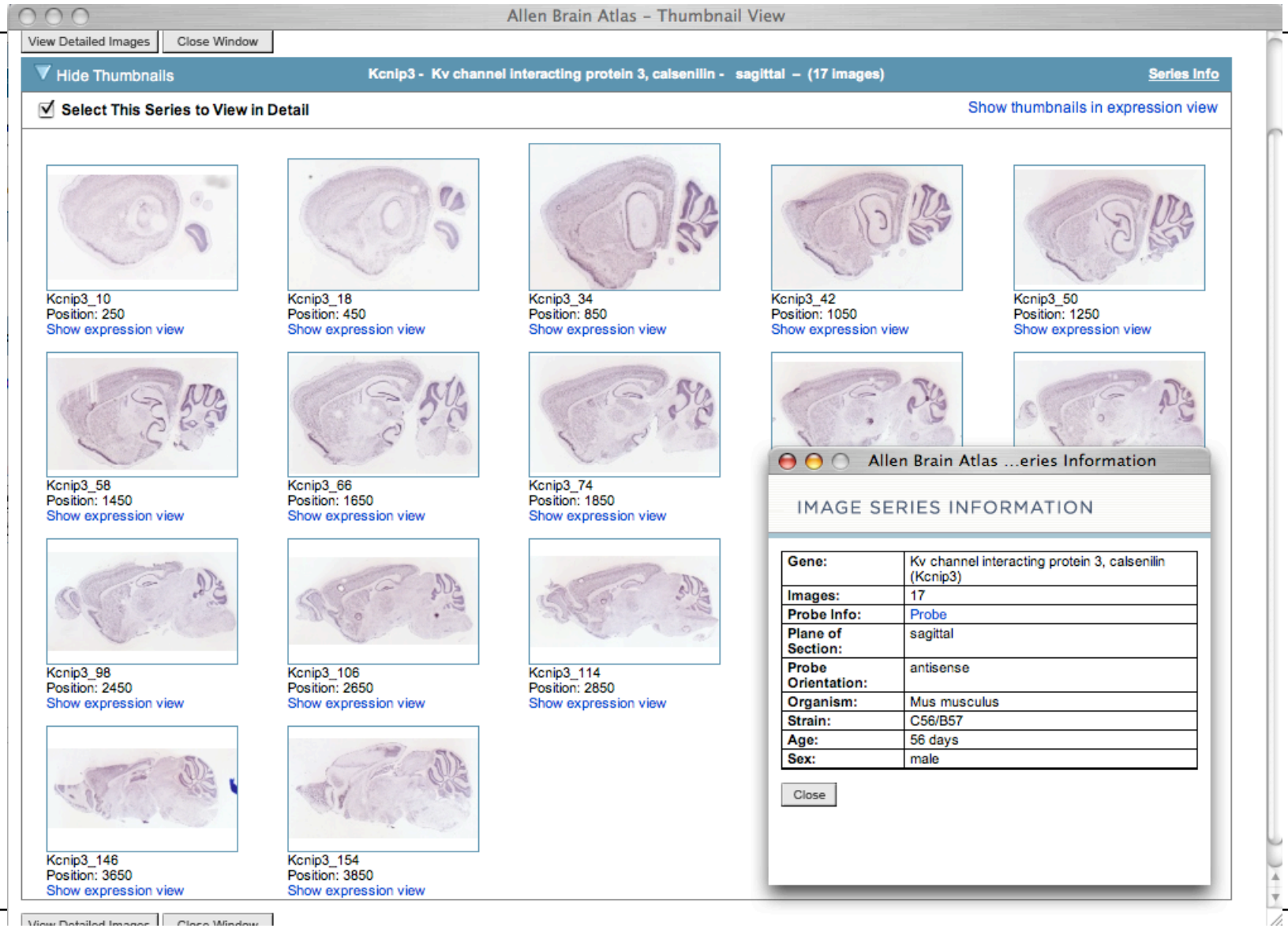
A remarkable scientific achievement. Mouse brains cut into thin slices and stained for the presence of gene expression.

20,000 genes, 400000 images at high resolution.

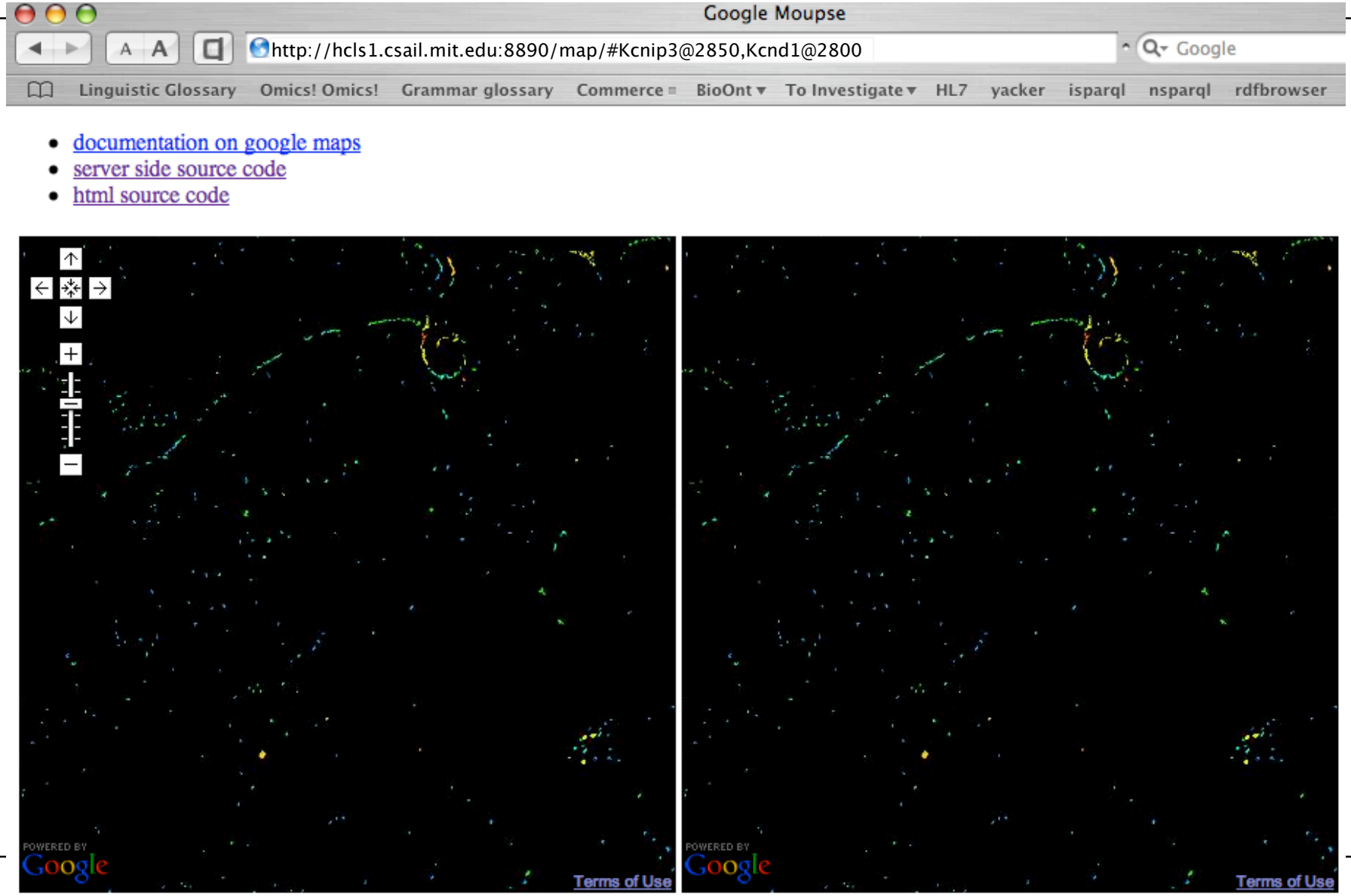
Currently available only through an HTML interface.

Scrape 80K web pages to extract the information, convert to RDF

The Allen Brain Atlas



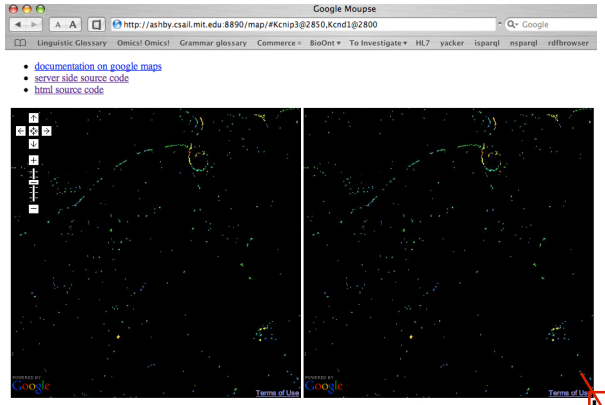
Google Maps/SPARQL/Allen Brain Atlas



How it works (standing on the shoulders of giants)

http://hcls1/map/#Kcnip3@2850,Kcnd1@2800

Javascript



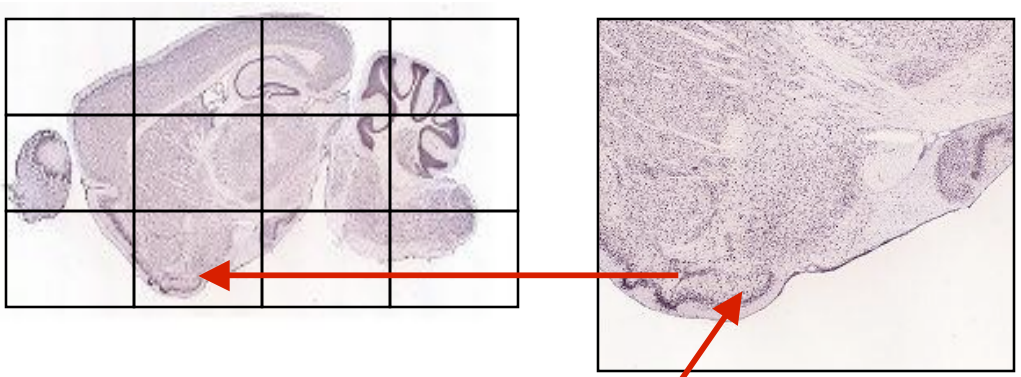
SPARQL
AJAX

Query

URL

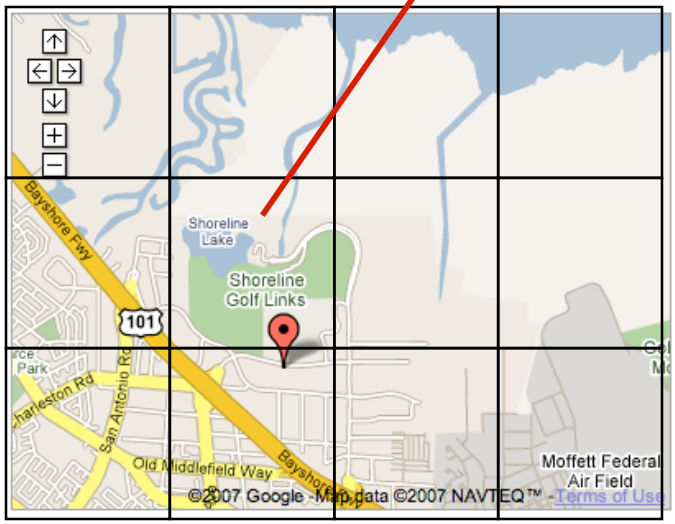
Neurocommons Servers

Allen Brain Institute Servers

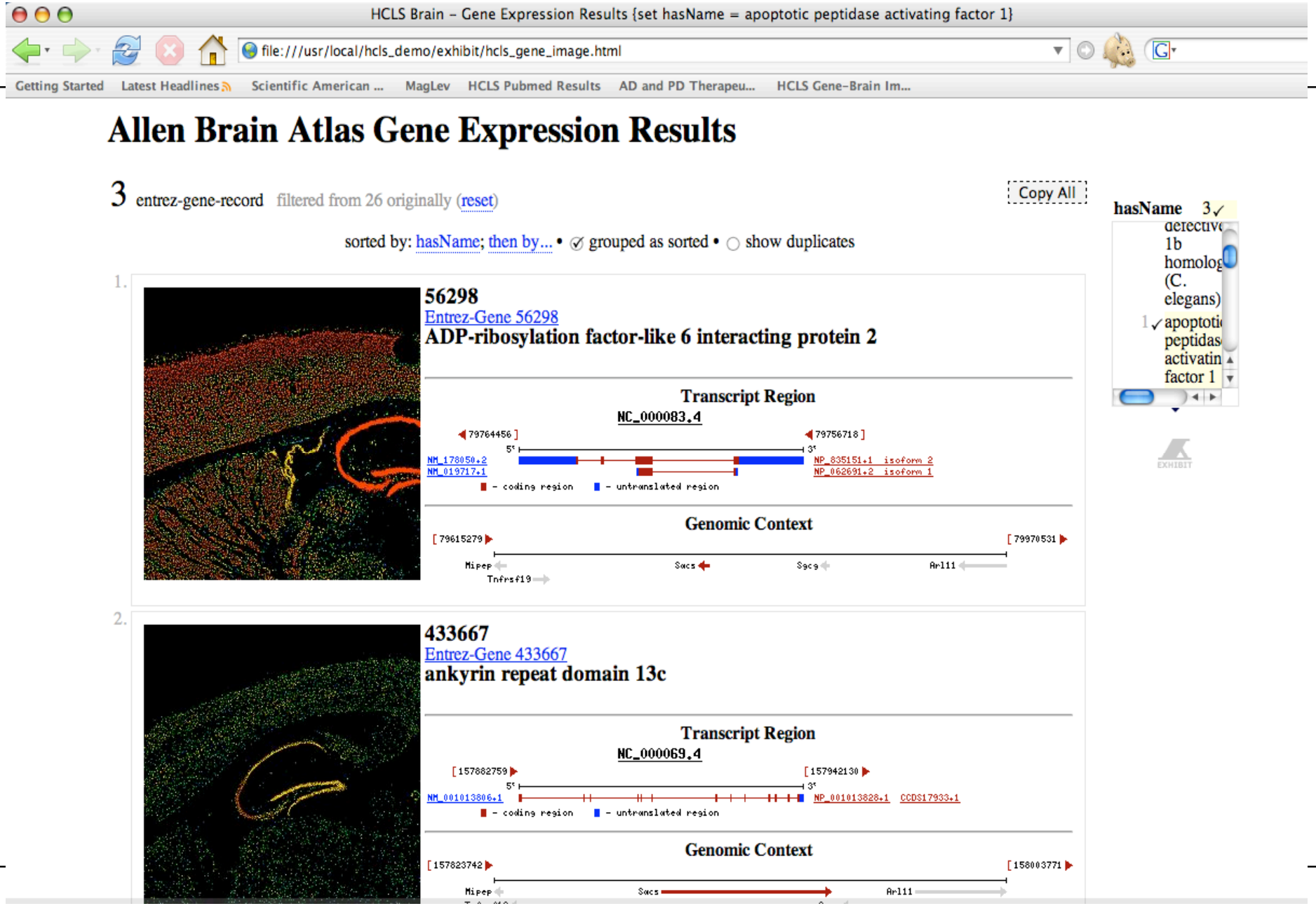


http://www.brainmap.org://….0205032816_B.aff/TileGroup3/1-0-1.jpg

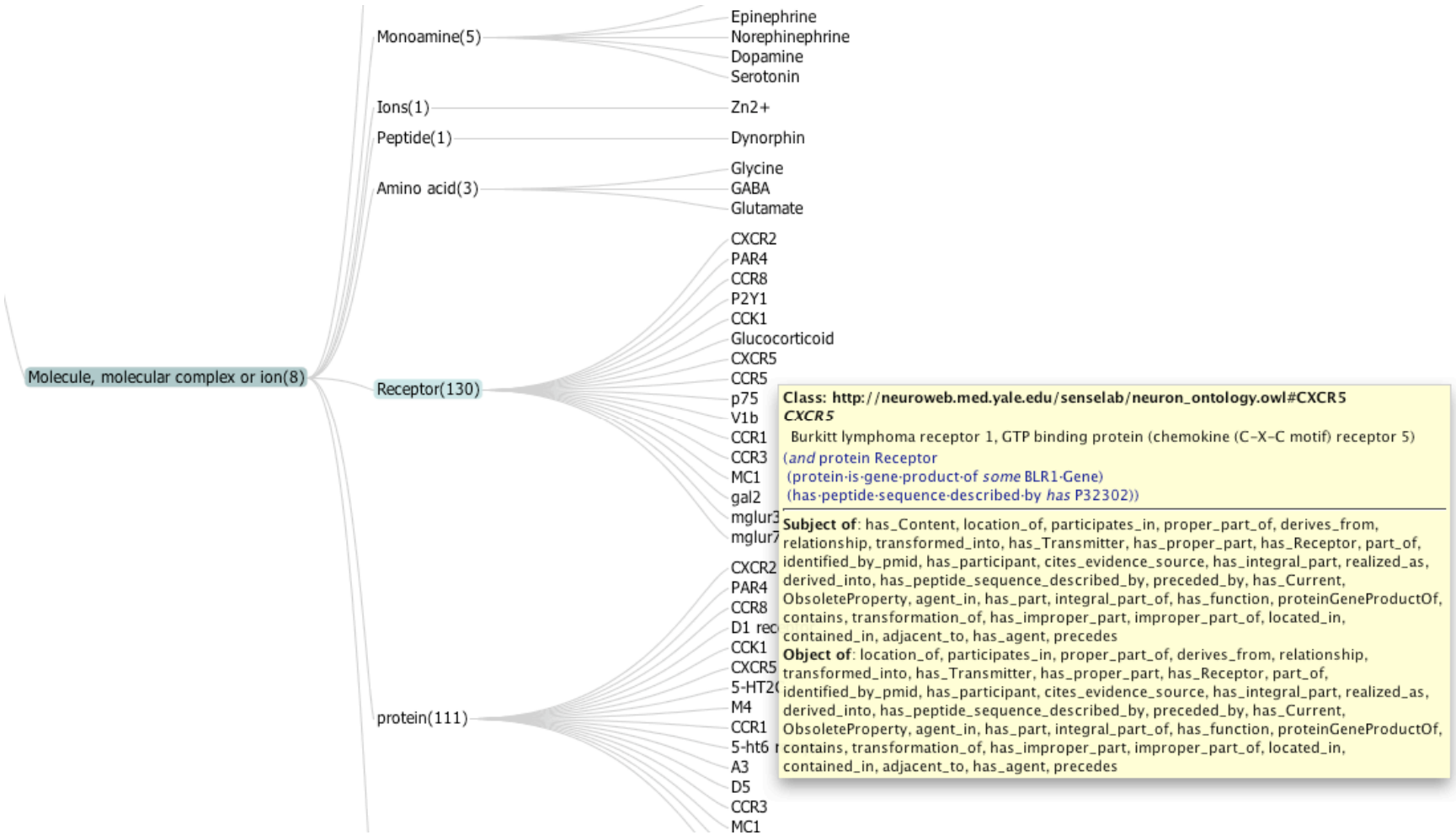
Google
Maps
API



But you can also use Exhibit to visualize



Exploring ontologies with LSW



About LSW

Jena for general SW manipulation (http://jena.sourceforge.net/)

Pellet for OWL reasoning (http://pellet.owldl.com/)

Prefuse for visualizations (http://prefuse.org/)

Armed-bear Common Lisp (http://armedbear.org/abcl.html)
(works under JVM)

So: Multi-language semantic web development environment

Mostly used for reading/writing OWL/RDF, checking consistency, querying with full inference using SPARQL, DIG, visualization, and prototyping new ideas.

LSW: http://svn.mumble.net:8080/svn/lsw/trunk/

Background Technology

So far about 350M triples in Openlink Virtuoso (~20Gb)

Commodity Hardware: 2x2core duo/2 disks/8G Ram

Biggest so far is MeSH associations to articles (200M triples)

Smaller, from 10K to 10M triples/source

A small fraction of biological knowledge!

(Don’t forget - you can still interoperate with data from relational databases)

Here's the good part!

You can play (for four more weeks at least) and download all data and install it yourself! (Thanks HP!)

Form: http://hcls1.csail.mit.edu:8890/nsparql/
Endpoint: http://hcls1.csail.mit.edu:8890/sparql/

Form: http://hcls2.csail.mit.edu:8890/nsparql/
Endpoint: http://hcls2.csail.mit.edu:8890/sparql/

More information:
http://esw.w3.org/topic/HCLS/Banff2007Demo
= http://tinyurl.com/**ywalvd**

We are actively looking for organizations to sponsor hosting this resource permanently

Acknowledgements

HCLS Demo Contributors

- John Barkley (NIST)
- Olivier Bodenreider (NLM, NIH)
- Bill Bug (Drexel University College of Medicine)
- Huajun Chen (Zhejiang University)
- Paolo Ciccarese (SWAN)
- Kei Cheung (SenseLab, Yale)
- Tim Clark (SWAN)
- Don Doherty (Brainstage Research Inc.)
- Kerstin Forsberg (AstraZeneca)
- Ray Hookaway (HP)
- Vipul Kashyap (Partners Healthcare)
- June Kinoshita (AlzForum)
- Joanne Luciano (Harvard Medical School)
- Scott Marshall (University of Amsterdam)
- Chris Mungall (NCBO)
- Eric Neumann (Teranode)
- Eric Prud'hommeaux (W3C)
- Jonathan Rees (Science Commons)
- Alan Ruttenberg (Science Commons)
- Matthias Samwald (Medical University of Vienna)

HCLS Demo Contributors

- Susie Stephens (Eli Lilly)
- Mike Travers
- Gwen Wong (SWAN)
- Elizabeth Wu (SWAN)

Data Providers

- Judith Blake (MGD)
- Mikail Bota (BAMS)
- David Hill (MGD)
- Oliver Hoffman (CL)
- Minna Lehvaslaiho (CL)
- Colin Knep (Alzforum)
- Maryanne Martone (CCDB)
- Susan McClatchy (MGD)
- Simon Twigger (RGD)
- Allen Brain Institute

Vendor Support

- **OpenLink** - Kingsley Idehen, Ivan Mikhailov, Orri Erling, Mitko Iliev, Patrick van Kleef
- **HP** - Ray Hookaway, Jeannine Crockford