Linking Up:

An overview of the W3C Semantic Web Health Care and Life Sciences Interest Group

M. Scott Marshall University of Amsterdam

Birth of the HCLSIG

- W3C Workshop on Semantic Web for Life
 Sciences, October 2004, Cambridge, MA, USA
- 115 participants
- http://www.w3.org/2004/07/swls-ws.html

HCLSIG Charter in early 2005

Recharter in June 2008

- Co-chairs are Chimezie Ogbuji (Cleveland Clinic Foundation), M. Scott Marshall (University of Amsterdam), Susie Stephens (Eli Lilly)
- W3C team contact is Eric Prud'hommeaux (W3C)
- 85 formal participants across pharma, health care, academia, and technology companies
- Mailing list of >500 people

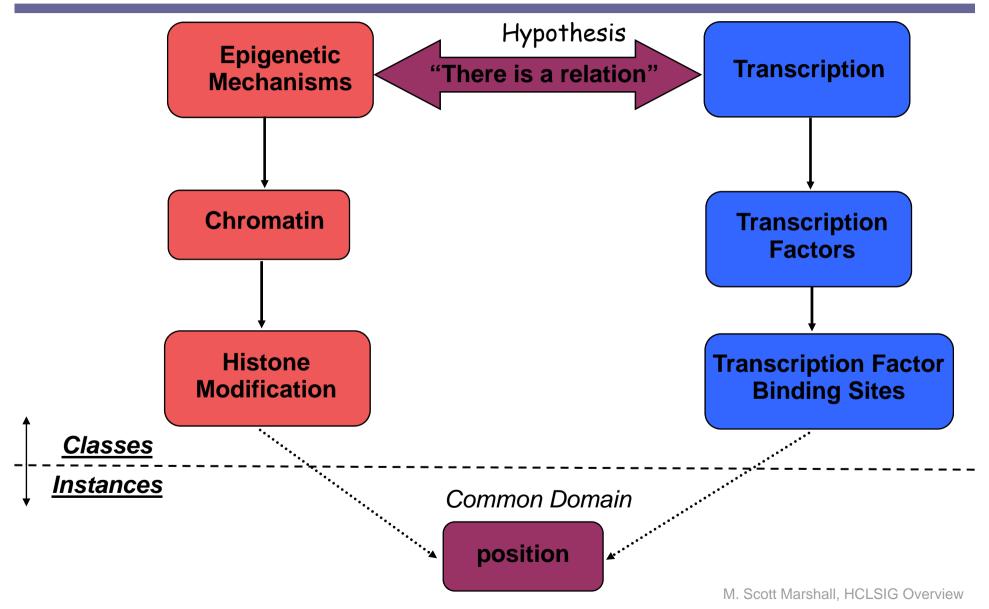
What is the Mission of HCLS IG?

 The mission of the Semantic Web Health Care and Life Sciences Interest Group, part of the Semantic Web Activity, is to develop, advocate, and support the use of Semantic Web technologies for biological science, translational medicine and health care. These domains stand to gain tremendous benefit by adoption of Semantic Web technologies, as they depend on the interoperability of information from many domains and processes for efficient decision support.

Vision: Concept-based interfaces

- The scientist should be able to work in terms of commonly used concepts.
- The scientist should be able to work in terms of personal concepts and hypotheses.
 - Not be forced to map concepts to the terms that have been chosen for a given application by the application builder.

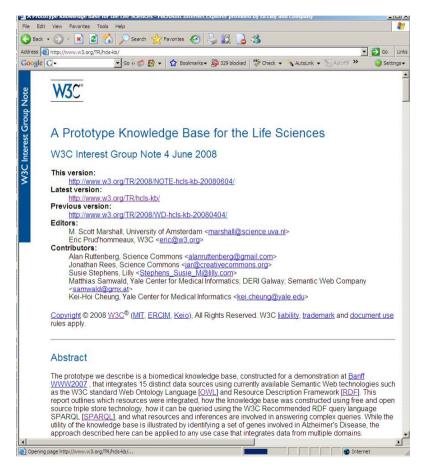
Interface Sketch: Finding a basis for relation



What does the group do?

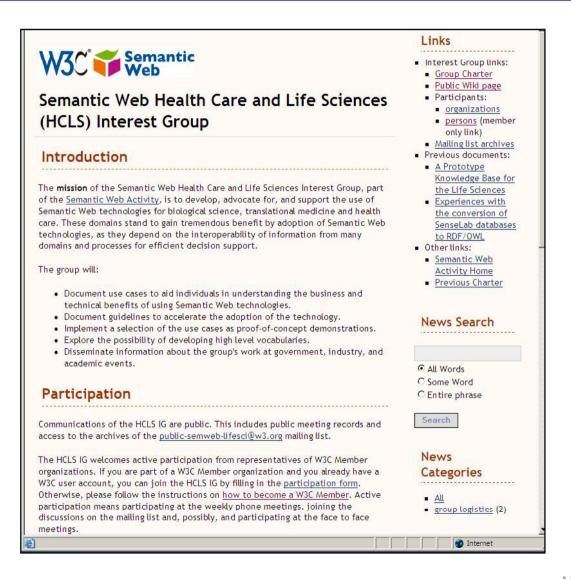
- Document use cases to aid individuals in understanding the business and technical benefits of using Semantic Web technologies
- Document guidelines to accelerate the adoption of the technology
- Implement a selection of the use cases as proof-of-concept demonstrations
- Explore the possibility of **developing** high-level **vocabularies**
- Disseminate information about the group's work at government, industry, and academic events

HCLS IG Documents

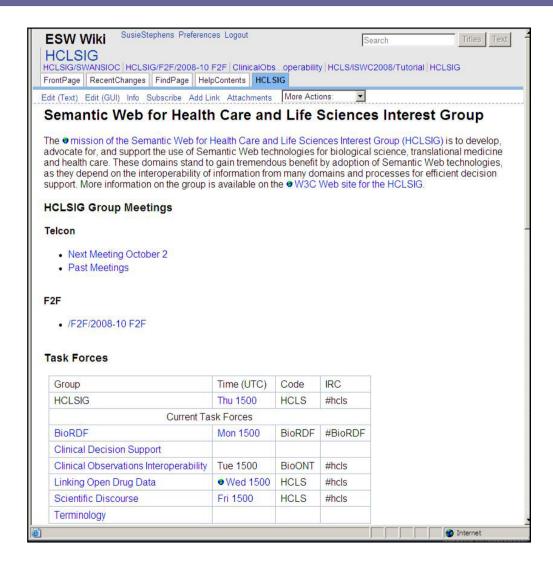




HCLS Web Site



HCLS Wiki



Motivation for New Members

- Benefits to getting involved include:
 - early access to use cases and best practice
 - influence standard recommendations
 - cost effective exploration of new technology through collaboration
 - network with others working with Semantic Web
- Get involved by contacting the chairs:
 - team-hcls-chairs@w3.org

Meetings

- Formal F2F, January 2006, Cambridge
- Formal F2F, October 2006, Amsterdam
- Workshop, ISWC November 2006, Banff
- Informal F2F (Demo) 3 x March/April, 2007, Cambridge
- Workshop, WWW 2007, May 2007, Banff
- Informal F2F (URI), July 2007, Cambridge
- Formal F2F, November 2007, Cambridge
- Formal F2F, October 2008, Mandelieu

(You are HERE)

Publications

Submitted:

Life Sciences on the Semantic Web: The Neurocommons and Beyond

Alan Ruttenberg, Jonathan Rees, Matthias Samwald, M. Scott Marshall

In Print:

- Advancing translational research with the Semantic Web
 Alan Ruttenberg, Tim Clark, William Bug, Matthias Samwald, Olivier Bodenreider,
 Helen Chen, Donald Doherty, Kerstin Forsberg, Yong Gao, Vipul Kashyap, June
 Kinoshita, Joanne Luciano, M Scott Marshall, Chimezie Ogbuji, Jonathan Rees, Susie
 Stephens, Gwendolyn T Wong, Elizabeth Wu, Davide Zaccagnini, Tonya
 Hongsermeier, Eric Neumann, Ivan Herman and Kei-Hoi Cheung,
 BMC Bioinformatics 2007, 8(Suppl 3):S2
- An Ontology-based approach for Data Integration An Application in Biomedical Research
 Vipul Kashyap, Kei-Hoi Cheung, Don Doherty, Matthias Samwald, M. Scott Marshall, Joanne Luciano, Susie Stephens, Ivan Hermanand Ray Hookway, Book Chapter in Cardoso, J., Hepp, M., Lytras, M. (Eds.) "Real-world Applications of Semantic Web Technology and Ontologies", Springer 2007.

 M. Scott Marshall, HCLSIG Overview

Presentations -> 2007

- WWW2007 Demo
- ISMB 2007 Demo
- ISMB BioOntology SIG Poster 2007

NIH Blueprint NIF Workshop

- Society for Neuroscience Poster Nov 2007
- Selection of presentation venues of members showing HCLS work
 - **Bridging Pharma and IT Drug Discovery Technology of Innovative Therapeutics 1st European Semantic Web Conference** Bio-IT World **Norwegian Semantic Web Day** InfoTech Pharma **Modern Drug Discovery and Development Summit Massachusetts Biotechnology Panel** eScience Institute; RDF, Ontologies and Meta-Data Workshop **Virginia Biotechnology Summit Systems Biology Semantic Web Gathering** Allen Institute for Brain Sciences Informatics and Interactomes in Huntington's Disease **Ontology for Biomedical Informatics Workshop Clinical Trial Ontology Workshop Jackson Laboratories Pubmed Plus**

Presentations in 2008

- Asian Semantic Web Conference, HCLS Workshop
- ISWC 2008 Semantic Web for Health Care and Life Sciences Tutorial
- SWAT4LS 2008 Workshop

Reasoning Web Summer School 2008

- WWW2008 Workshop
- ISMB 2008 Demo
- Selection of presentation venues of members showing HCLS work:
 - Drug Discovery Technology of Innovative Therapeutics
 - European Semantic Web Conference
 - □ Bio-IT World
 - □ InfoTech Pharma 2008

Linked Data Principles



- 1. Use URIs as names for things.
- 2. Use HTTP URIs so that people can look up those names.
- 3. When someone looks up a URI, provide useful RDF information.
- 4. Include RDF statements that link to other URIs so that they can discover related things.
 - Tim Berners-Lee 2007
 - http://www.w3.org/DesignIssues/LinkedData.html

Task Forces

- BioRDF integrated neuroscience knowledge base
 - Task lead Kei Cheung
- Clinical Decision Support expression and use of guidelines
 - Task lead TBD
- Clinical Observations Interoperability patient recruitment in trials
 - Task lead Vipul Kashyap
- Linking Open Drug Data aggregation of Web-based drug data
 - Task lead Chris Bizer
- Scientific Discourse building communities through networking
 - Task leads Tim Clark, John Breslin
- Terminology Semantic Web representation of existing resources
 - Task lead John Madden

BioRDF Task

BioRDF participants

- The task is lead by Kei Cheung (Yale)
- Has approximately 20 participants

BioRDF activities include:

- Explore the effectiveness of current tools for making data available as RDF/OWL
- Build a life sciences demo that spans from bench to bedside using RDF/OWL to help scientist better understand the value of the Semantic Web
- Document our finding to help accelerate the adoption of the Semantic Web by others

BioRDF Publications

- A Prototype Knowledge Base for the Life Sciences http://www.w3.org/TR/hcls-kb/
- Experience with the Conversion of SenseLab Databases to RDF/OWL http://www.w3.org/TR/hcls-senselab/
- More Information on the group is available at
 - http://esw.w3.org/topic/HCLSIG BioRDF Subgroup

BioRDF - Answering Questions

- Goals: Get answers to questions posed to a body of collective knowledge in an effective way
- Knowledge used: Publicly available databases, and text mining
- Strategy: Integrate knowledge using careful modeling, exploiting Semantic Web standards and technologies

COI Goals

- Demonstrate the value of Semantic Web specifications to bridge the divide between clinical practice and clinical research
- Develop collaboratively a proof of concept
- Get a variety of stakeholders to participate as a prelude to acceptance and adoption

COI Use Cases

- Wide variety of use cases were investigated and discussed
 - Patient Recruitment
 - Adverse Event Detection
 - Tracking Patient Through a Clinical Trial
- Decision: Focus on Patient Recruitment

Linking Open Drug Data (LODD)

Primary Objectives

- 1. Survey publicly available data sets about drugs
- 2. Explore how the data sets could be linked together
- 3. Explore interesting questions that could be answered if the data sets are linked

LODD Participants

- Kristin Tolle (Microsoft)
- Eric Prud'hommeaux (W3C)
- Don Doherty (Brainstage)
- Susie Stephens (Lilly)
- Bosse Anderssen (AZ)
- M. Scott Marshall (University of Amsterdam)
- Chris Bizer (Freie Universitat Berlin)
- Glen Newton (National Research Council Canada)
- Michel Dumontier (Carleton University)
- TN Bhat (NIST)

Data Sets for the LODD Task

- ClinicalTrials.gov
- LillyTrials
- Orange Book
- RxNorm
- Drug Bank
- Diseasome
- Drugome

- DrugDB
- Drug Ontology
- PubChem
- chEBI
- National Drug Code
- ChemBlast
- KEGG Drug
- VA NDF-RT

Plus ca change..

- ..plus c'est la meme chose.
- Models change:
 - Epigenetics: The code isn't only in the DNA
 - Polypharmacology: We should aim for ligands with multiple targets.
- Terminologies change:
 - In Personal Medical Records:
 - Per user: patient, clinician

Why Synonym Services?

Example: CBP has now officially been renamed KAT3A p300 is now KAT3B

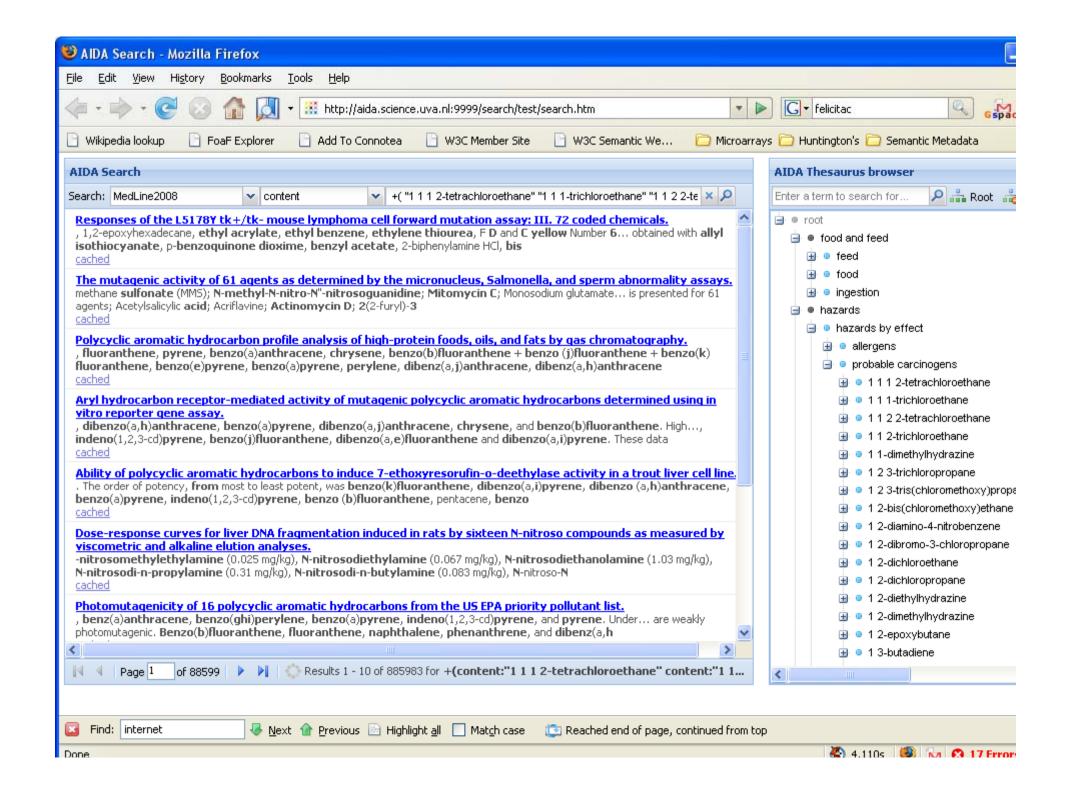
See new nomenclature rules for chromatin remodellers: Cell, Vol 131, 633-636, 16 November 2007

Some common themes

- Mapping from Terminology to Model
- Mapping from Terminology to Terminology
- Mapping from Model to Model
- ...

Raises some questions:

- What to use for the mapping representation?
- How do we create the mapping/alignment?



The Question Answering Process

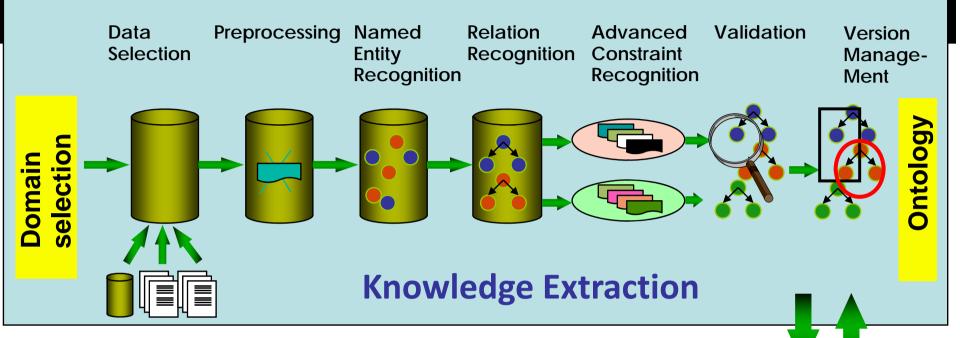
- Start with terminology
- Translate into query language
- Query is compiled into a representation that can unify with the data representation

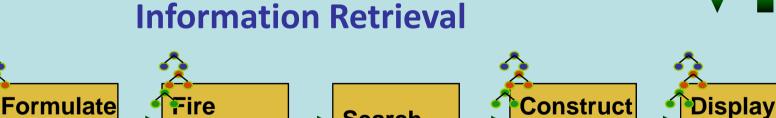
Add new knowledge to Knowledge Base



Adaptive Information Disclosure







Search

query query **User support:** Query **Alternatives Expansion**

Disambiguation

Filtering Relevancescore

answer

Link to Concept tree

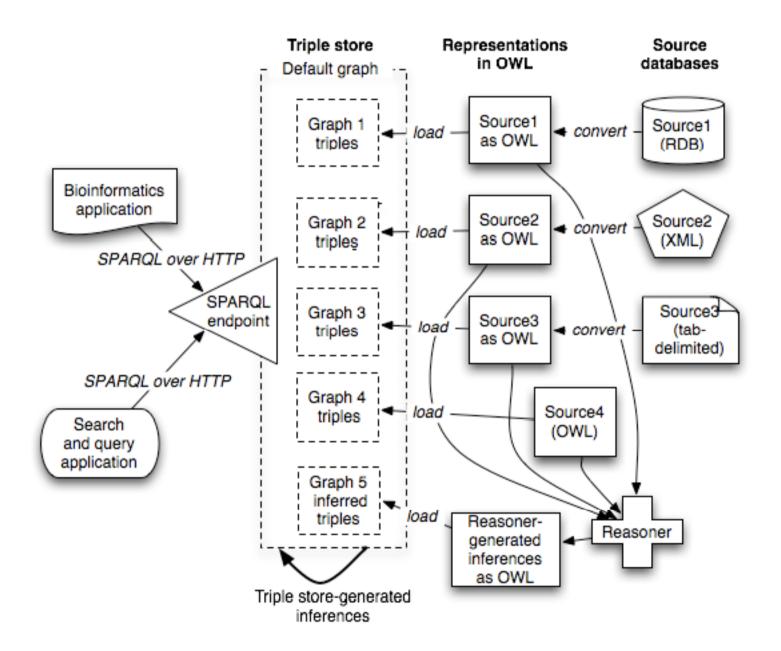
results

Scientific Question

"find me genes involved in signal transduction that are related to pyramidal neurons"

A SPARQL query for processes involved in pyramidal neurons

```
prefix go: <a href="http://purl.org/obo/owl/GO#">http://purl.org/obo/owl/GO#>
prefix rdfs: <a href="http://www.w3.org/2000/01/rdf-schema#">http://www.w3.org/2000/01/rdf-schema#>
prefix owl: <a href="http://www.w3.org/2002/07/owl#">prefix owl: <a href="http://www.w3.org/2002/07/owl#">http://www.w3.org/2002/07/owl#</a>
prefix mesh: <a href="http://purl.org/commons/record/mesh/">http://purl.org/commons/record/mesh/>
prefix sc: <a href="http://purl.org/science/owl/sciencecommons/">http://purl.org/science/owl/sciencecommons/</a>>
                                                                                                                                                                                                      Mesh: Pyramidal Neurons
prefix ro: <a href="http://www.obofoundry.org/ro/ro.owl#>"> prefix ro: <a href="http://www.obofoundry.org/ro/ro.owl#">http://www.obofoundry.org/ro/ro.owl#></a>
select ?genename ?processname
where
{ graph <http://purl.org/commons/bels/pubmesh>
    { ?paper ?p mesh:D017966 .
       ?article sc:identified by pmid?paper.
                                                                                                                                                                                                      Pubmed: Journal Articles
        ?gene sc:describes_gene_or_gene_product_mentioned by ?article.
  graph <a href="mailto:rg/commons/hcls/goa">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.
                                                                                                                                                                                                     Entrez Gene: Genes
       ?res2 owl:someValuesFrom ?process.
   graph <a href="mailto:ref">ref</a> / ref</a> /
    {{?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.
                                                                                                                                                                                                     GO: Signal Transduction
  graph <a href="mailto:re/commons/hcls/gene">http://purl.org/commons/hcls/gene</a>
   { ?gene rdfs:label ?genename }
  graph <a href="mailto:ref">ref</a> / purl.org/commons/hcls/20070416>
   { ?process rdfs:label ?processname}
                                                                                                                   Inference required
```



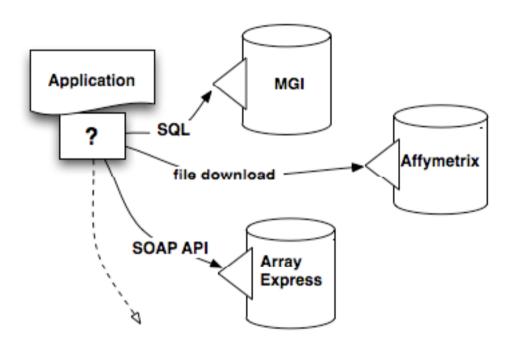
Next Phase

- More Use Cases
- More Users (of Knowledge Base, our approaches)
- Store Knowledge in Rules
- Easier Interface
 - Convenient term lookup from Knowledge Base
 - Automatic translation of terms in Interface
 - Saving and finding similar queries (collaborative filtering)
- Distributed Knowledge (Queries)
- Reasoning across large Knowledge Bases

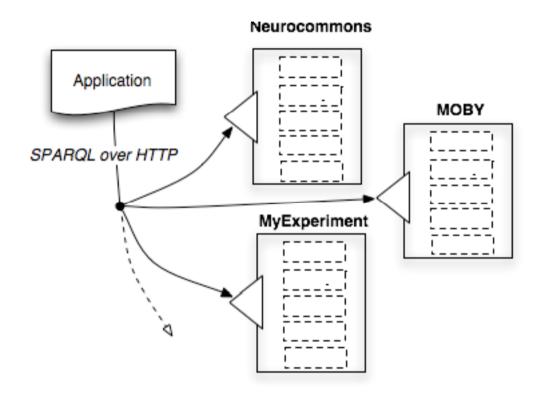
SPARQL-DL

```
1. SELECT ?gene_name ?process_name
2. WHERE
3. { PropertyValue(?pubmed_record, ?related, mesh:D017966)
4. PropertyValue(?article, sc:identified_by_pmid , ?pubmed_record)
5. PropertyValue(?gene_record, sc:describes_gene_or_gene_product_mentioned_by, ?article)
6. SubClassOf(?protein, some(ro:has_function, some(ro:realized_as, ?process)))
7. SubClassOf(?process, or(go:GO_0007166, some(ro:part_of, go:GO_0007166))
8. SubClassOf(?protein, some(sc:is_protein_gene_product_of_dna_described_by, {?gene_record}))
9. Annotation(?gene_record, rdfs:label, ?gene_name)
10. Annotation(?process, rdfs:label, ?process_name)
11. }
```

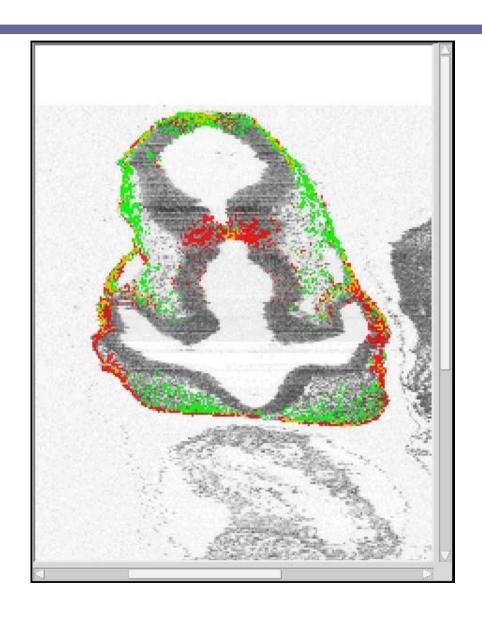
Distributed Query - Before



Distributed Query - After



Someday, we should be able to find this as evidence for a fact in the Knowledge Base



The End

"Science is built up of facts, as a house is built of stones; but an accumulation of facts is no more a science than a heap of stones is a house."

Henri Poincaré,Science and Hypothesis, 1905