

# The Semantic Web: A Medical Perspective

Patrick Sinclair  
University of Southampton

[pass@ecs.soton.ac.uk](mailto:pass@ecs.soton.ac.uk)  
<http://www.metade.org>

# metade / half

Português

English

Before I start, I just want to explain that I'm half portuguese, half english.

\*

I'll be doing the talk in English, mas tambem falo portugues.  
Se houver alguma coisa que nao entenderam, podem perguntar em Portugues.

# metade / half

Bacalhau à Braz



<http://flickr.com/photos/davipt/367229297/>

Fish and chips



<http://flickr.com/photos/samsmith/35825678/>

Before I start, I just want to explain that I'm half portuguese, half english.

\*

I'll be doing the talk in English, mas tambem falo portugues.  
Se houver alguma coisa que nao entenderam, podem perguntar em  
Portugues.



I also want to say that I'm a computer scientist.

All my medical knowledge comes from watching TV!

So apologies if I get any of the medicine wrong.

# The Semantic Web: A Medical Perspective

The Semantic web is a recent development to express information on the web in a way that computers can process it more effectively.

So what's interesting about the Semantic Web and medicine?

What are the issues in the medical world that could be helped with the Semantic web?

# data, data, data...

In this talk we'll be looking at issues regarding data.

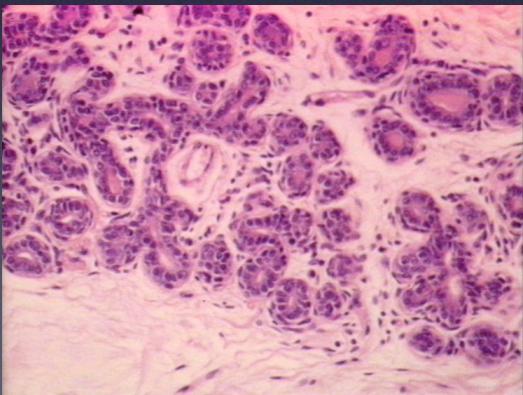
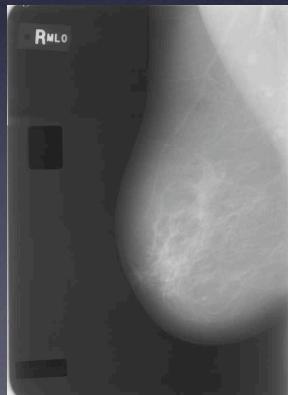
# Quantities of Data

Obviously there's a lot of data!

Scalable - across web?

# Different types of data

- patient records
- clinical records
- lab results



There are lots of different types of data.

In particular, there's a lot of interesting multimedia data:

images - like x-rays, mris, mammograms

There's lots of potential for videos too.



# Limited Resources

<http://flickr.com/photos/monkeyc/112342184/>

There is always a lack of resources.

So it's important that we can deal with all this data effectively.

We need to make the most of the medical workers' and researchers' valuable time.



## More effective tools

<http://flickr.com/photos/waynemah/77745389/>

We could do with better tools and techniques to manage all of this data.

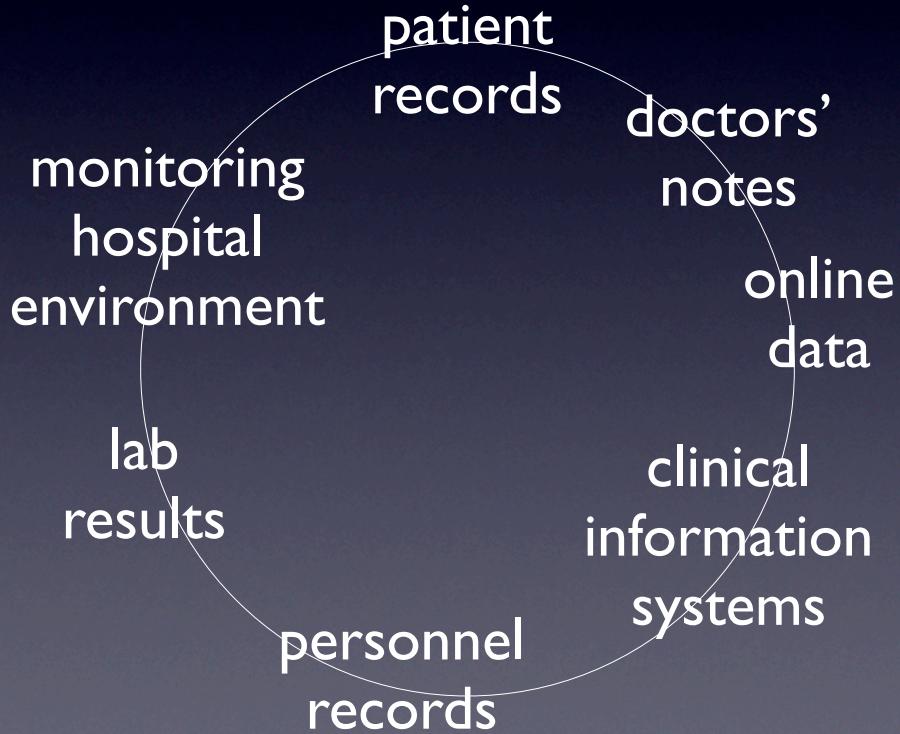
We need to make computers and machines do more hard work, so that people can focus on the important details.

We might use automatic techniques to help diagnose cases.

# Sharing Data

An important issue is being able to share data.

# Why is sharing data a problem?



Why is sharing data a problem?

There's so many different types of data used in the practice of medicine.

Some are fairly obvious - you obviously need to have patient records and doctors notes and lab results.

But often these are distributed across departments.

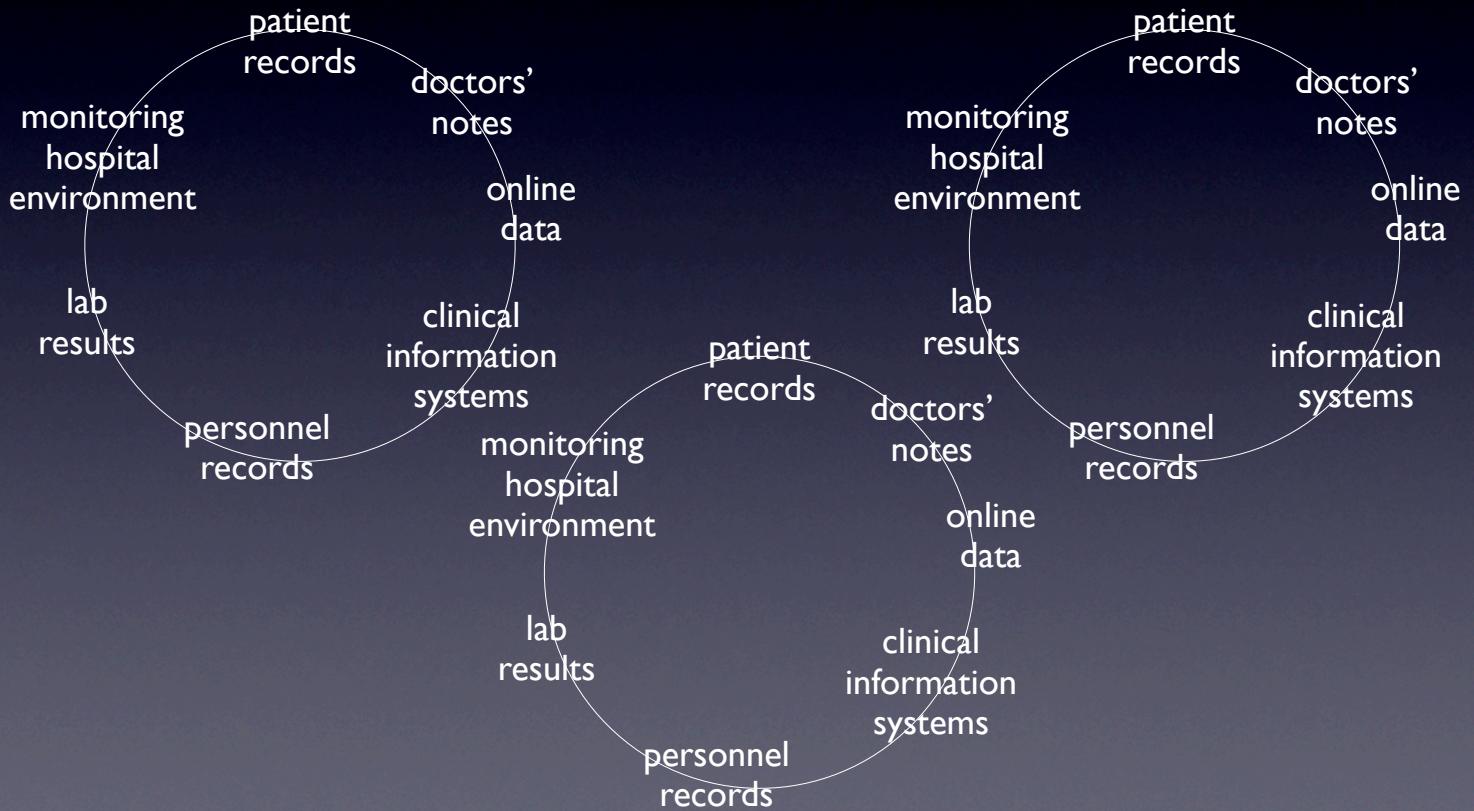
Specialists, for example a radiologist and a pathologist, will meet, share relevant data, but then take their data back with them.

So each specialist won't be able to consult crucial information after the meeting!

To detect infectious, gastrointestinal or respiratory illnesses you might need to monitor emergency room visits, water contamination and air quality.

There might also be helpful data online, such as pollens and allergens.

# Sharing across different institutions



And obviously the problem gets a lot worse when you need to share across different institutions!

# Semantic Interoperability

## Technical

- System A talks XYZ and System B talks ZYX
- Internal system unavailable to the outside world

## Social

- radiographers and pathologists use different terms
- issues of trust?
- issues of money?
- licensing?

There are lots of issues when trying to share data.

There are technical problems.

Different computer systems will have different ways to represent information.  
So you need to get them talking to each other.

It might be that a system is unavailable to the outside world.

But there are also social problems.

Different specialists use different terms to describe the same thing.

There are issues of trust - you might want to let patients access their records but how do you know that it's the right person?

And then there's cost in linking the different systems together.  
It might also be that organisations decide to charge for their data, or have restrictions enforced on the use of the data.



# Silos of Data

<http://flickr.com/photos/scriptingnews/140068566/>

The result is isolated silos of data.

Data gets placed in a specific system or place and it's hard to get at.

“Curing and preventing  
diseases requires a  
synthesis of understanding  
across disciplines”

*Advancing translational research with the Semantic Web*  
Ruttenberg et al., 2007

And this is terrible!  
Because to research and cure diseases you need understanding  
across different specialities.  
Without access to the necessary data, this will not happen.

# The Study of Neurodegenerative Diseases

- psychiatry
- neurology
- microscopic anatomy
- neuronal physiology
- biochemistry
- genetics
- molecular biology
- bioinformatics

<http://www.flickr.com/photos/amonerris/418347301/>

For example, the study of neurodegenerative diseases spans several disciplines.

# Alzheimer's Disease

Alzheimers' is a terrible disease.

There is no common agreement on exactly how it is caused, where to intervene to treat it, or to prevent it.

<http://www.flickr.com/photos/kk/64242903/>  
<http://www.flickr.com/photos/annedehaas/138509062/>

<http://www.flickr.com/photos/amonerris/418347301/>

<http://www.flickr.com/photos/annedehaas/138509062/>  
<http://www.flickr.com/photos/tags/alzheimer/interesting/>

# Study on Alzheimer's Disease

Lesne et al. 2006

- mouse genetics
- cell biology
- animal neuropsychology
- protein biochemistry
- neuropathology
- and other areas

A recent study published on Alzheimers' disease illustrates the typical situation.

The hypothesis integrated research from a number of different disciplines.

Researchers have come to the conclusion that in many cases no single lab, library or data repository contains the information they need.

<http://flickr.com/photos/rick-in-rio/399225005/in/photostream/> - mouse

<http://flickr.com/photos/shammonds/295120456/> - cell biology

<http://flickr.com/photos/neurollero/58139828/> - neuropsychology

<http://flickr.com/photos/viknanda/403371973/> - protein biochemistry

<http://flickr.com/photos/elixir/245997787/> - neuropathology



So it's crucial to be able  
to share information!

<http://flickr.com/photos/gaijinbiker/46032173/>

So it's crucial to be able to share information.

And this is where the Semantic Web comes in.

The goal of the semantic web is to improve the sharing of data on the web.

# World Wide Web

But before we look at the Semantic Web in detail, let's talk about the World Wide Web...

# The web is awesome!

The web is awesome!

It's completely changed how we access, publish and share information.

<http://www.yahoo.com/health>  
1996

And it's incredible to see how far it's come in such a short time!

- \* This is what Yahoo Health looked like in 1996
- \* And this is what they have now.

[Migraine Resource Center]



## Top:Health

[Options](#)

Search all of Yahoo  Search only in Health

- [Reuters Health News](#)
- [Today's News: Health & Wellness](#)
- [Indices](#) (23)
- [Sub Category Listing](#)

- 
- [Alternative Medicine](#) (297) NEW!
  - [Career and Employment@](#)
  - [Companies@](#)
  - [Conferences](#) (14)
  - [Death and Dying@](#)
  - [Dentistry](#) (166) NEW!
  - [Disabilities@](#)
  - [Diseases and Conditions](#) (1902) NEW!
  - [Education](#) (43)
  - [Emergency Services](#) (81) NEW!
  - [Environmental Health](#) (79) NEW!
  - [Fitness](#) (79)
  - [General Health](#) (86)

And it's incredible to see how far it's come in such a short time!

- \* This is what Yahoo Health looked like in 1996
- \* And this is what they have now.



[Sign In](#)  
New User? Sign Up

[Home](#) | [Healthy Living](#) ▾ | [Diseases & Conditions](#) ▾ | [Health News](#) | [Groups & Boards](#) | [Drug Guide](#)

Search Yahoo! Health for:

[SEARCH](#)

ADVER

## What Goes In



What you put in plays a role in every part of your day, as well as your long-term health. Find out how to eat smarter and have some fun doing it.

- A Matter of Fat
- Instant Energy
- The Longevity Diet
- Eat Right Without Thinking
- More Food, Less Fat
- Eat This, Not That Summer Edition
- Ultimate Smoothie Selector

[More About Nutrition](#)

Sponsored by:



[PAUSE](#)

## HEALTH EXPERT ADVICE [» See All Experts](#)

Use the arrows below to scroll through all of our experts.



**Sex Camp**

By Dr. Laura Berman

## IN THE SPOTLIGHT



**Asthma**

Find out if you're related, rather than

## TRANSFORMING



**Choosing Surgery**

Sherry lost 17% of her body weight after surgery and now she's a

[PLAY](#)

And it's incredible to see how far it's come in such a short time!

\* This is what Yahoo Health looked like in 1996

\* And this is what they have now.



# The Web is designed for people!

And it's success, from both a social as well as a technical level, is that it is designed for people.

Most information on the web is intended for people to read directly.

Information is connected – linked – but it is compartmentalised on different sites.

And the web is page granular - everything is based around the concept of individual web pages.

The web is a global, universal information space for documents.

# Have to piece together search results, web pages and information

Because web pages are intended for people to read, it's hard for a computer to process the information.

So this means that when you look for something online, you end up doing a lot of hard work!

You have to piece together the search results, and keep a track of the web pages you visited.

# Allen Brain Atlas

<http://www.brainatlas.org/aba/>

I want to show you an example of what I mean.

The Allen Brain Atlas is a freely available scientific resource developed by the Allen Institute.

It provides maps of the expression of approximately 20,000 genes in the mouse brain.

It's a remarkable scientific achievement.

They've cut mouse brains into thin slices and stained the slices to test which gene expressions are present.

Select up to three gene series to view in detail. Click a thumbnail to open image in resizable viewer.

[View Detailed Images](#)

[Close Window](#)

### Kcnip3 - Kv channel interacting protein 3, calsenilin - sagittal -- (17 images)

[Series Info](#)

Hide Thumbnails

[Show thumbnails in expression view](#)

Select This Series to View in Detail



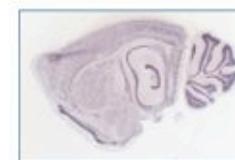
Kcnip3\_10  
Position: 250  
[Show expression view](#)



Kcnip3\_18  
Position: 450  
[Show expression view](#)



Kcnip3\_34  
Position: 850  
[Show expression view](#)



Kcnip3\_42  
Position: 1050  
[Show expression view](#)



Kcnip3\_50  
Position: 1250  
[Show expression view](#)



Kcnip3\_58  
Position: 1450  
[Show expression view](#)



Kcnip3\_66  
Position: 1650  
[Show expression view](#)



Kcnip3\_74  
Position: 1850  
[Show expression view](#)



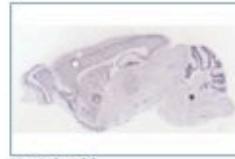
Kcnip3\_82  
Position: 2050  
[Show expression view](#)



Kcnip3\_90  
Position: 2250  
[Show expression view](#)



Kcnip3\_98  
Position: 2450  
[Show expression view](#)



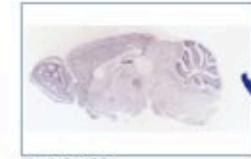
Kcnip3\_106  
Position: 2650  
[Show expression view](#)



Kcnip3\_114  
Position: 2850  
[Show expression view](#)



Kcnip3\_122  
Position: 3050  
[Show expression view](#)



Kcnip3\_138  
Position: 3450  
[Show expression view](#)



[Done](#)



So on the site you get access to information on these 20,000 genes.  
There are 400,000 images at high resolution.

So you can search for genes, and view the images,

\*

We can also view information about a given gene.

It's a fantastic resource.

But say we want to take this data and use it in another application or link it to another data set.

We're forced to manually copy/paste the data ourselves, reformat it.

It's a lot of hard work.

And it's the kind of work the computer should do for us.

But the valuable data is trapped in the HTML markup that represents the page.

When you view the source of the page \*\*, all the computer can process is how to make a table \* with some textual stuff inside!

It doesn't know what the stuff is.

Select up to three gene series to view in detail. Click a thumbnail to open image in resizable viewer.

[View Detailed Images](#)

[Close Window](#)

Kcnip3 - Kv channel interacting protein 3, calsenilin - sagittal - (17 images)

[Series Info](#)

[▼ Hide Thumbnails](#)

Allen Brain Atlas ...eries Information

[Show thumbnails in expression view](#)

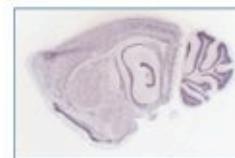
### IMAGE SERIES INFORMATION

Gene:	Kv channel interacting protein 3, calsenilin (Kcnip3)
Images:	17
Probe Info:	Probe
Plane of Section:	sagittal
Probe Orientation:	antisense
Organism:	Mus musculus
Strain:	C56/B57
Age:	56 days
Sex:	male

[Close](#)



34  
850  
expression view



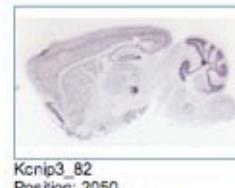
Kcnip3\_42  
Position: 1050  
[Show expression view](#)



Kcnip3\_50  
Position: 1250  
[Show expression view](#)



74  
1850  
expression view



Kcnip3\_82  
Position: 2050  
[Show expression view](#)



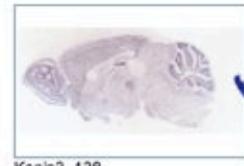
Kcnip3\_90  
Position: 2250  
[Show expression view](#)



Kcnip3\_98  
Position: 2450  
[Show expression view](#)



Kcnip3\_106  
Position: 2650  
[Show expression view](#)



Kcnip3\_114  
Position: 2850  
[Show expression view](#)



Kcnip3\_122  
Position: 3050  
[Show expression view](#)



Kcnip3\_138  
Position: 3450  
[Show expression view](#)

Done



So on the site you get access to information on these 20,000 genes.  
There are 400,000 images at high resolution.

So you can search for genes, and view the images,

\*

We can also view information about a given gene.

It's a fantastic resource.

But say we want to take this data and use it in another application or link it to another data set.

We're forced to manually copy/paste the data ourselves, reformat it.

It's a lot of hard work.

And it's the kind of work the computer should do for us.

But the valuable data is trapped in the HTML markup that represents the page.

When you view the source of the page \*\*, all the computer can process is how to make a table \* with some textual stuff inside!

It doesn't know what the stuff is.

Select up to three gene series to view in detail. Click a thumbnail to open image in resizable viewer.

[View Detailed Images](#)

[Close Window](#)

Kcnip3 - Kv channel interacting protein 3, calsenilin - sagittal -- (17 images)

[Series Info](#)

[▼ Hide Thumbnails](#)

Allen Brain Atlas ...eries Information

[Show thumbnails in expression view](#)

### IMAGE SERIES INFORMATION

Gene:	Kv channel interacting protein 3, calsenilin (Kcnip3)
Images:	17
Probe Info:	Probe
Plane of Section:	sagittal
Probe Orientation:	antisense
Organism:	Mus musculus
Strain:	C56/B57
Age:	56 days
Sex:	male

[Close](#)

[Reload](#)

[View Source](#)

[Save Page As...](#)

[Print Page...](#)

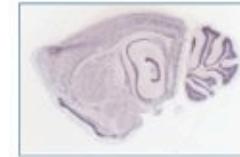
Kcnip3\_98  
Position: 2450  
[Show expression view](#)

Kcnip3\_106  
Position: 2650  
[Show expression view](#)

Kcnip3\_114  
Position: 2850  
[Show expression view](#)

Kcnip3\_122  
Position: 3050  
[Show expression view](#)

Kcnip3\_138  
Position: 3450  
[Show expression view](#)



So on the site you get access to information on these 20,000 genes.  
There are 400,000 images at high resolution.

So you can search for genes, and view the images,

\*

We can also view information about a given gene.

It's a fantastic resource.

But say we want to take this data and use it in another application or link it to another data set.

We're forced to manually copy/paste the data ourselves, reformat it.

It's a lot of hard work.

And it's the kind of work the computer should do for us.

But the valuable data is trapped in the HTML markup that represents the page.

When you view the source of the page \*\*, all the computer can process is how to make a table \* with some textual stuff inside!

It doesn't know what the stuff is.

ALLEN INSTITUTE

Select up to three genes

View Detailed Images

Hide Thumbnails

IMAGE

Gene: Kcnip3\_98

Images: Position: 2450

Probe Info: Show expression

Plane of Section: sagittal

Probe Orientation: antisense

Organism: Mus musculus

Strain: C56/B57

Age: 56 days

Sex: male

Close

Done

```

<table cellpadding="0" cellspacing="0" border="0" class="tblState1" style="" id="tblInfo">
    <tr valign="top">
        <td class="seriesInfLt">Gene:</td>
        <td class="seriesInfRt">Kv channel interacting protein 3, calsenilin (Kcnip3)</td>
    </tr>
    <tr valign="top">
        <td class="seriesInfLt">Images:</td>
        <td class="seriesInfRt">17</td>
    </tr>
    <tr valign="top">
        <td class="seriesInfLt">Probe Info:</td>
        <td class="seriesInfRt">
            <a href="/viewProbe.do?imageSeriesId=68862060&probeName=RP_050301_01_D12" onclick="window.open(this.href,'probeInfo_RP_050301_01_D12','scrollbars=yes,resizable=no,width=500,height=560,dependent=yes');return false;" onfocus="this.blur()" title="View Riboprobe Data">Probe</a></td>
        </tr>
        <tr valign="top">
            <td class="seriesInfLt">Plane of Section:</td>
            <td class="seriesInfRt">sagittal</td>
        </tr>
        <tr valign="top">
            <td class="seriesInfLt">Probe Orientation:</td>
            <td class="seriesInfRt">antisense</td>
        </tr>
        <tr valign="top">
            <td class="seriesInfLt">Organism:</td>
            <td class="seriesInfRt">Mus musculus</td>
        </tr>
        <tr valign="top">
            <td class="seriesInfLt">Strain:</td>
            <td class="seriesInfRt">C56/B57</td>
        </tr>
        <tr valign="top">
            <td class="seriesInfLt">Age:</td>
            <td class="seriesInfRt">56 days</td>
        </tr>
        <tr valign="top">
            <td class="seriesInfLt">Sex:</td>
            <td class="seriesInfRt">male</td>
        </tr>
    </table>

```

Series Info

0 1250 pression view

0 2250 pression view

38 3450 pression view

So on the site you get access to information on these 20,000 genes.  
 There are 400,000 images at high resolution.

So you can search for genes, and view the images,

\*

We can also view information about a given gene.

It's a fantastic resource.

But say we want to take this data and use it in another application or link it to another data set.

We're forced to manually copy/paste the data ourselves, reformat it.

It's a lot of hard work.

And it's the kind of work the computer should do for us.

But the valuable data is trapped in the HTML markup that represents the page.

When you view the source of the page \*\*, all the computer can process is how to make a table \* with some textual stuff inside!

It doesn't know what the stuff is.

ALLEN INSTITUT

Select up to three genes

View Detailed Images

Hide Thumbnails

IMAGE

Gene: Kcnip3\_98

Images: Position: 2450

Probe Info: Show expression

Plane of Section: sagittal

Probe Orientation: antisense

Organism: Mus musculus

Strain: C56/B57

Age: 56 days

Sex: male

Close

Kcnip3\_98  
Position: 2450  
Show expression

Done

[/viewProbe.do?imageSeriesId=68862060&probeName=RP\\_050301\\_01\\_D12" onclick="window.open\(this.href,'probeInfo\\_RP\\_050301\\_01\\_D12','scrollbars=yes,resizable=no,width=500,height=560,dependent=yes'\);return false;" onfocus="this.blur\(\);">Probe](#)

Series Info

0 1250 pression view

10 2250 pression view

38 3450 pression view

Table source code:

```
<table cellpadding="0" cellspacing="0" border="0" class="tblState1" style="" id="tblInfo">
<tr valign="top">
<td class="seriesInflT">Gene:</td>
<td class="seriesInfrT">Kv channel interacting protein 3, calsenilin (Kcnip3)</td>
</tr>
<tr valign="top">
<td class="seriesInflT">Images:</td>
<td class="seriesInfrT">17</td>
</tr>
<tr valign="top">
<td class="seriesInflT">Probe Info:</td>
<td class="seriesInfrT">
<a href="/viewProbe.do?imageSeriesId=68862060&probeName=RP_050301_01_D12" onclick="window.open(this.href,'probeInfo_RP_050301_01_D12','scrollbars=yes,resizable=no,width=500,height=560,dependent=yes');return false;" onfocus="this.blur();">Probe</a></td>
</tr>
<tr valign="top">
<td class="seriesInflT">Plane of Section:</td>
<td class="seriesInfrT">sagittal</td>
</tr>
<tr valign="top">
<td class="seriesInflT">Probe Orientation:</td>
<td class="seriesInfrT">antisense</td>
</tr>
<tr valign="top">
<td class="seriesInflT">Organism:</td>
<td class="seriesInfrT">Mus musculus</td>
</tr>
<tr valign="top">
<td class="seriesInflT">Strain:</td>
<td class="seriesInfrT">C56/B57</td>
</tr>
<tr valign="top">
<td class="seriesInflT">Age:</td>
<td class="seriesInfrT">56 days</td>
</tr>
<tr valign="top">
<td class="seriesInflT">Sex:</td>
<td class="seriesInfrT">male</td>
</tr>
</table>
```

So on the site you get access to information on these 20,000 genes.  
 There are 400,000 images at high resolution.

So you can search for genes, and view the images,  
 \*

We can also view information about a given gene.

It's a fantastic resource.

But say we want to take this data and use it in another application or link it to another data set.

We're forced to manually copy/paste the data ourselves, reformat it.  
 It's a lot of hard work.

And it's the kind of work the computer should do for us.  
 But the valuable data is trapped in the HTML markup that represents the page.

When you view the source of the page \*\*, all the computer can process is how to make a table \* with some textual stuff inside!

It doesn't know what the stuff is.

# Web Application Architecture



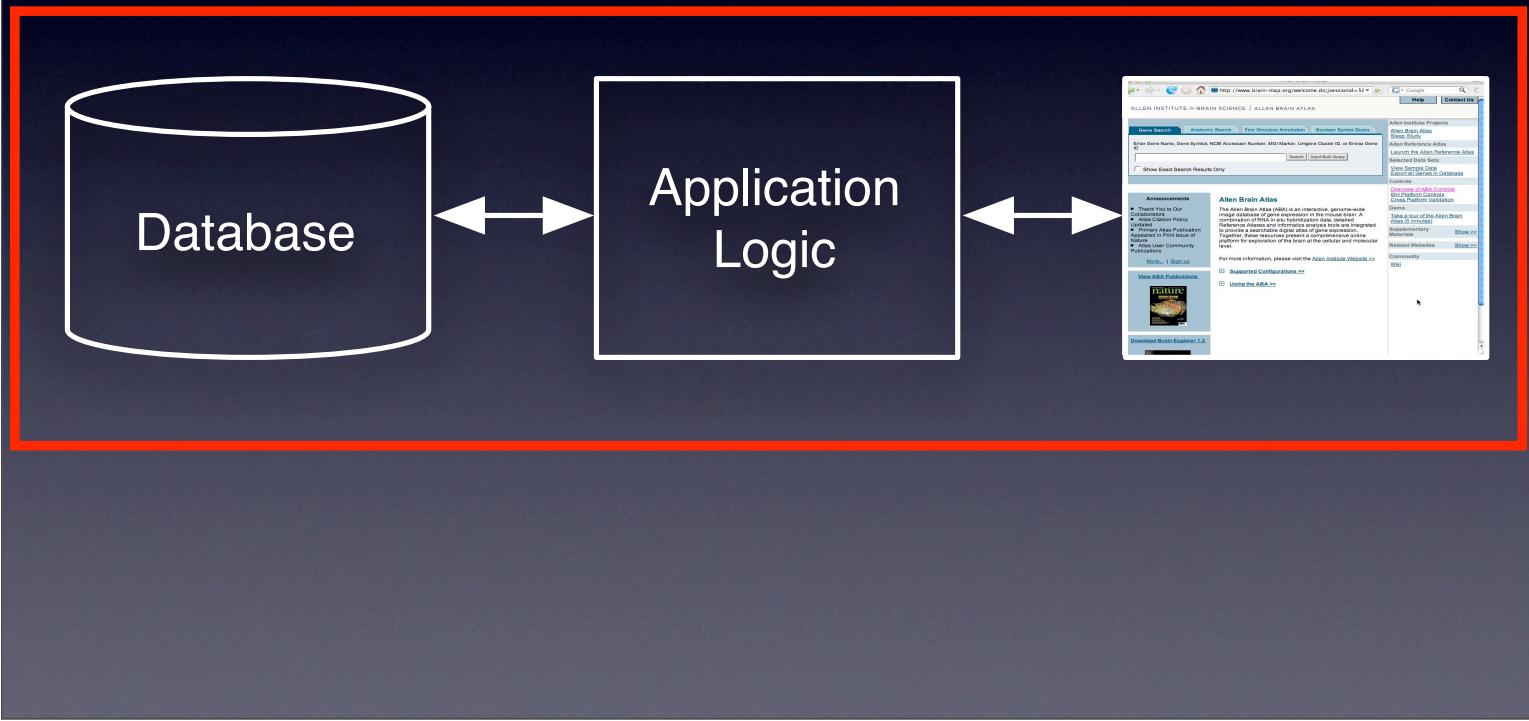
And you can't go behind the scenes either.

This is the typical web application architecture.  
The data is generally stored in a database.  
There's some application code that takes the data and renders it into web pages.  
Then there's the user with their browser looking at the site.

\*

But the data is trapped!  
You almost never get access to the database.  
And as I just showed, there's not much you can do to extract information in the browser.

# Web Application Architecture



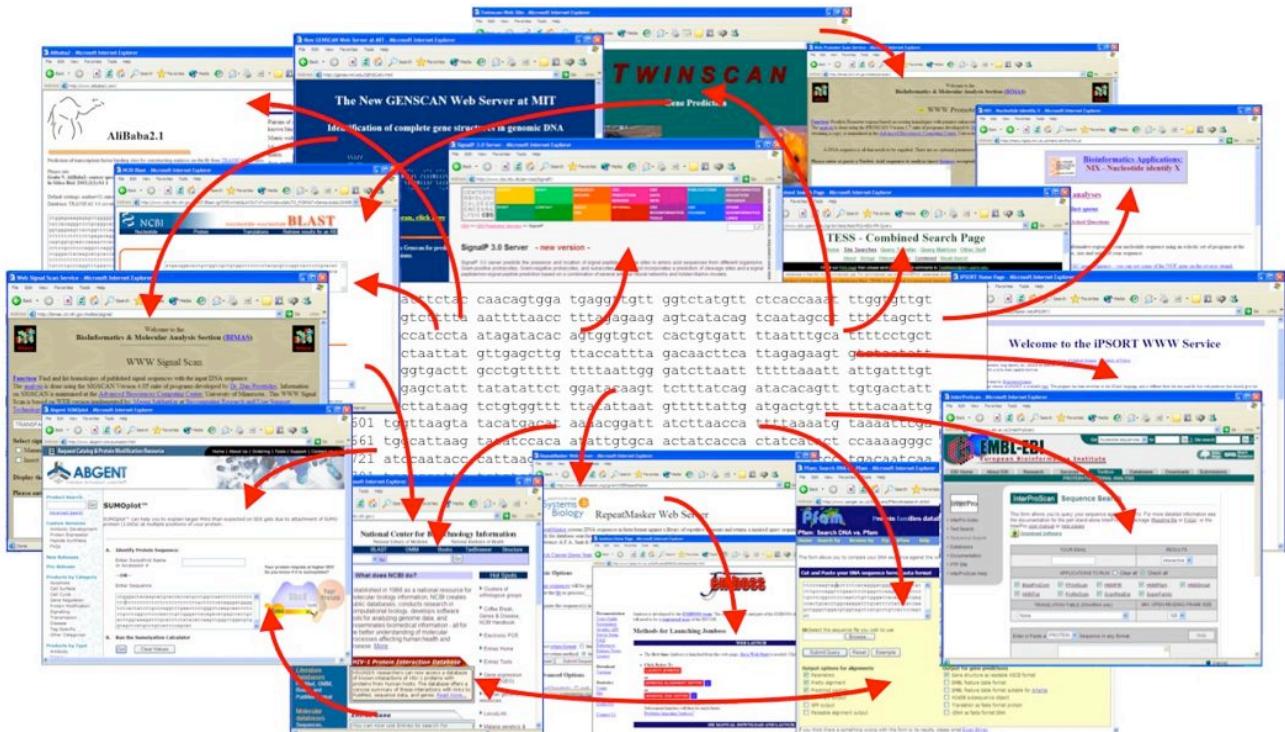
And you can't go behind the scenes either.

This is the typical web application architecture.  
The data is generally stored in a database.  
There's some application code that takes the data and renders it into web pages.  
Then there's the user with their browser looking at the site.

\*

But the data is trapped!  
You almost never get access to the database.  
And as I just showed, there's not much you can do to extract information in the browser.

# Need for Semantics



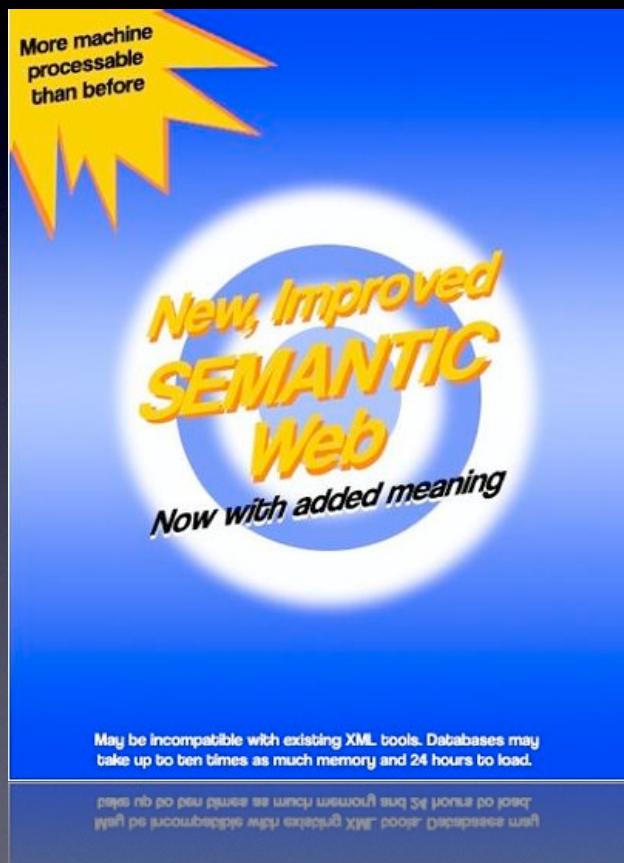
*Data Integration in myGrid with Taverna*  
Duncan Hull, University of Manchester, UK  
<http://www.cs.man.ac.uk/~hulld/>

So what happens is people end up managing all of the different data sources manually.

They are forced to copy and paste data from different data sources and application themselves.

So what we need is a way to describe the data so that our computers can do more of this hard work for us.

# The Semantic Web



<http://flickr.com/photos/dullhunk/303503677/>

These kinds of issues have led to the development the Semantic web.

The Semantic Web is more than just putting data on the web.

# It's all about linking data!

It is about making links, so that a person or machine can explore the web of data.

When you have linked data you can follow the links to find other related data.

<http://www.w3.org/DesignIssues/LinkedData.html>

# It's not rocket science!

“ Is this rocket science? Well, not really. The Semantic Web, like the World Wide Web, is just taking well established ideas, and making them work interoperability over the Internet. ”

Tim Berners-Lee

Business Case for the Semantic Web  
<http://www.w3.org/DesignIssues/Business>

And the approach you take is actually pretty simple.

It's just building upon the existing web infrastructure.

It's not rocket science!

# Four Simple Rules

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 information.
4. Include links to other URIs so that they can discover more things.

There are four basic rules you need to follow.

# I. Use URLs as names for things

“Uniform Resource Identifier (URI), is a compact string of characters used to identify or name a resource.”

Wikipedia  
<http://en.wikipedia.org/wiki/URI>

The first step is to be able to identify things globally and uniquely.  
We use Uniform Resource Identifiers, URIs, to do this.

Think of a URI like a barcode in a shop.  
But made out of a sequence of letters.

And the idea is that everything gets a URI!

There are an infinite number of possible URIs.  
But each URI can name only one thing.  
You create different URIs to name different things

If you want to talk about something, use the URI.  
In a shop, if I say I want the ‘apple laptop’ it could be one of many models.  
But if I use the barcode, the computer can find the precise model I’m talking about.

YOU DON’T HAVE TO USE THE SAME URI - only if you’re talking about exactly the same thing.

## 2. Use HTTP URIs so that people can look up those names.

You use HTTP URIs so that people can look up those names.

HTTP is the Hypertext Transfer protocol - it's how the web works.

You're probably familiar a URL - which defines the location of a web page.

It starts with defining a protocol - in this case http.

\*

This is the name of the server

\*

This is the path on the server to the page you want to describe.

So a URI can identify resources, and they can tell you where you can find that resource.

But on the Semantic Web we use URIs not to find pages, but to bits of data.

\*

So this might be a URI for a person

\*

Here is a URI for the Mouse species.

So by using Web URIs, we just reuse existing infrastructure.

We don't need to create a brand new system for distributing information.

This was one of the problems with previous approaches, you had to get everyone to agree to the new system.

## 2. Use HTTP URIs so that people can look up those names.

http://

You use HTTP URIs so that people can look up those names.

HTTP is the Hypertext Transfer protocol - it's how the web works.

You're probably familiar a URL - which defines the location of a web page.

It starts with defining a protocol - in this case http.

\*

This is the name of the server

\*

This is the path on the server to the page you want to describe.

So a URI can identify resources, and they can tell you where you can find that resource.

But on the Semantic Web we use URIs not to find pages, but to bits of data.

\*

So this might be a URI for a person

\*

Here is a URI for the Mouse species.

So by using Web URIs, we just reuse existing infrastructure.

We don't need to create a brand new system for distributing information.

This was one of the problems with previous approaches, you had to get everyone to agree to the new system.

## 2. Use HTTP URIs so that people can look up those names.

`http://www.my.server.com`

You use HTTP URIs so that people can look up those names.

HTTP is the Hypertext Transfer protocol - it's how the web works.

You're probably familiar a URL - which defines the location of a web page.

It starts with defining a protocol - in this case http.

\*

This is the name of the server

\*

This is the path on the server to the page you want to describe.

So a URI can identify resources, and they can tell you where you can find that resource.

But on the Semantic Web we use URIs not to find pages, but to bits of data.

\*

So this might be a URI for a person

\*

Here is a URI for the Mouse species.

So by using Web URIs, we just reuse existing infrastructure.

We don't need to create a brand new system for distributing information.

This was one of the problems with previous approaches, you had to get everyone to agree to the new system.

## 2. Use HTTP URIs so that people can look up those names.

`http://www.my.server.com/page.html`

You use HTTP URIs so that people can look up those names.

HTTP is the Hypertext Transfer protocol - it's how the web works.

You're probably familiar a URL - which defines the location of a web page.

It starts with defining a protocol - in this case http.

\*

This is the name of the server

\*

This is the path on the server to the page you want to describe.

So a URI can identify resources, and they can tell you where you can find that resource.

But on the Semantic Web we use URIs not to find pages, but to bits of data.

\*

So this might be a URI for a person

\*

Here is a URI for the Mouse species.

So by using Web URIs, we just reuse existing infrastructure.

We don't need to create a brand new system for distributing information.

This was one of the problems with previous approaches, you had to get everyone to agree to the new system.

## 2. Use HTTP URIs so that people can look up those names.



[http://dbpedia.org/resource/Tim\\_Berners-Lee](http://dbpedia.org/resource/Tim_Berners-Lee)

<http://www.my.server.com/page.html>

You use HTTP URIs so that people can look up those names.

HTTP is the Hypertext Transfer protocol - it's how the web works.

You're probably familiar a URL - which defines the location of a web page.

It starts with defining a protocol - in this case http.

\*

This is the name of the server

\*

This is the path on the server to the page you want to describe.

So a URI can identify resources, and they can tell you where you can find that resource.

But on the Semantic Web we use URIs not to find pages, but to bits of data.

\*

So this might be a URI for a person

\*

Here is a URI for the Mouse species.

So by using Web URIs, we just reuse existing infrastructure.

We don't need to create a brand new system for distributing information.

This was one of the problems with previous approaches, you had to get everyone to agree to the new system.

## 2. Use HTTP URIs so that people can look up those names.



[http://dbpedia.org/resource/Tim\\_Berners-Lee](http://dbpedia.org/resource/Tim_Berners-Lee)

<http://www.my.server.com/page.html>



<http://dbpedia.org/resource/Mouse>

You use HTTP URIs so that people can look up those names.

HTTP is the Hypertext Transfer protocol - it's how the web works.

You're probably familiar a URL - which defines the location of a web page.

It starts with defining a protocol - in this case http.

\*

This is the name of the server

\*

This is the path on the server to the page you want to describe.

So a URI can identify resources, and they can tell you where you can find that resource.

But on the Semantic Web we use URIs not to find pages, but to bits of data.

\*

So this might be a URI for a person

\*

Here is a URI for the Mouse species.

So by using Web URIs, we just reuse existing infrastructure.

We don't need to create a brand new system for distributing information.

This was one of the problems with previous approaches, you had to get everyone to agree to the new system.

### 3. When someone looks up a URI, provide useful information.

<http://dbpedia.org/resource/Mouse>

The next step is that when someone looks up a URI you provide them with useful information.

But this is information aimed primarily at machines versus people.

So let's look up the resource Mouse using this URI.

(Let's look at a resource from DBpedia.

Wikipedia is the free, open source encyclopaedia project.

Anyone can join up, where anyone can add or edit information.

DBpedia is a great project, they are extracting the structured information from Wikipedia and serving it using Semantic Web techniques.)

<<http://dbpedia.org/data/Mouse>> <<http://www.w3.org/2000/01/rdf-schema#label>> "RDF description of Mouse" .  
<<http://dbpedia.org/data/Mouse>> <<http://xmlns.com/foaf/0.1/primaryTopic>> <<http://dbpedia.org/resource/Mouse>> .  
<<http://dbpedia.org/resource/Ascendancy>> <<http://dbpedia.org/property/input>> <<http://dbpedia.org/resource/Mouse>> .  
<[http://dbpedia.org/resource/A\\_Fork\\_in\\_the\\_Tale](http://dbpedia.org/resource/A_Fork_in_the_Tale)> <<http://dbpedia.org/property/input>> <<http://dbpedia.org/resource/Mouse>> .  
<<http://dbpedia.org/resource/Babyz>> <<http://dbpedia.org/property/input>> <<http://dbpedia.org/resource/Mouse>> .  
<[http://dbpedia.org/resource/Blue\\_Ice\\_%28computer\\_game%29](http://dbpedia.org/resource/Blue_Ice_%28computer_game%29)> <<http://dbpedia.org/property/input>> <<http://dbpedia.org/resource/Mouse>> .  
<[http://dbpedia.org/resource/Gus\\_Goes\\_to\\_Cybertown](http://dbpedia.org/resource/Gus_Goes_to_Cybertown)> <<http://dbpedia.org/property/input>> <<http://dbpedia.org/resource/Mouse>> .  
<[http://dbpedia.org/resource/Mall\\_Tycoon\\_2](http://dbpedia.org/resource/Mall_Tycoon_2)> <<http://dbpedia.org/property/input>> <<http://dbpedia.org/resource/Mouse>> .  
<[http://dbpedia.org/resource/Mixed-Up\\_Fairy\\_Tales](http://dbpedia.org/resource/Mixed-Up_Fairy_Tales)> <<http://dbpedia.org/property/input>> <<http://dbpedia.org/resource/Mouse>> .  
<<http://dbpedia.org/resource/Mouse>> <<http://dbpedia.org/abstract>> "#REDIRECT Mus musculus"@it .  
<<http://dbpedia.org/resource/Mouse>> <<http://dbpedia.org/abstract>> "De naam muis wordt gebruikt voor allerlei kleine zoogdieren, ..."@nl  
<<http://dbpedia.org/resource/Mouse>> <<http://dbpedia.org/abstract>> "Este art\u00EDculo trata sobre el animal rat\u00E1n. Para el rat\u00E1n..."@es  
<<http://dbpedia.org/resource/Mouse>> <<http://dbpedia.org/abstract>> "La souris ou souris commune (Mus musculus) est un petit rongeur de..."@fr  
<<http://dbpedia.org/resource/Mouse>> <<http://dbpedia.org/abstract>> "M\u00E4rkte Murinae \u00E4r en underfamilj i familjen ..."@sv  
<<http://dbpedia.org/resource/Mouse>> <<http://dbpedia.org/abstract>> "Potoczne okre\u015Blenie mysz stosujemy przede wszystkim ..."@pl  
<<http://dbpedia.org/resource/Mouse>> <<http://dbpedia.org/abstract>> "Ver texto.0 g\u00E9nero Mus inclui esp\u00E9cies a que normalmente..."@pt  
<<http://dbpedia.org/resource/Mouse>> <<http://dbpedia.org/abstract>> "Ver texto.0 g\u00E9nero Mus inclui esp\u00E9cies a que normalmente..."@de  
<<http://dbpedia.org/resource/Mouse>> <<http://dbpedia.org/property/abstract>> "A mouse (plural mice) is a mammal that belongs to one of..."@en  
<<http://dbpedia.org/resource/Mouse>> <<http://dbpedia.org/property/classis>> "[[Mammal]]ia" .  
<<http://dbpedia.org/resource/Mouse>> <<http://dbpedia.org/property/color>> "pink" .  
<<http://dbpedia.org/resource/Mouse>> <<http://dbpedia.org/property/familia>> <<http://dbpedia.org/resource/Muridae>> .  
<<http://dbpedia.org/resource/Mouse>> <[http://dbpedia.org/property/fossil\\_range](http://dbpedia.org/property/fossil_range)> "Late [[Miocene]] - Recent" .  
<<http://dbpedia.org/resource/Mouse>> <<http://dbpedia.org/property/genus>> "\*\*\*\*'Mus'\*\*\*\*" .  
<<http://dbpedia.org/resource/Mouse>> <[http://dbpedia.org/property/genus\\_authority](http://dbpedia.org/property/genus_authority)> "[[Carolus Linnaeus|Linnaeus]], 1758" .  
<<http://dbpedia.org/resource/Mouse>> <<http://dbpedia.org/property/image>> <[http://upload.wikimedia.org/wikipedia/commons/e/e7/house\\_mou](http://upload.wikimedia.org/wikipedia/commons/e/e7/house_mou)> .  
<<http://dbpedia.org/resource/Mouse>> <[http://dbpedia.org/property/image\\_caption](http://dbpedia.org/property/image_caption)> "House mouse, ''[[Mus musculus]]''' .  
<<http://dbpedia.org/resource/Mouse>> <[http://dbpedia.org/property/image\\_width](http://dbpedia.org/property/image_width)> "250px" .  
<<http://dbpedia.org/resource/Mouse>> <<http://dbpedia.org/property/linktargetcount>> "135"^^<<http://www.w3.org/2001/XMLSchema#int>> .  
<<http://dbpedia.org/resource/Mouse>> <<http://dbpedia.org/property/name>> "Mouse" .  
<<http://dbpedia.org/resource/Mouse>> <<http://dbpedia.org/property/ordo>> "[[Rodent]]ia" .  
<<http://dbpedia.org/resource/Mouse>> <<http://dbpedia.org/property/phylum>> <<http://dbpedia.org/resource/Chordata>> .  
<<http://dbpedia.org/resource/Mouse>> <<http://dbpedia.org/property/reference>> <<http://brainmaps.org/index.php?p=speciedata&species=mus>> .  
<<http://dbpedia.org/resource/Mouse>> <<http://dbpedia.org/property/reference>> <<http://greenfield.fortunecity.com/hummingbird/173/animal>> .  
<<http://dbpedia.org/resource/Mouse>> <<http://dbpedia.org/property/reference>> <[http://members.madasafish.com/~cj.whitehound/Rats\\_Nest/S](http://members.madasafish.com/~cj.whitehound/Rats_Nest/S)> .  
<<http://dbpedia.org/resource/Mouse>> <<http://dbpedia.org/property/reference>> <<http://pavelicious.boo.pl/fls/mice/>> .  
<<http://dbpedia.org/resource/Mouse>> <<http://dbpedia.org/property/reference>> <<http://web.singnet.com.sg/~jerymaus/mice.htm>> .  
<<http://dbpedia.org/resource/Mouse>> <<http://dbpedia.org/property/reference>> <<http://www.fancymice.info>> .  
<<http://dbpedia.org/resource/Mouse>> <<http://dbpedia.org/property/reference>> <<http://www.miceandrats.com/miceas.htm>> .  
<<http://dbpedia.org/resource/Mouse>> <<http://dbpedia.org/property/reference>> <<http://www.nature.com/news/2005/050718/full/050718-2.htm>> .  
<<http://dbpedia.org/resource/Mouse>> <<http://dbpedia.org/property/reference>> <<http://www.petco.com>> .  
<<http://dbpedia.org/resource/Mouse>> <<http://dbpedia.org/property/reference>> <<http://www.petwebsite.com/mice/mice.htm>> .  
<<http://dbpedia.org/resource/Mouse>> <<http://dbpedia.org/property/reference>> <<http://www.worldscibooks.com/lifesci/etextbook/p393/p393>>

And you get back data!

And it's structured in a way that computers can understand!

Mouse at DBpedia.org  
<http://dbpedia.org/resource/Mouse>

**Depiction of \$label** A mouse (plural mice) is a mammal that belongs to one of numerous species of small rodents. The best known mouse species is the common house mouse (*Mus musculus*). It is found in nearly all countries and, as the laboratory mouse, serves as an important model organism in biology; it is also a popular pet. The American white-footed mouse (*Peromyscus leucopus*) and the deer mouse (*Peromyscus maniculatus*) also sometimes live in houses. These species of mice live commensally with humans.

Property	Value
?:abel	Mouse
p:abstract	A mouse (plural mice) is a mammal that belongs to one of numerous species of small rodents. The best known mouse species is the common house mouse ( <i>Mus</i> ... <a href="#">»more»</a> (en))
?:abstract	#REDIRECT <i>Mus musculus</i> (it) De naam muis wordt gebruikt voor allerlei kleine zoogdieren, in het bijzonder de huismuis ( <i>Mus domesticus</i> ), maar ook vele andere soorten, vooral knaagdieren. ... <a href="#">»more»</a> (nl) Este artículo trata sobre el animal ratón. Para el ratón como hardware, véase Mouse. Se denomina ratón a varias especies de pequeños roedores del género ... <a href="#">»more»</a> (es) La souris ou souris commune ( <i>Mus musculus</i> ) est un petit rongeur de la famille des muridés et du genre <i>Mus</i> . Il existe une quarantaine d'espèces de ce genre. Certaines ... <a href="#">»more»</a> (fr) Möss (Murinae) är en underfamilj i familjen råttdjur och även beteckning för andra musliknande arter i ordningen gnagare. Möss är nu förtiden ett relativt vanligt sällskapsdjur. Tammöss lever ca 1,5-3 år." (sv) Potoczne określenie mysz stosujemy przede wszystkim do małych gryzoni z rodziny myszowatych (Muridae), a także przedmiotów, które takie gryzonie w jakimś ... <a href="#">»more»</a> (pl) Ver texto. O género <i>Mus</i> inclui espécies a que normalmente se apelida de ratos." (pt) <u>NOTOC</u> Das Wort Maus bezeichnet <sup>*</sup> (de) ネズミ上科ハツカネズミ分類 界: 動物界 Animalia 門: 脊索動物門 Chordata 亜門:脊椎動物亜門 Vertebrata 級: 哺乳綱 Mammalia 目:ネズミ目(齧歯目) Rodentia 上科:ネズミ上科 Myomorpha 科 本文参照 ネズミ（鼠）とは、ネズミ目（齧歯目）ネズミ上科に属する哺乳動物の総称である。ハツカネズミ、ドブネズミなど、1,000種以上が含まれる一大グループを形成している。ネズミのほとんどが夜行性で、よく人間が寝ている間に人間の食料などを食べたりするので、寝盗み（寝ている間にこっそりと人間の食料を盗む）が転じて、ネズミという名がついたといわれる。また、ネズミは齧歯

You can take this data and make it more readable for people.

But it's not very user friendly - it's just a list of properties and values describing the mouse species.

- About Pedialix
- Register as search plugin
- Content released under GFD-Licence
- Original article at en.wikipedia.org
- Changes

Search

Toolbox [+]

In other languages [+]

**Ads by Google**[Wikipedia Wiki](#)

Get Wikipedia Wiki!  
Search with Ask for anything.

[www.Ask.com](http://www.Ask.com)[Wikipedia](#)

Search Wikipedia! The more you eBay, the less you pay  
[www.ebay.co.uk](http://www.ebay.co.uk)

# Mouse

For other uses, see [Mouse \(disambiguation\)](#).

A **mouse** (Plural **mice**) is a **mammal** that belongs to one of numerous **species** of small **rodents**. The best known mouse species is the **common house mouse** (*Mus musculus*). It is found in nearly all countries and, as the laboratory mouse, serves as an important **model organism** in biology; it is also a popular **pet**. The **American white-footed**

Mouse	
Fossil range: Late Miocene - Recent	
House mouse, <i>Mus musculus</i>	
Scientific classification	
Kingdom:	Animalia
Phylum:	Chordata
Class:	Mammalia
Order:	Rodentia
Superfamily:	Muroidea
Family:	Muridae
Subfamily:	Murinae
Genus:	<i>Mus</i>

Categories: Articles lacking sources from August 2006 | All articles lacking sources | Animals kept as pets | Muroid rodents | Old World rats and mice

**Contents** [hide]

- 1 Body & behavior
- 2 Taxonomy of the genus *Mus*
- 3 Laboratory mice
- 4 History
- 5 Mice as feeder animals
- 6 Mice as pets
- 7 Mice in popular culture
- 8 Trivia
- 9 See also
- 10 References
- 11 External links

Most popular articles [+]

Most popular incoming links [-]

Most popular outgoing links [-]

And you can also use the data to build well formatted web pages that people can use - just like the current web.

## 4. Include links to other URLs so that they can discover more things.

The next part is that you link to other URLs so that you can discover more things.

And this is where it gets interesting!

# RDF

## Resource Description Framework

There are standards that define how you should do this.

We start with RDF.

It stands for the Resource Description Framework.

And it's one of the key technologies for the Semantic Web.

# RDF

subject - predicate - object

RDF describes data in triples: subject, predicate, object.

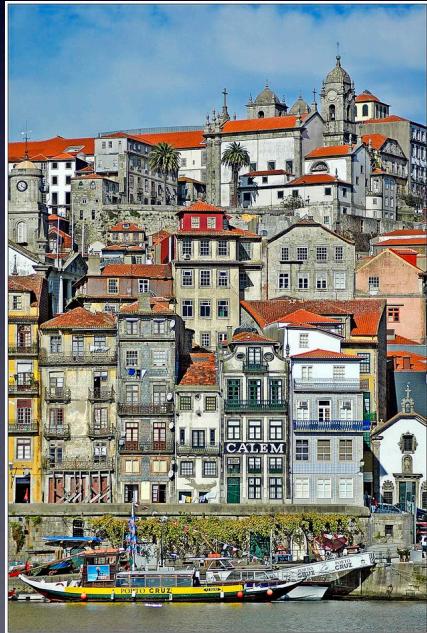
- \* Let's try to describe some facts about Porto.
- \* In English, we would say: Porto has a population of 249633.
- \* The subject of this statement is 'Porto', the predicate is 'population', the object is the number.

And in RDF we're using URIs.

- \* So there's a URI for Porto.
- \* There's a URI for the predicate, a property defining that a place has a population.
- \* And there's a literal value, a number.

# RDF

subject - predicate - object



<http://flickr.com/photos/apiidim/54528200/>

RDF describes data in triples: subject, predicate, object.

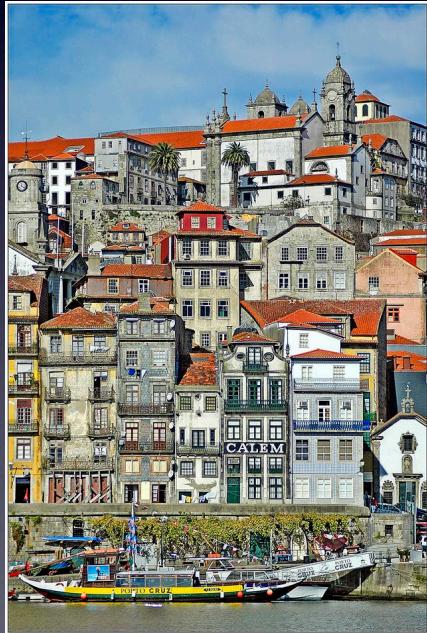
- \* Let's try to describe some facts about Porto.
- \* In English, we would say: Porto has a population of 249633.
- \* The subject of this statement is 'Porto', the predicate is 'population', the object is the number.

And in RDF we're using URIs.

- \* So there's a URI for Porto.
- \* There's a URI for the predicate, a property defining that a place has a population.
- \* And there's a literal value, a number.

# RDF

subject - predicate - object



<http://flickr.com/photos/apiidim/54528200/>

Porto has a population of 249633.

RDF describes data in triples: subject, predicate, object.

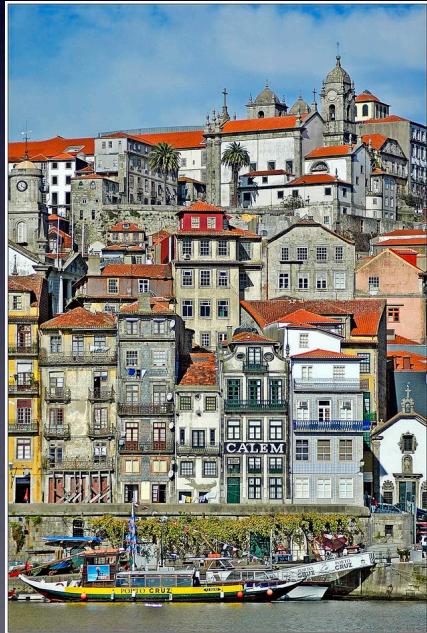
- \* Let's try to describe some facts about Porto.
- \* In English, we would say: Porto has a population of 249633.
- \* The subject of this statement is 'Porto', the predicate is 'population', the object is the number.

And in RDF we're using URIs.

- \* So there's a URI for Porto.
- \* There's a URI for the predicate, a property defining that a place has a population.
- \* And there's a literal value, a number.

# RDF

subject - predicate - object



<http://flickr.com/photos/apiidim/54528200/>

Porto has a population of 249633.

**Porto** has a **population** of **249633**.

RDF describes data in triples: subject, predicate, object.

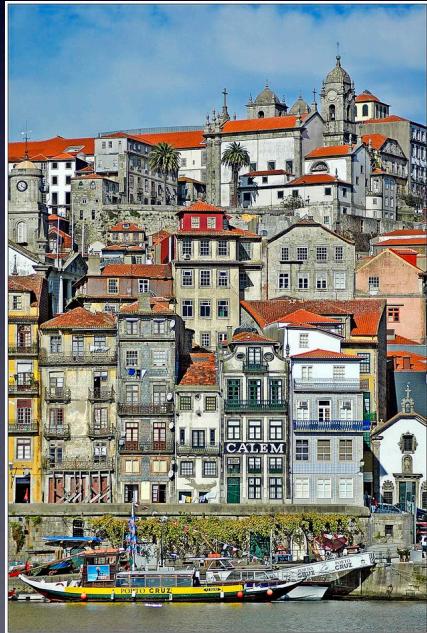
- \* Let's try to describe some facts about Porto.
- \* In English, we would say: Porto has a population of 249633.
- \* The subject of this statement is 'Porto', the predicate is 'population', the object is the number.

And in RDF we're using URIs.

- \* So there's a URI for Porto.
- \* There's a URI for the predicate, a property defining that a place has a population.
- \* And there's a literal value, a number.

# RDF

subject - predicate - object



<http://flickr.com/photos/apiidim/54528200/>

Porto has a population of 249633.

**Porto** has a **population** of **249633**.

<http://sws.geonames.org/2735943/>

RDF describes data in triples: subject, predicate, object.

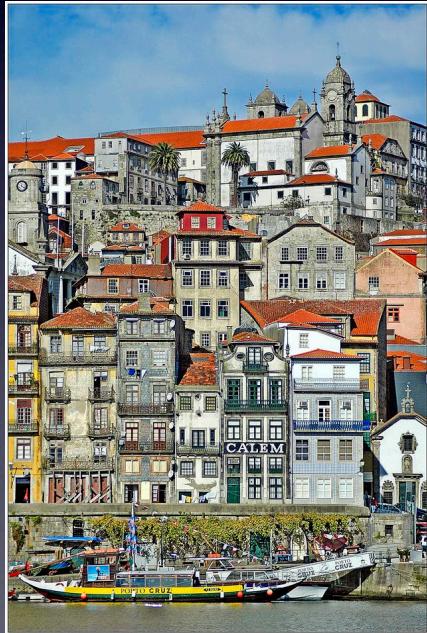
- \* Let's try to describe some facts about Porto.
- \* In English, we would say: Porto has a population of 249633.
- \* The subject of this statement is 'Porto', the predicate is 'population', the object is the number.

And in RDF we're using URIs.

- \* So there's a URI for Porto.
- \* There's a URI for the predicate, a property defining that a place has a population.
- \* And there's a literal value, a number.

# RDF

subject - predicate - object



<http://flickr.com/photos/apiidim/54528200/>

Porto has a population of 249633.

**Porto** has a **population** of **249633**.

<http://sws.geonames.org/2735943/>

<http://www.geonames.org/ontology#population>

RDF describes data in triples: subject, predicate, object.

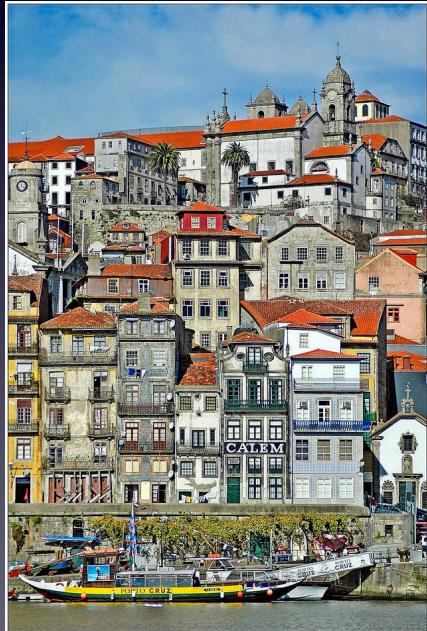
- \* Let's try to describe some facts about Porto.
- \* In English, we would say: Porto has a population of 249633.
- \* The subject of this statement is 'Porto', the predicate is 'population', the object is the number.

And in RDF we're using URIs.

- \* So there's a URI for Porto.
- \* There's a URI for the predicate, a property defining that a place has a population.
- \* And there's a literal value, a number.

# RDF

subject - predicate - object



<http://flickr.com/photos/apiidim/54528200/>

Porto has a population of 249633.

**Porto** has a **population** of **249633**.

<http://sws.geonames.org/2735943/>

<http://www.geonames.org/ontology#population>

249633

RDF describes data in triples: subject, predicate, object.

- \* Let's try to describe some facts about Porto.
- \* In English, we would say: Porto has a population of 249633.
- \* The subject of this statement is 'Porto', the predicate is 'population', the object is the number.

And in RDF we're using URIs.

- \* So there's a URI for Porto.
- \* There's a URI for the predicate, a property defining that a place has a population.
- \* And there's a literal value, a number.

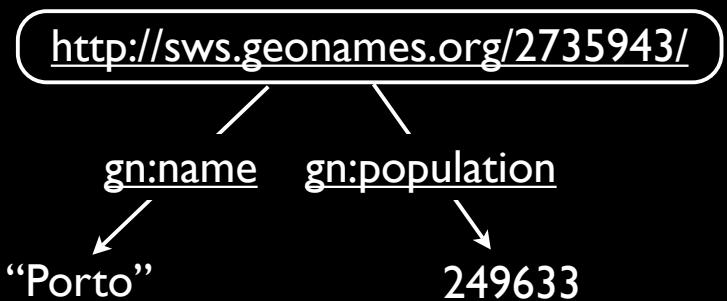
prefix gn: <<http://www.geonames.org/ontology#>>

<http://sws.geonames.org/2735943/>

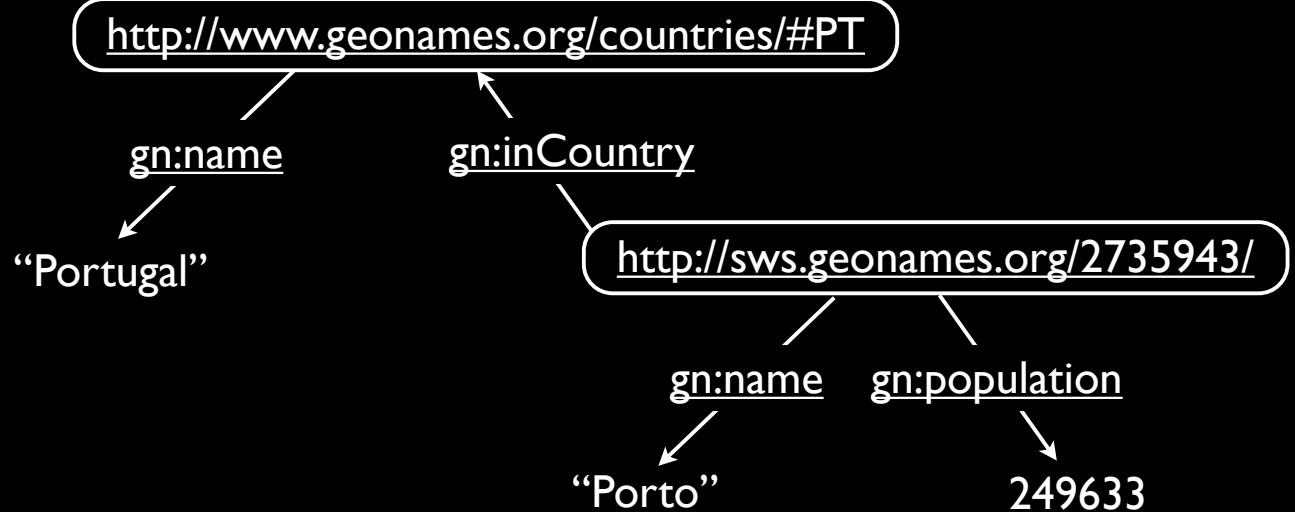
gn:population

249633

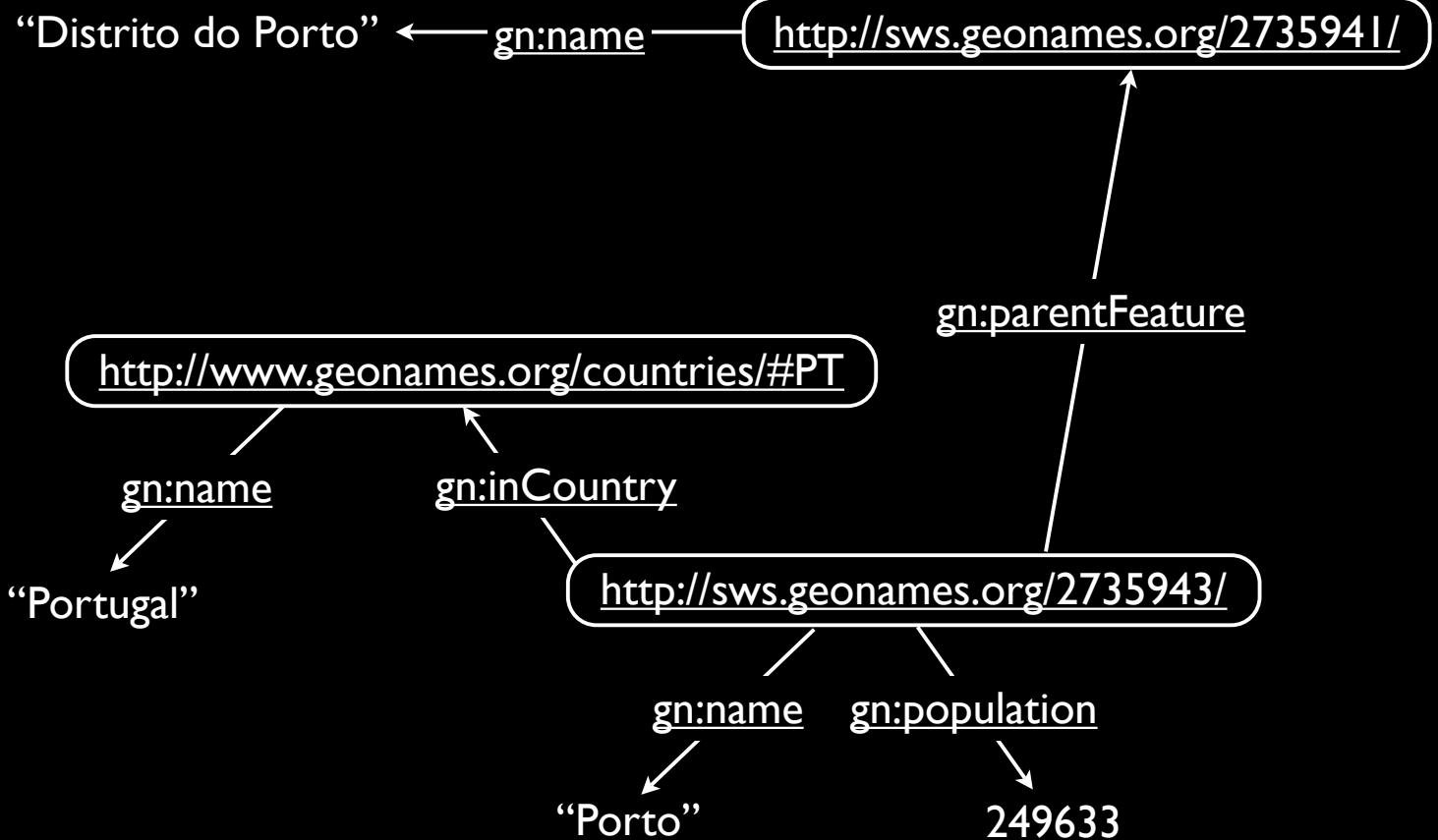
prefix gn: <<http://www.geonames.org/ontology#>>



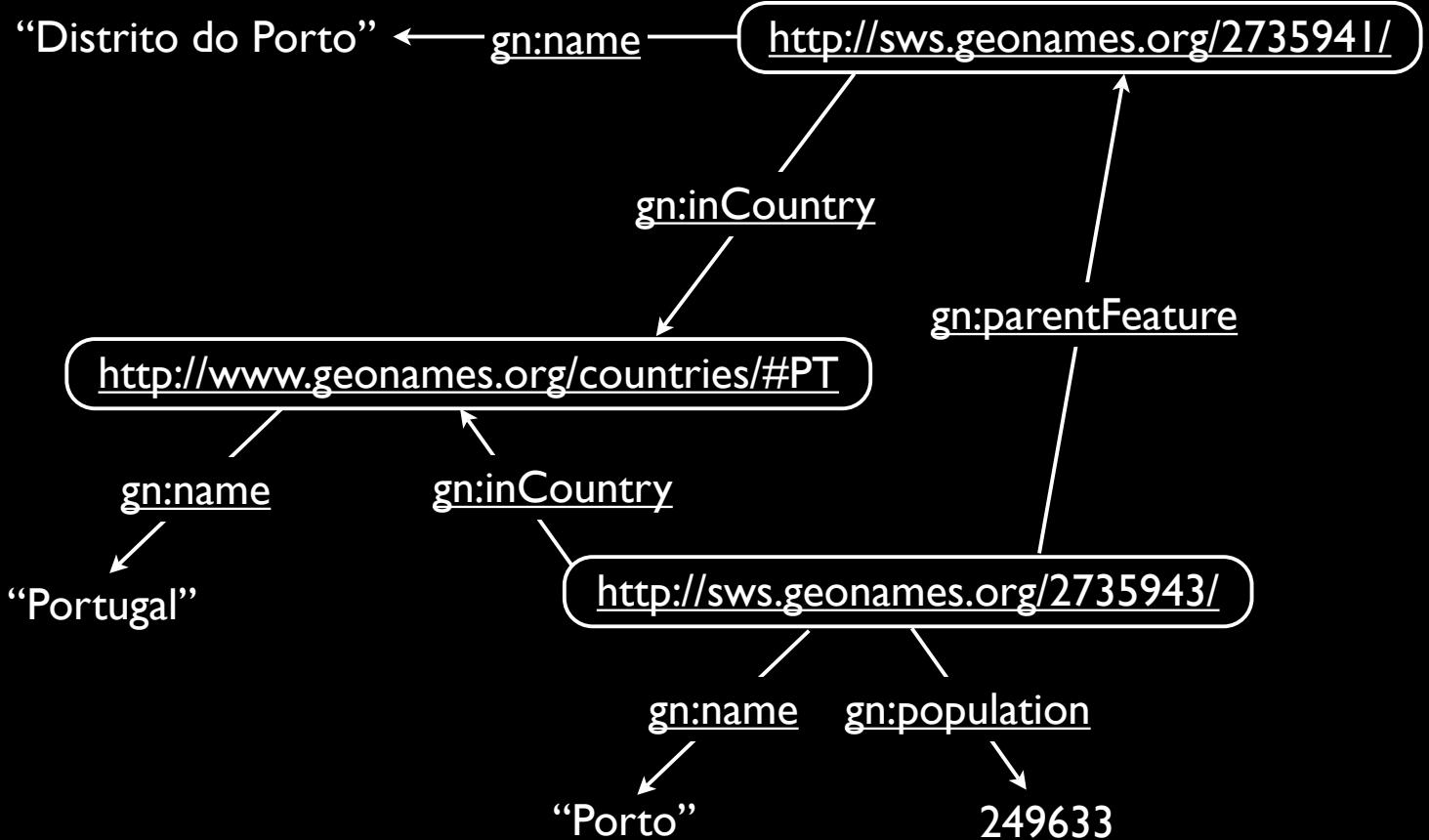
prefix gn: <<http://www.geonames.org/ontology#>>



prefix gn: <<http://www.geonames.org/ontology#>>



prefix gn: <<http://www.geonames.org/ontology#>>



# RDF is a graph

the cell Purkinje neuron has the molecule alpha4-1 subunit nicotinic receptor within

<http://www.w3.org/2000/01/rdf-schema#label> -> the cell Purkinje neuron has the molecule alpha4-1 subunit nicotinic receptor within

RDF data is a graph.

Each triple, (subject, predicate, object) is a labelled edge on a graph.

So instead of thinking about web sites, web pages and documents, think in terms of graphs.

Graph created using: <http://simile.mit.edu/welkin>

Data from: [http://esw.w3.org/topic/HCLS/HCLSIG\\_DemoHomePage\\_HCLSIG\\_Demo](http://esw.w3.org/topic/HCLS/HCLSIG_DemoHomePage_HCLSIG_Demo)

# Merging RDF

<http://flickr.com/photos/mcclanahan/4915491/>

The great thing about RDF is that you can merge graphs from different sources.

Nodes with the same URIs are considered identical.

There are no limitations on the graphs – you can mix any data you want!

Adding triples never changes the meaning of a graph  
i.e. you can't unsay things by adding triples  
and you can't invalidate earlier conclusions

# Traditional approaches

- Relational Database:
  - You typically have to predefine a structure
  - You have to stick with that structure
- XML:
  - There are no standard ways to represent relations between items

RDF is really flexible compared to traditional approaches such as databases and XML.

Limitations of XML:

- no standard way to represent relationships between items
- no standard way to query the web site other than polling

# Ontologies

“An ontology is a specification of a conceptualization.”

Tom Gruber

What is an Ontology?

<http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>

But it is still important to define structure in the data.  
You need to model complex information structures so that other people can understand what you're describing  
This allows them to query your data.

With the Semantic Web, we use ontologies to define structure.  
Ontologies are originally a philosophical concept, but they've hijacked by the computer scientists.

This is the classic definition of an ontology for computer scientist.

“An ontology is a data model that represents a set of concepts within a domain and the relationships between those concepts. It is used to reason about the objects within that domain.”

Wikipedia

<http://en.wikipedia.org/wiki/Ontology>

I prefer Wikipedia's definition!

# OWL



## Web Ontology Language

To describe ontologies on the semantic web, we use OWL, the web ontology language .

(By the way, this is the wise, but dyslexic, owl from the children's book Winnie the Pooh.

The rumour is that the people working on OWL didn't like the acronym WOL - web ontology language.

But then someone remembered the owl from winnie the pooh, who spelt his name wol.

So they decided to use the acronym OWL for web ontology language.

It's a computer scientist joke :)

- define “city” as a concept
- Porto is a city
- “city” is a subclass of “populated place”
- the property “population” only applies to “populated place” and is a numeric value

Let's go back to the city and places example

There are properties we want to define, like the concept city.  
So we can do queries like ‘give me all of the cities in the world’.

This is how you do it in OWL...

```
prefix owl: <http://www.w3.org/2002/07/owl#>
prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
prefix geonames: <http://www.geonames.org/ontology#>
```

- define “city” as a concept
- Porto is a city
- “city” is a subclass of “populated place”
- the property “population” only applies to “populated place” and is a numeric value

Let's go back to the city and places example

There are properties we want to define, like the concept city.  
So we can do queries like ‘give me all of the cities in the world’.

This is how you do it in OWL...

```
prefix owl: <http://www.w3.org/2002/07/owl#>
prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
prefix geonames: <http://www.geonames.org/ontology#>
```

- define “city” as a concept

```
geonames:City rdf:type owl:Class .
```

- Porto is a city

- “city” is a subclass of “populated place”
- the property “population” only applies to “populated place” and is a numeric value

Let's go back to the city and places example

There are properties we want to define, like the concept city.  
So we can do queries like ‘give me all of the cities in the world’.

This is how you do it in OWL...

```
prefix owl: <http://www.w3.org/2002/07/owl#>
prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
prefix geonames: <http://www.geonames.org/ontology#>
```

- define “city” as a concept

geonames:City rdf:type owl:Class .

- Porto is a city

<<http://sws.geonames.org/2735943/>> rdf:type geonames:City .

- “city” is a subclass of “populated place”

- the property “population” only applies to “populated place” and is a numeric value

Let's go back to the city and places example

There are properties we want to define, like the concept city.  
So we can do queries like ‘give me all of the cities in the world’.

This is how you do it in OWL...

prefix owl: <<http://www.w3.org/2002/07/owl#>>  
prefix rdf: <<http://www.w3.org/1999/02/22-rdf-syntax-ns#>>  
prefix geonames: <<http://www.geonames.org/ontology#>>

- define “city” as a concept

geonames:City rdf:type owl:Class .

- Porto is a city

<<http://sws.geonames.org/2735943/>> rdf:type geonames:City .

- “city” is a subclass of “populated place”

geonames:City rdfs:subClassOf geonames:PopulatedPlace .

- the property “population” only applies to “populated place” and is a numeric value

Let's go back to the city and places example

There are properties we want to define, like the concept city.  
So we can do queries like ‘give me all of the cities in the world’.

This is how you do it in OWL...

```
prefix owl: <http://www.w3.org/2002/07/owl#>
prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
prefix geonames: <http://www.geonames.org/ontology#>
```

- define “city” as a concept

geonames:City rdf:type owl:Class .

- Porto is a city

<<http://sws.geonames.org/2735943/>> rdf:type geonames:City .

- “city” is a subclass of “populated place”

geonames:City rdfs:subClassOf geonames:PopulatedPlace .

- the property “population” only applies to “populated place” and is a numeric value

geonames:population rdfs:type owl:DatatypeProperty .

geonames:population rdfs:domain geonames:PopulatedPlace .

geonames:population rdfs:range xsd:integer .

Let's go back to the city and places example

There are properties we want to define, like the concept city.  
So we can do queries like ‘give me all of the cities in the world’.

This is how you do it in OWL...

# OWL Examples

- “married to” is a **symmetric** property
- “father of” is not
- “human” is a subclass of “animal”
- “person” is equivalent to “human”
- there can only be one biological mother
- “son of” is the inverse of “father of”

*A no-nonsense introduction to “Semantic Web” technologies*

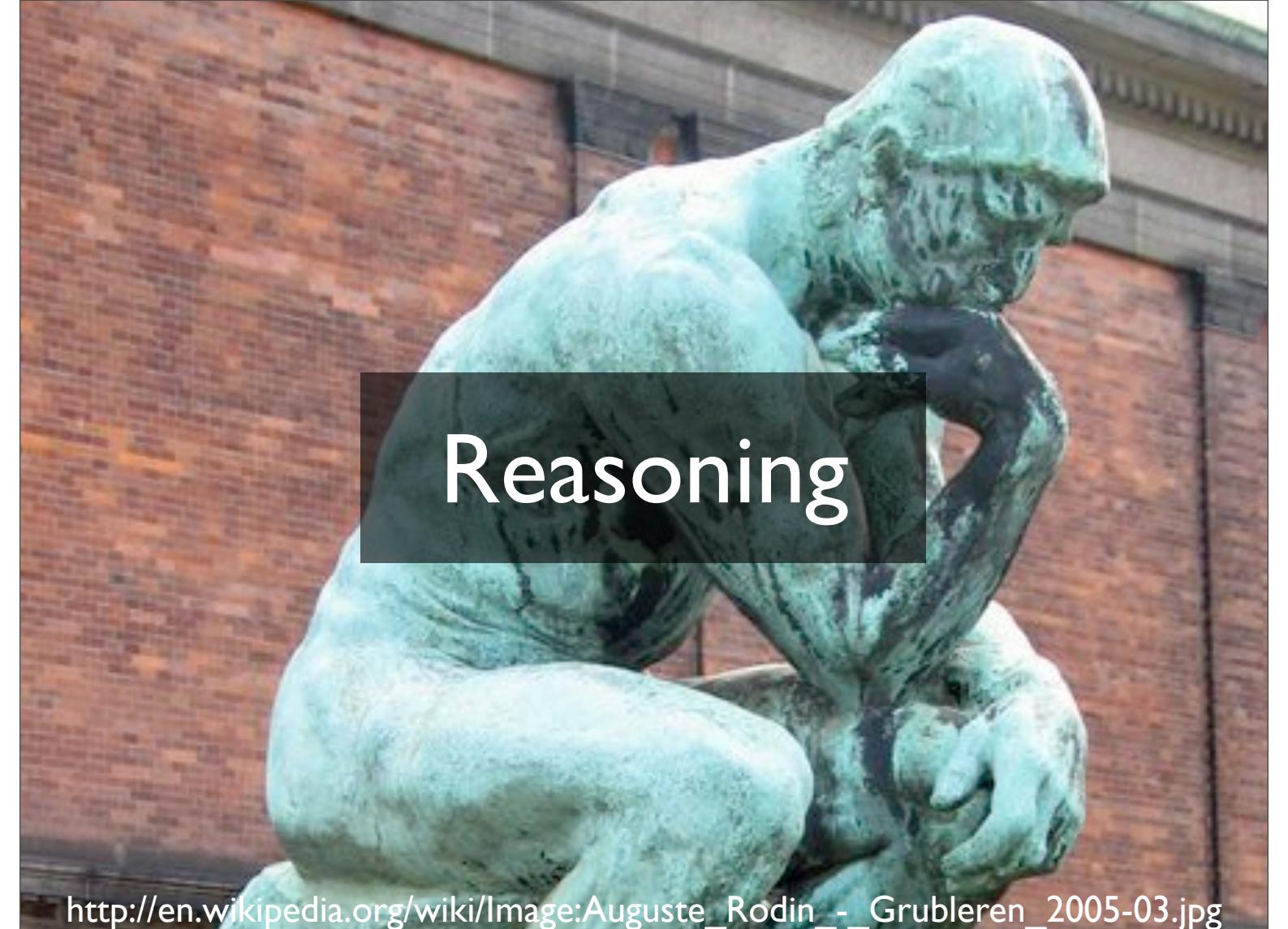
Stefano Mazzocchi, 2007

<http://www.betaversion.org/%7Estefano/papers/ac2006.1.pdf>

And you can do some pretty powerful things with OWL.

So you can say that ...

<http://www.betaversion.org/%7Estefano/papers/ac2006.1.pdf>



# Reasoning

[http://en.wikipedia.org/wiki/Image:Auguste\\_Rodin\\_-\\_Grubleren\\_2005-03.jpg](http://en.wikipedia.org/wiki/Image:Auguste_Rodin_-_Grubleren_2005-03.jpg)

By specifying this detailed information about concepts and properties, we can reason over the information.

That is, we can draw conclusions based on the facts we know from the restrictions and concept/property descriptions.



Jon Hare  
University of Southampton



I want to show a simple example of this.

It's from an image search search engine project at Southampton.

The Kennel Club is an organisation in London that is involved in the promotion of dogs and dog activity.

So they have a nice photo collection of dogs that we did some experiments on.

We're doing a search for hound.  
It's the basic term for hunting dog – cão de caça.

When doing a keyword search, it matches the word 'hound' in the photo descriptions.

So 95 images are returned.

When we search using the concept, many more images are returned.

# Reasoning applications in Medicine and Bioinformatics

- Querying tools
- Automatic classification techniques
  - Protein classification using ontology classification
- Checking data is consistent

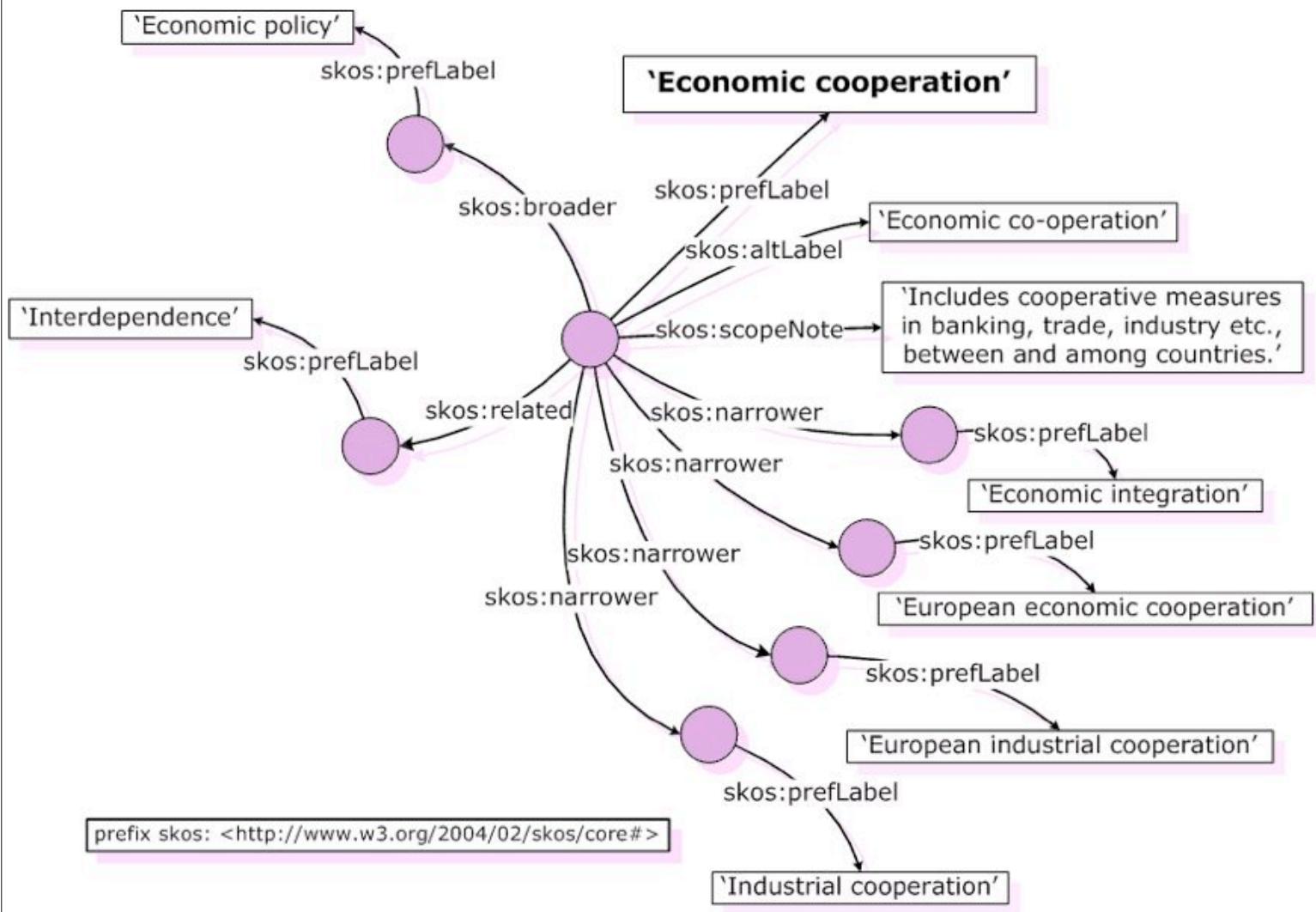
# SKOS

Simple Knowledge Organization System

<http://www.w3.org/2004/02/skos/>

Another useful data modelling tool for the semantic web is SKOS.

This lets you define more simplistic knowledge organisation systems, like a thesaurus.



A thesaurus describe concepts.

You typically use a thesaurus to index material – e.g. publications.

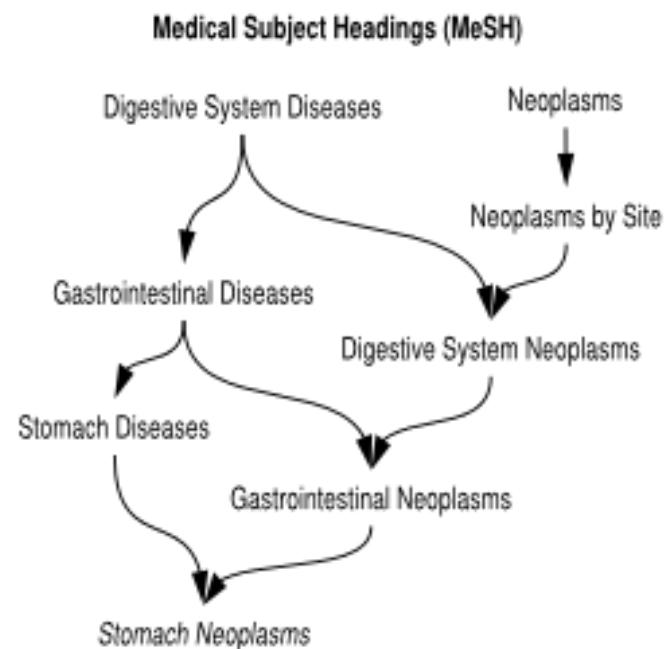
Concepts can have labels, and descriptions.

The interesting thing is describing the relations.

- broader (i.e. more general)
- narrower (more specific)
- related concepts

# Medical Subject Headings (MeSH)

- Thesaurus of medical terms
- Used for indexing journals and books
- Available in SKOS format at:  
<http://thesauri.cs.vu.nl/>



*Collaborative thesaurus tagging the Wikipedia way*  
Jakob Voss

For example, the Medical Subject Headings, a thesaurus of medical terms has been converted to SKOS format.

# SPARQL

## SPARQL Query Language for RDF

So we can define all of this data in a standard format.  
But we need to be able to query the data to find what we want.

For that we have SPARQL - the RDF query language.

Typically you store RDF data in a database, called a triplestore

prefix gn: <<http://www.geonames.org/ontology#>>

<<http://sws.geonames.org/2735943/>> gn:name “Porto” .

prefix gn: <<http://www.geonames.org/ontology#>>

<<http://sws.geonames.org/2735943/>> gn:name “Porto” .

SELECT ?name

WHERE

{

<<http://sws.geonames.org/2735943/>> gn:name ?name .

}

prefix gn: <<http://www.geonames.org/ontology#>>

<<http://sws.geonames.org/2735943/>> gn:name “Porto” .

SELECT ?name

WHERE

{

<<http://sws.geonames.org/2735943/>> gn:name ?name .

}

name

“Porto”

prefix gn: <<http://www.geonames.org/ontology#>>

```
SELECT ?capital ?country
WHERE {
  ?x gn:name ?capital.
  ?y gn:name ?country.
  ?x gn:capitalCity ?y.
  ?y gn:continent <http://sws.geonames.org/6255146/>.
}
```

prefix gn: <<http://www.geonames.org/ontology#>>

```
SELECT ?capital ?country
WHERE {
  ?x gn:name ?capital.
  ?y gn:name ?country.
  ?x gn:capitalCity ?y.
  ?y gn:continent <http://sws.geonames.org/6255146/>.
}
```

capital	country
“Algiers”	“Algeria”
“Cairo”	“Egypt”
“Luanda”	“Angola”
“Nairobi”	“Kenya”

```

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}
}

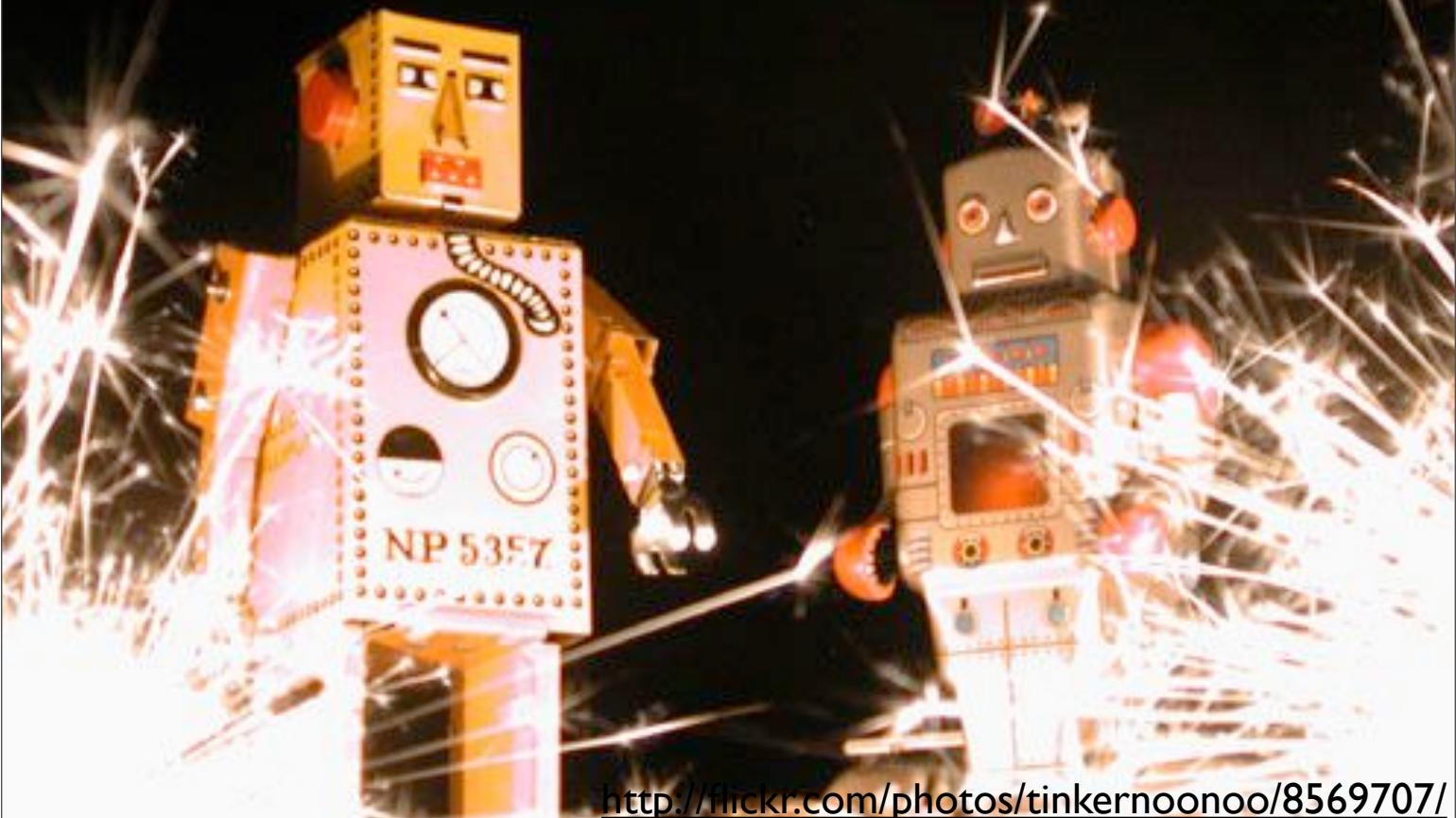
```

Obviously writing these queries can get very complicated.

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

and so is processing the results

# SPARQL is for machines!



<http://flickr.com/photos/tinkernoonoo/8569707/>

But that's because SPARQL is for machines!

# What about the users?

mSpace

<http://www.mspace.fm>

<http://beta.mspace.fm>

The screenshot shows the mSpace Classical Music Explorer application. On the left, there's a vertical sidebar labeled "INTEREST". The main area has three main sections:

- Browser:** A panel with three tabs: "Era", "Composer", and "Piece". The "Era" tab is active, showing a list of historical periods: Renaissance, Baroque, Classical, Romantic, Modern, and Contemporary. The "Contemporary" item is selected and highlighted with a blue background.
- Information:** A panel containing a heading "mSpace Explorer" and a descriptive text block. The text reads: "This is a demonstration of an mSpace explorer that allows you to explore large sets of contextual information using several unique features to help you find what you may need or be interested in, even if you don't know anything about the domain, beyond the fact that it interests you."
- Preview:** A panel on the right which is currently empty.

Below the main interface, a caption reads: "mSpace Classical Music Explorer. Part of the mSpace network."

The mSpace application is generating a complex sparql query for us and displaying the results in this really nice interface

A photograph showing a DJ's hands on a turntable and mixer. The turntable has two black vinyl records on it. The mixer is silver with various knobs and buttons. The background is dark.

# Remix Information

[http://flickr.com/photos/smoking\\_gun/309199668/](http://flickr.com/photos/smoking_gun/309199668/)

But there's more to SPARQL than querying

There are great features for remixing information.

# Describing Resources

```
DESCRIBE ?x  
WHERE  
{  
    ?x gn:name “Porto” .  
}
```

prefix gn: <<http://www.geonames.org/ontology#>>

```
<http://sws.geonames.org/2735943/> gn:name “Porto” .  
<http://sws.geonames.org/2735943/> gn:population 249633 .  
<http://sws.geonames.org/2735943/> gn:parentFeature <http://sws.geonames.org/2735941/> .  
<http://sws.geonames.org/2735943/> gn:inCountry <http://www.geonames.org/countries/#PT> .
```

...

# Reshape data

```
prefix gn: <http://www.geonames.org/ontology#>
prefix wkp: <http://www.wikipedia.org/ontology#>
```

```
CONSTRUCT {
  <http://sws.geonames.org/2735943/> wkp:placeName ?name
}
WHERE {
  ?x gn:name ?name
}
```

You can also reshape the data.

Here we want to map our data to another system  
So we need to use their term instead.

\*

This query creates a new RDF graph that uses the new term instead of the old one.

# Reshape data

```
prefix gn: <http://www.geonames.org/ontology#>
prefix wkp: <http://www.wikipedia.org/ontology#>
```

```
CONSTRUCT {
  <http://sws.geonames.org/2735943/> wkp:placeName ?name
}
WHERE {
  ?x gn:name ?name
}

<http://sws.geonames.org/2735943/> wkp:placeName "Porto" .
```

You can also reshape the data.

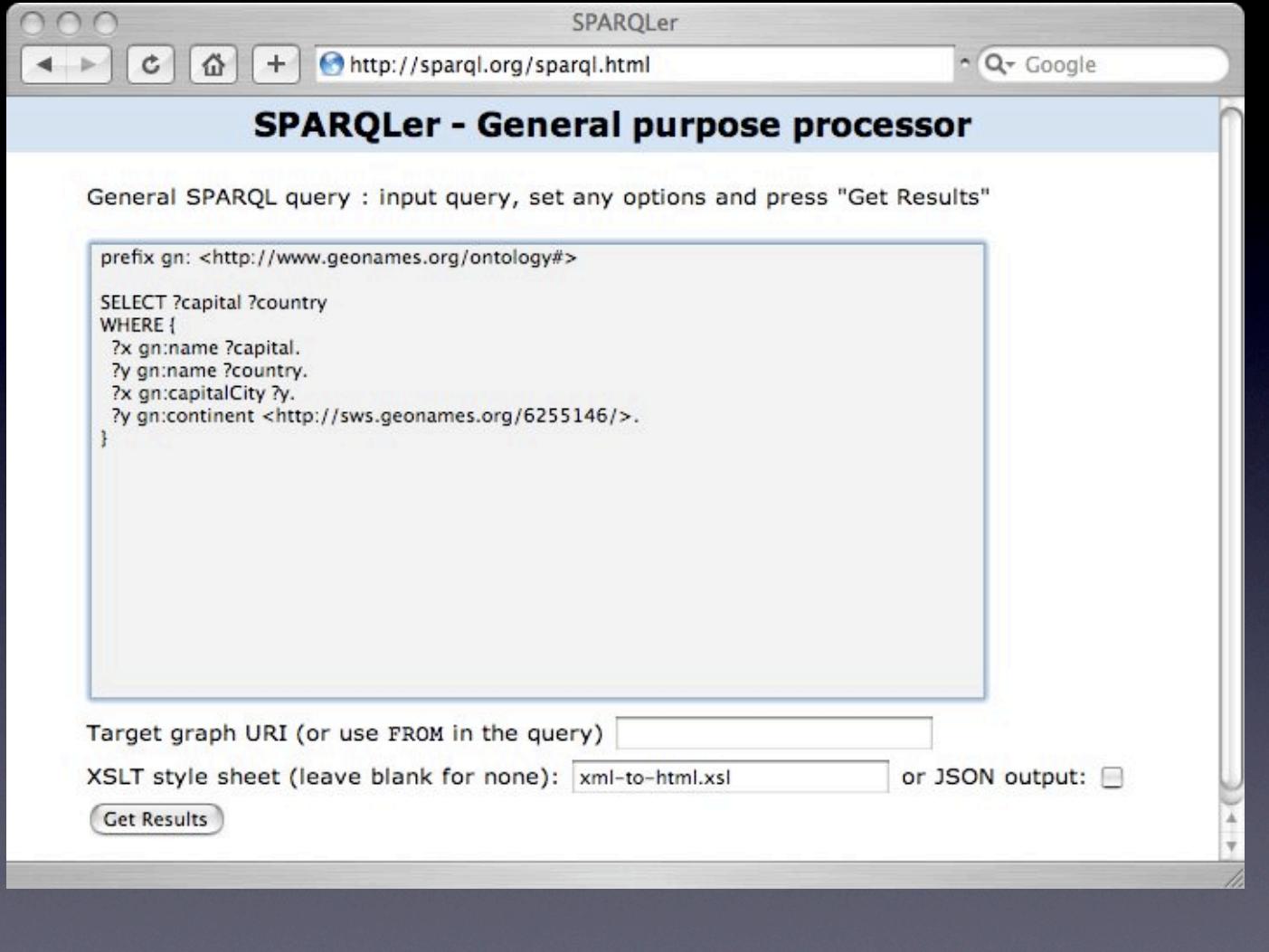
Here we want to map our data to another system  
So we need to use their term instead.

\*

This query creates a new RDF graph that uses the new term instead of the old one.

# SPARQL Protocol

The SPARQL standard is more than just the query language. The SPARQL standard also defines a protocol for executing SPARQL queries over the web.



This is defined as a RESTful web service interface.

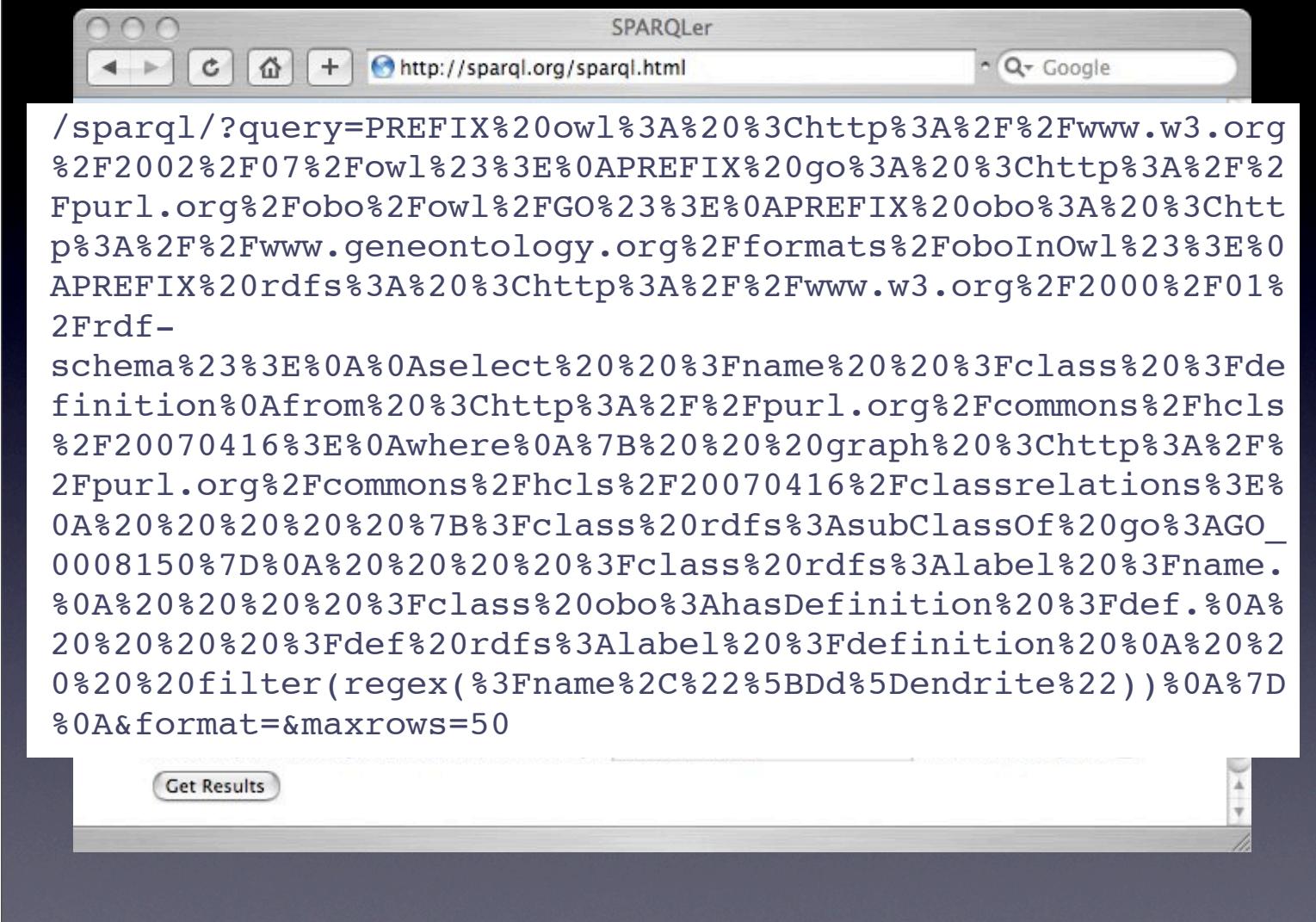
Basically, it defines a standard HTML form for doing queries, like on a normal web page.

\*

But one that computer programs can also access!

And there are standard way of presenting querying results.

So you can build SPARQL applications and just point them at a dataset!



This is defined as a RESTful web service interface.

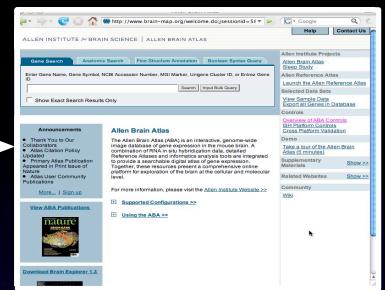
Basically, it defines a standard HTML form for doing queries, like on a normal web page.

\*

But one that computer programs can also access!

And there are standard way of presenting querying results.

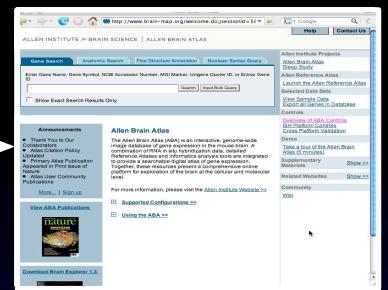
So you can build SPARQL applications and just point them at a dataset!



And this frees the data!

...

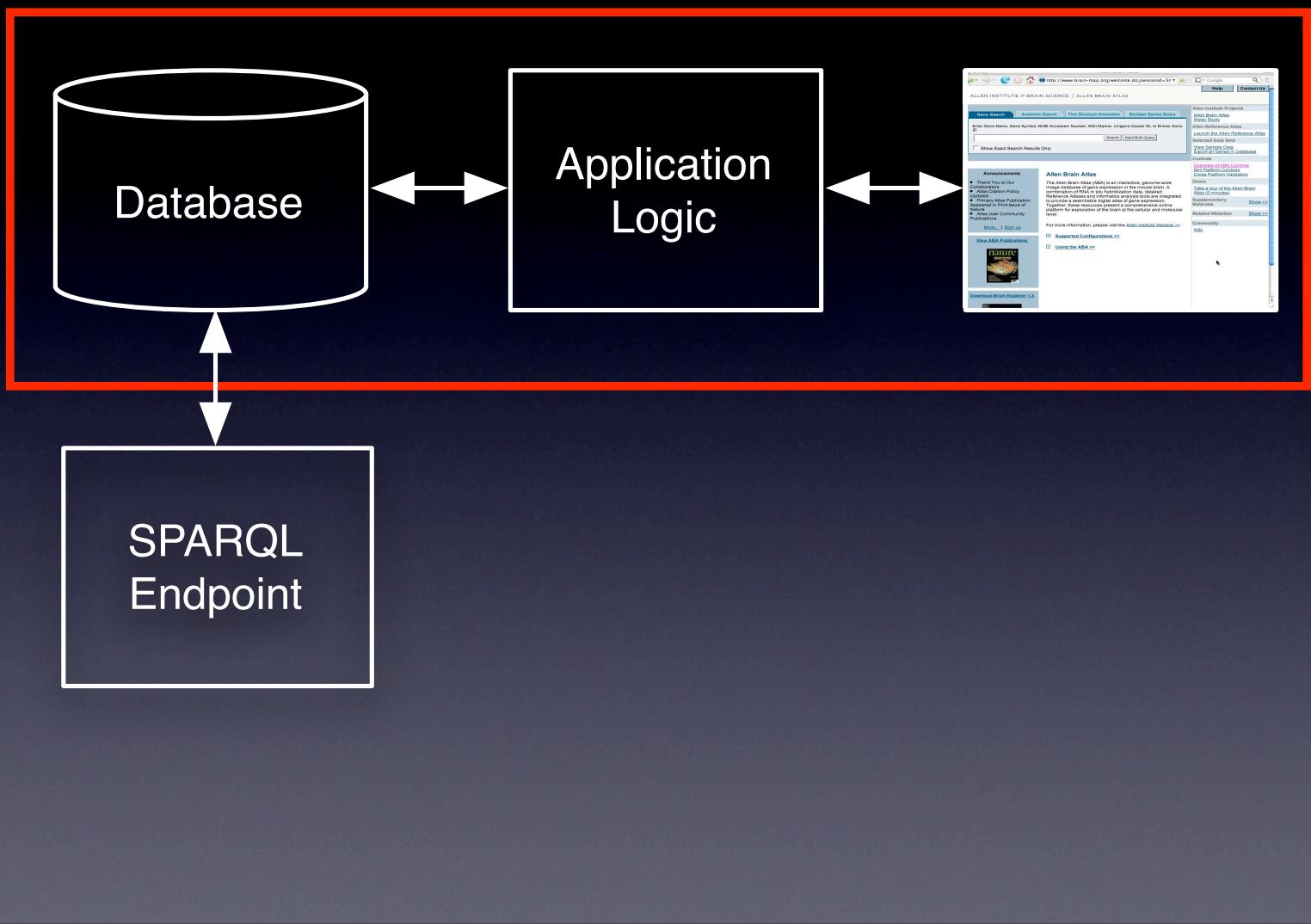
combine different sparql end points



And this frees the data!

...

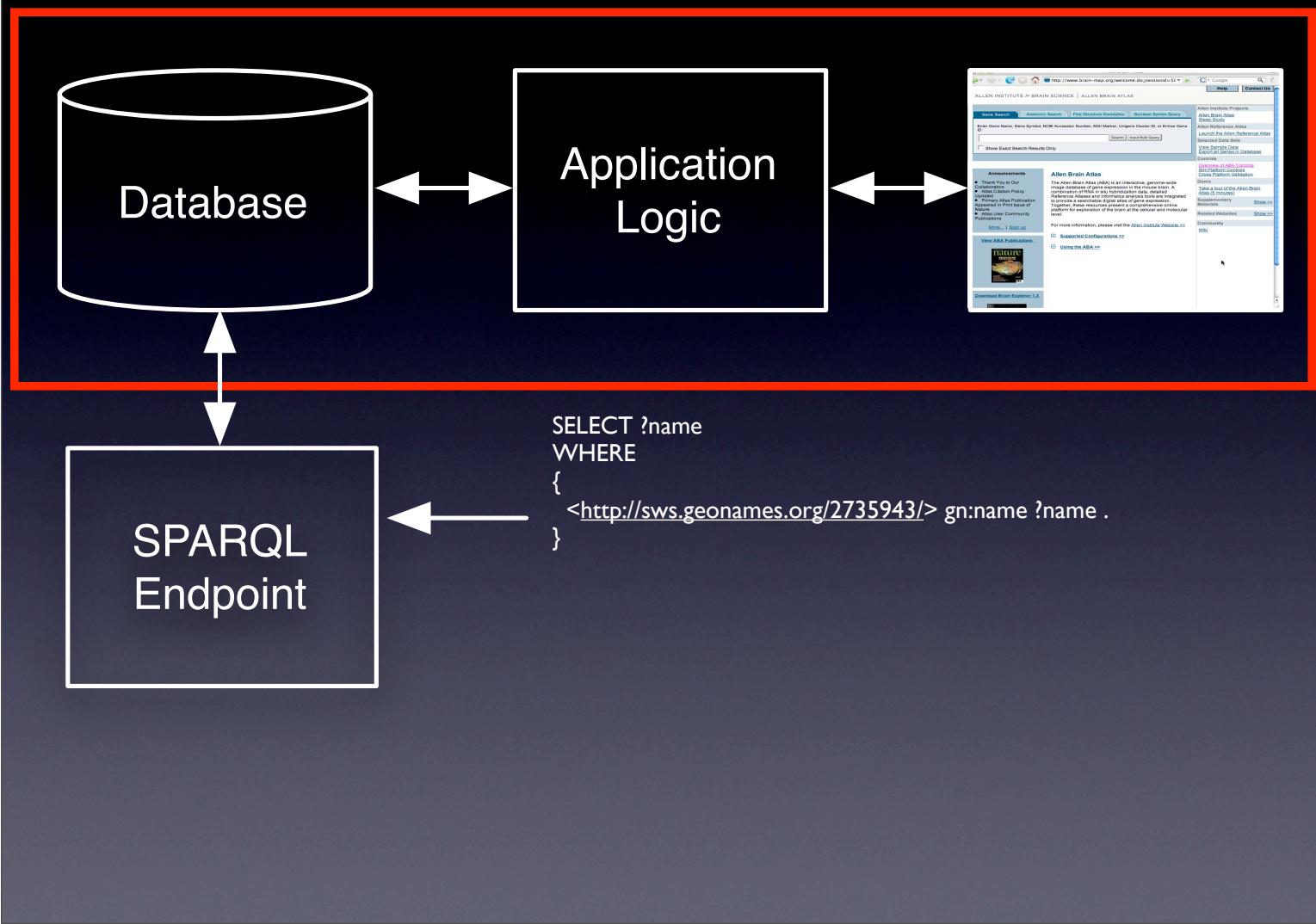
combine different sparql end points



And this frees the data!

...

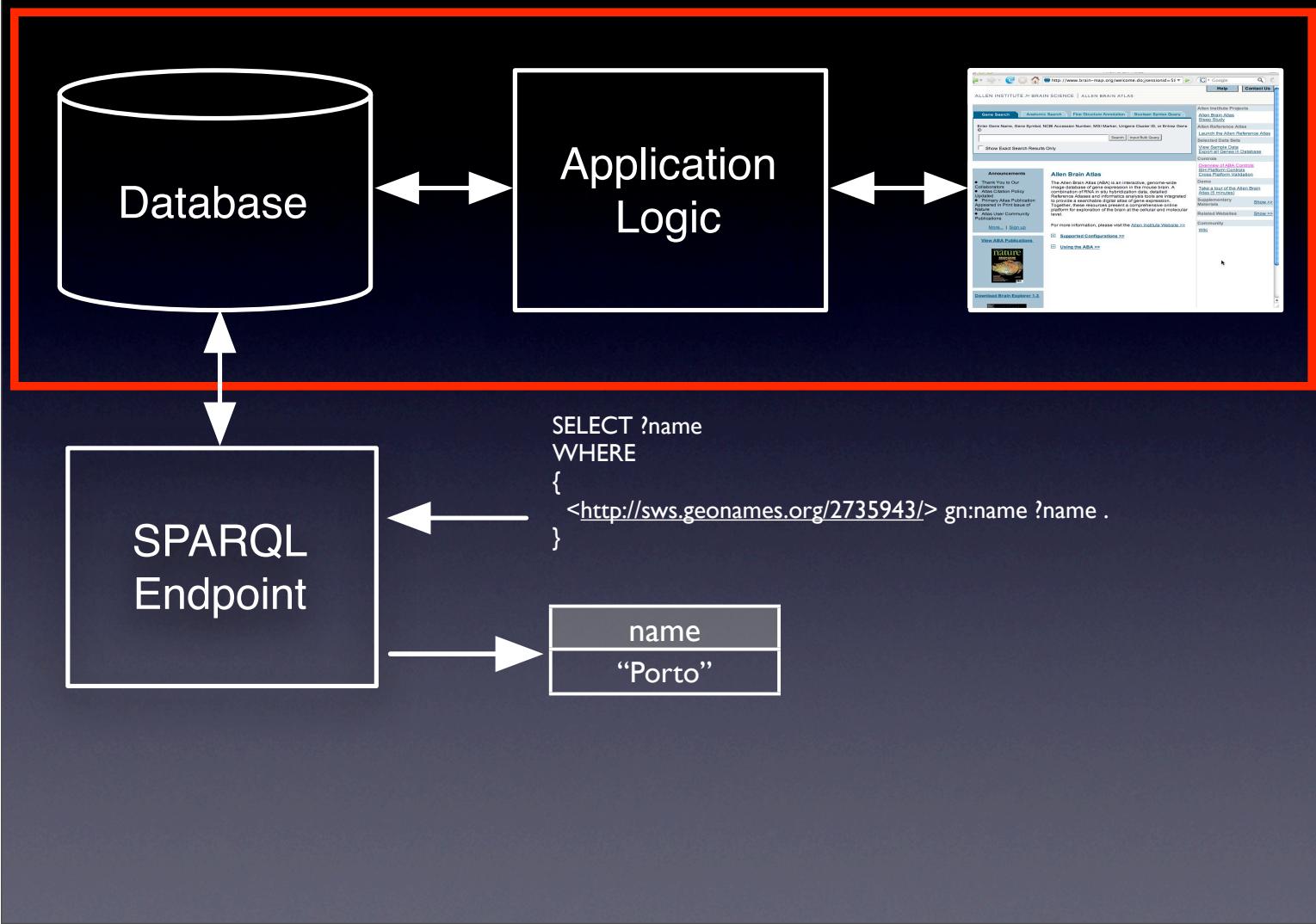
combine different sparql end points



And this frees the data!

...

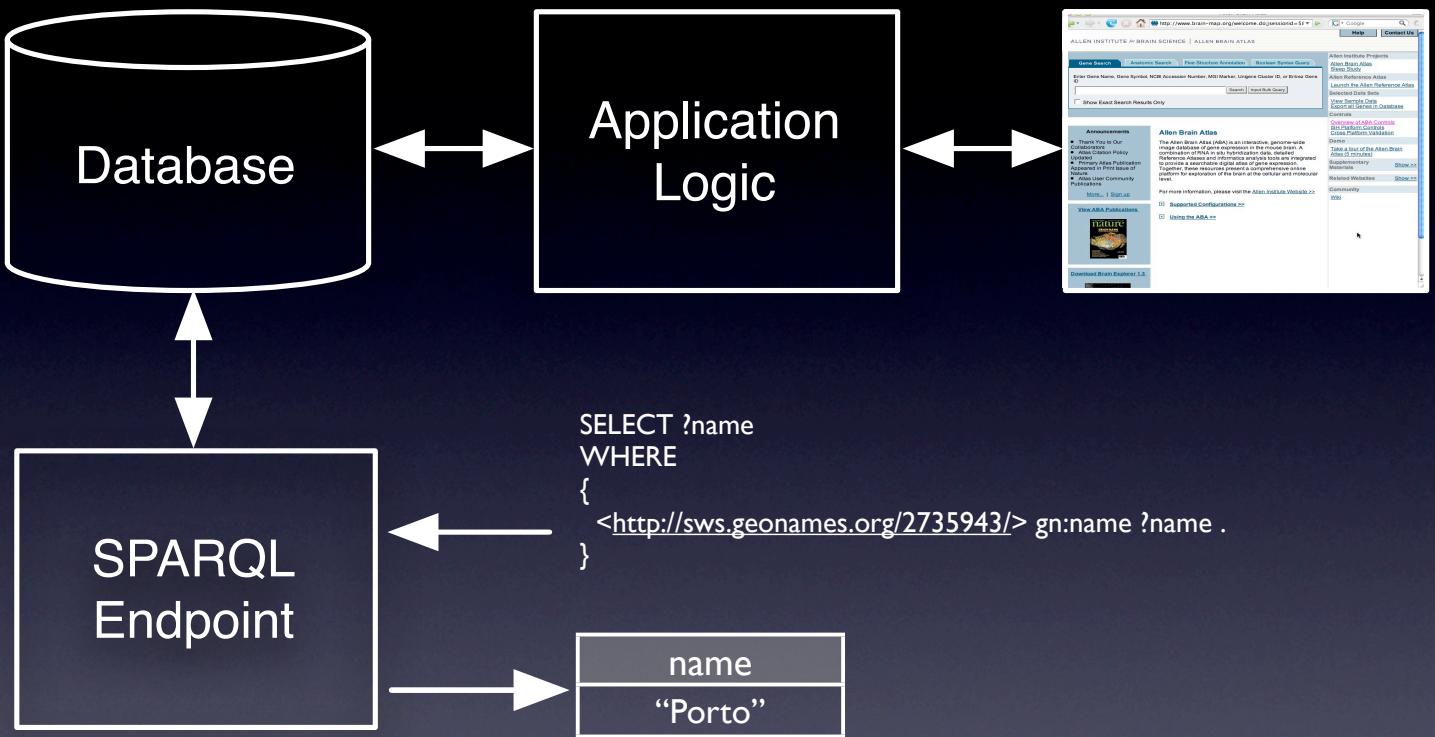
combine different sparql end points



And this frees the data!

...

combine different sparql end points



And this frees the data!

...

combine different sparql end points

# Semantic Web Case Studies

# Benefits of Semantic Web to Life Sciences

- Fusion of data across many scientific disciplines
- Easier recombination of data
- Querying of data at different levels of granularity
- Capture provenance of data through annotation
- Perform inference across data sets
- Machine processable approach
- Data can be assessed for inconsistencies

# Allen Brain Atlas Semantic Mashup

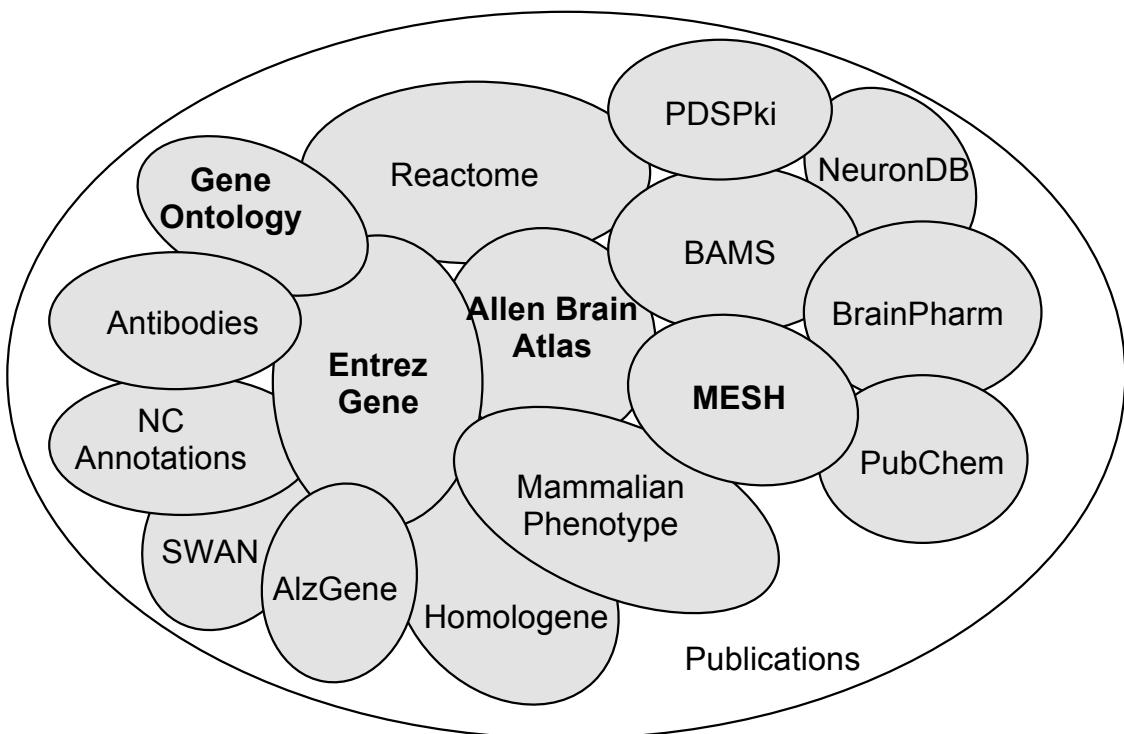
<http://esw.w3.org/topic/HCLS/Banff2007Demo>

For example, remember the Allen Brain Atlas.

There's been work on extracting the information from the site, and converting it into RDF.

# Integration and analysis of heterogeneous data sets

Hypothesis, Genome, Pathways, Molecular Properties, Disease, etc.



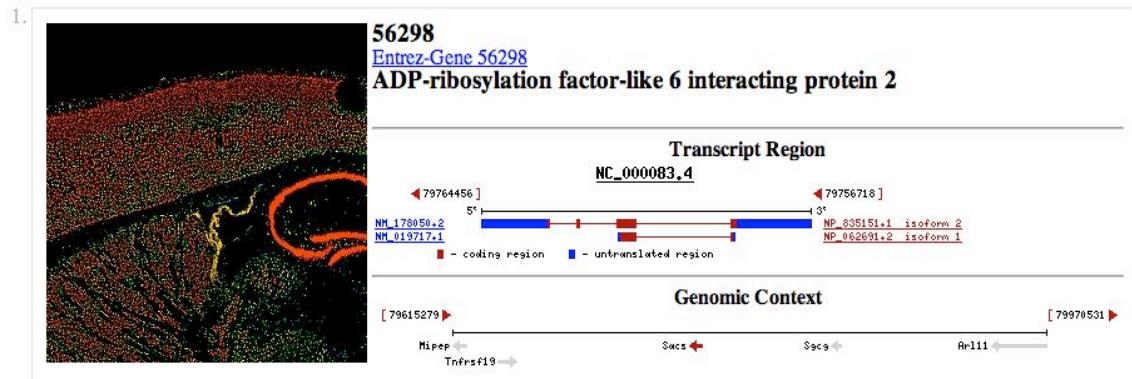
So now you can query the data using SPARQL and combine it with other data sets!

# Allen Brain Atlas Gene Expression Results

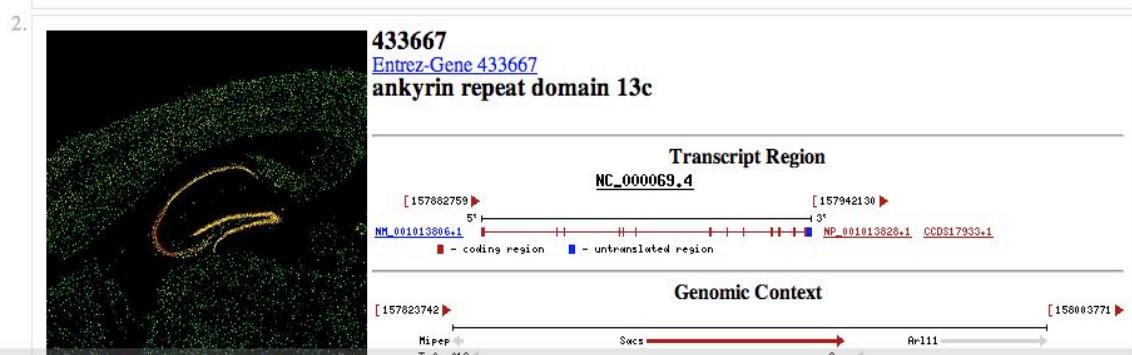
3 entrez-gene-record filtered from 26 originally (reset)

Copy All

sorted by: hasName; then by... •  grouped as sorted •  show duplicates



hasName 3✓  
deregulated  
1b  
homolog  
(C.  
elegans)  
1✓ apoptotic  
peptidas  
activatin  
factor 1



And you can take the data and present it in different user interfaces.

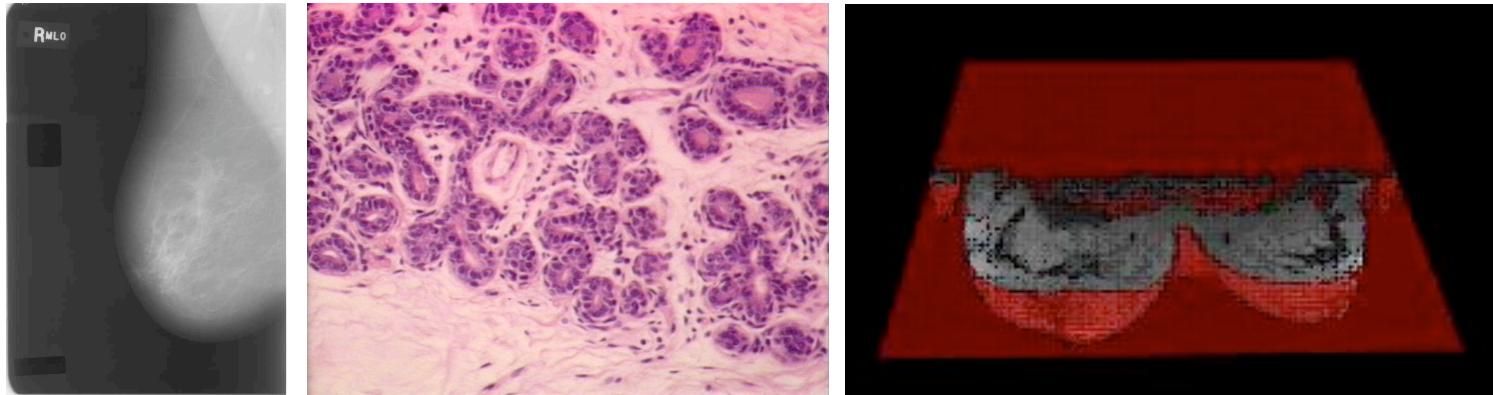


<http://www.aktors.org/miakt/>

---

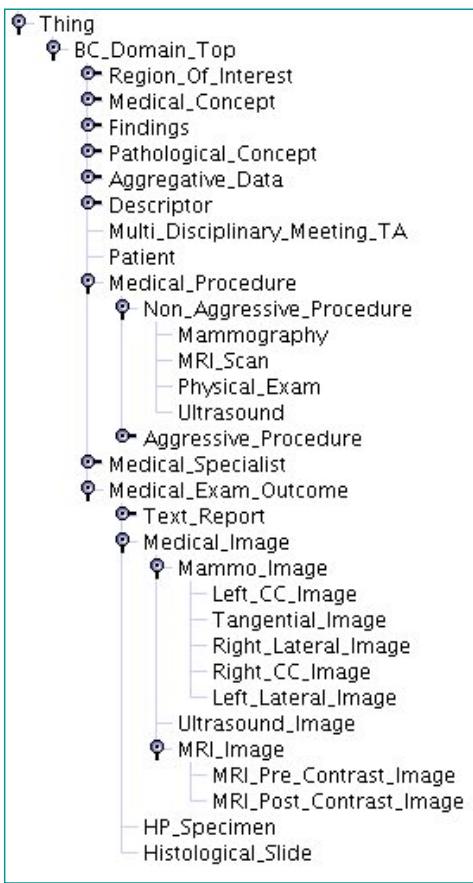
MiAKT was a project Southampton was involved in.

# Assistance in breast cancer screening and diagnosis



And we were using semantic web technologies to assist in breast cancer screening and diagnosis.

# Semantic Web techniques for managing information

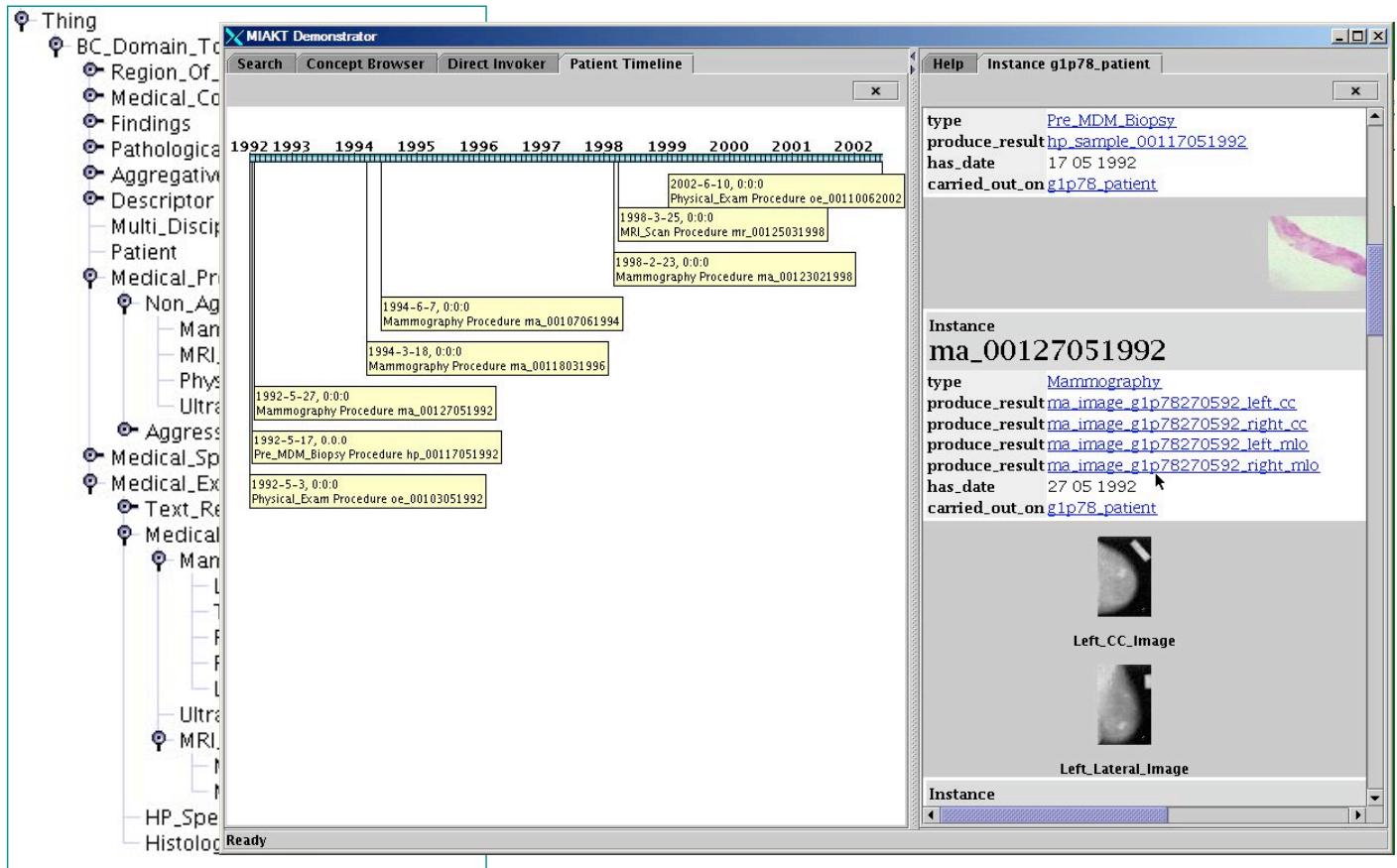


We defined an ontology, and stored all of our data using RDF.

Shared information between different specialists, that incorporated data from different systems.

It also let us develop interesting user interfaces, for example using a timeline to show how a case developed.

# Semantic Web techniques for managing information

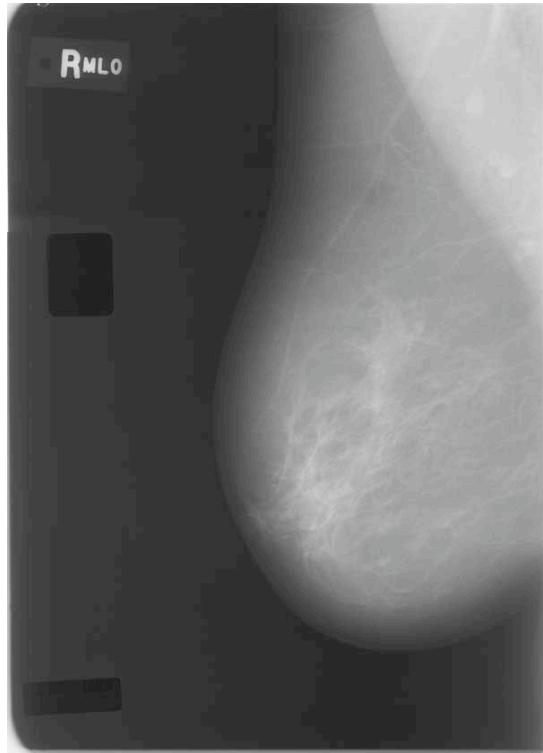


We defined an ontology, and stored all of our data using RDF.

Shared information between different specialists, that incorporated data from different systems.

It also let us develop interesting user interfaces, for example using a timeline to show how a case developed.

# Linking content features to semantic annotations

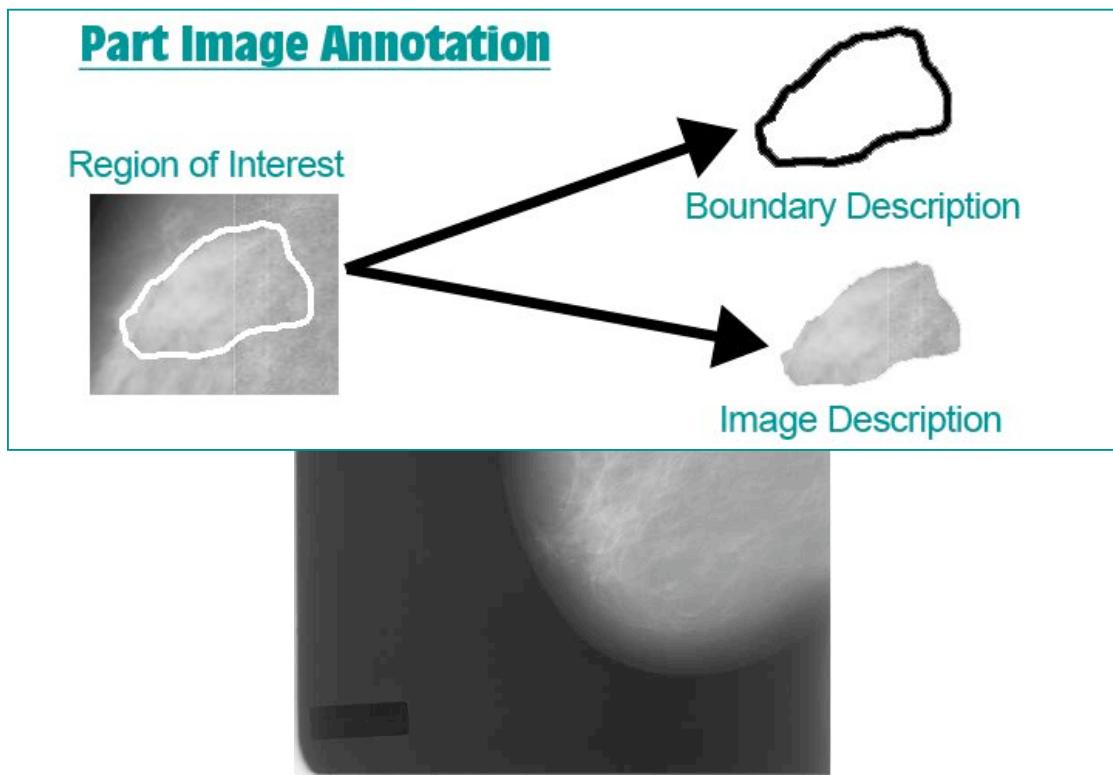


We also did computer vision analysis on the mammography data.

- \* So we might let users annotate areas of interest.
- \* And then we link these areas of interest to semantic annotations automatically.

All using RDF – you can mix and match all kinds of data!

# Linking content features to semantic annotations

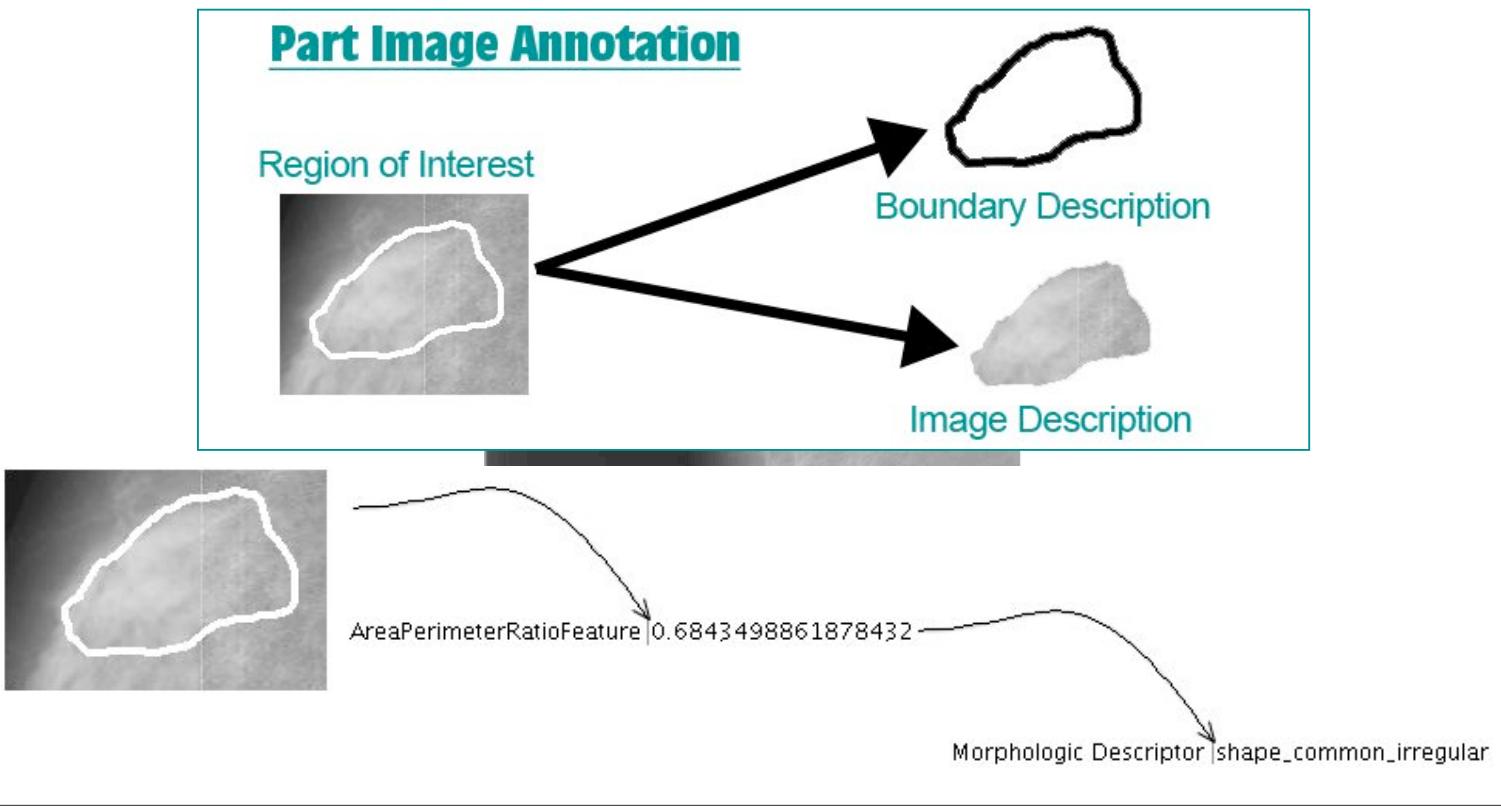


We also did computer vision analysis on the mammography data.

- \* So we might let users annotate areas of interest.
- \* And then we link these areas of interest to semantic annotations automatically.

All using RDF – you can mix and match all kinds of data!

# Linking content features to semantic annotations



We also did computer vision analysis on the mammography data.

- \* So we might let users annotate areas of interest.
- \* And then we link these areas of interest to semantic annotations automatically.

All using RDF – you can mix and match all kinds of data!

# Classification of patient cases using machine learning

<b>Get Suggested Classification</b>	<b>malignant (69.45231676%)</b>
<b>Instance</b>	
<b>00011_right_cc_abnor_1</b>	
<b>type</b>	<a href="#">Mammo_ROI</a>
<b>has_low_level_image_feature</b>	<a href="#">contour_1093944759974_1</a>
<b>has_morph_feature</b>	<a href="#">shape_common_irregular</a>
<b>has_morph_feature</b>	<a href="#">margin_common_spiculated</a>
<b>is_finding</b>	<a href="#">mass_00011_right_cc_abnor_1</a>

And we could use these semantic annotations with machine learning techniques to perform aspects of the patients' diagnosis automatically.

# Natural language generation of patient reports

```
<rdf:Description rdf:about='http://www.aktors.org/miakt/ontologies/bcio/  
Breast_Cancer_Imaging.daml#00021_patient'>  
    <rdf:type rdf:resource='http://www.aktors.org/miakt/ontologies/bcio/  
Breast_Cancer_Imaging.daml#Patient'/'>  
        <NS2:has_age>72</NS2:has_age>  
        <NS2:involved_in_ta rdf:resource=  
Breast_Cancer_Imaging.daml#ta-soton-10  
2128853'/'>  
</rdf:Description>
```

## Automatically Generated Patient Record

The 65 years old patient is involved in a triple assessment procedure. The triple assessment procedure contains a mammography exam. The mammography exam on a 15 10 1992 produced a right lateral image, a left lateral image, a right CC image , and a left CC image. The mammography exam is carried out on the patient. The left CC image has a left lateral side and a craniocaudal view. The left lateral image has a left lateral side and an oblique mediolateral view. The right CC image contains an abnormality and the right CC image has a right lateral side and a craniocaudal view. The abnormality has an irregular shape, a definitely malignant assessment , and a spiculated margin. The right lateral image contains an abnormality and the right lateral image has a right lateral side

And we used the structured RDF data to automatically generate natural language patient reports!



# Challenges

[http://flickr.com/photos/good\\_day/60867797/](http://flickr.com/photos/good_day/60867797/)

# Public Perception

It's taken a while for people to understand what the semantic web is doing.

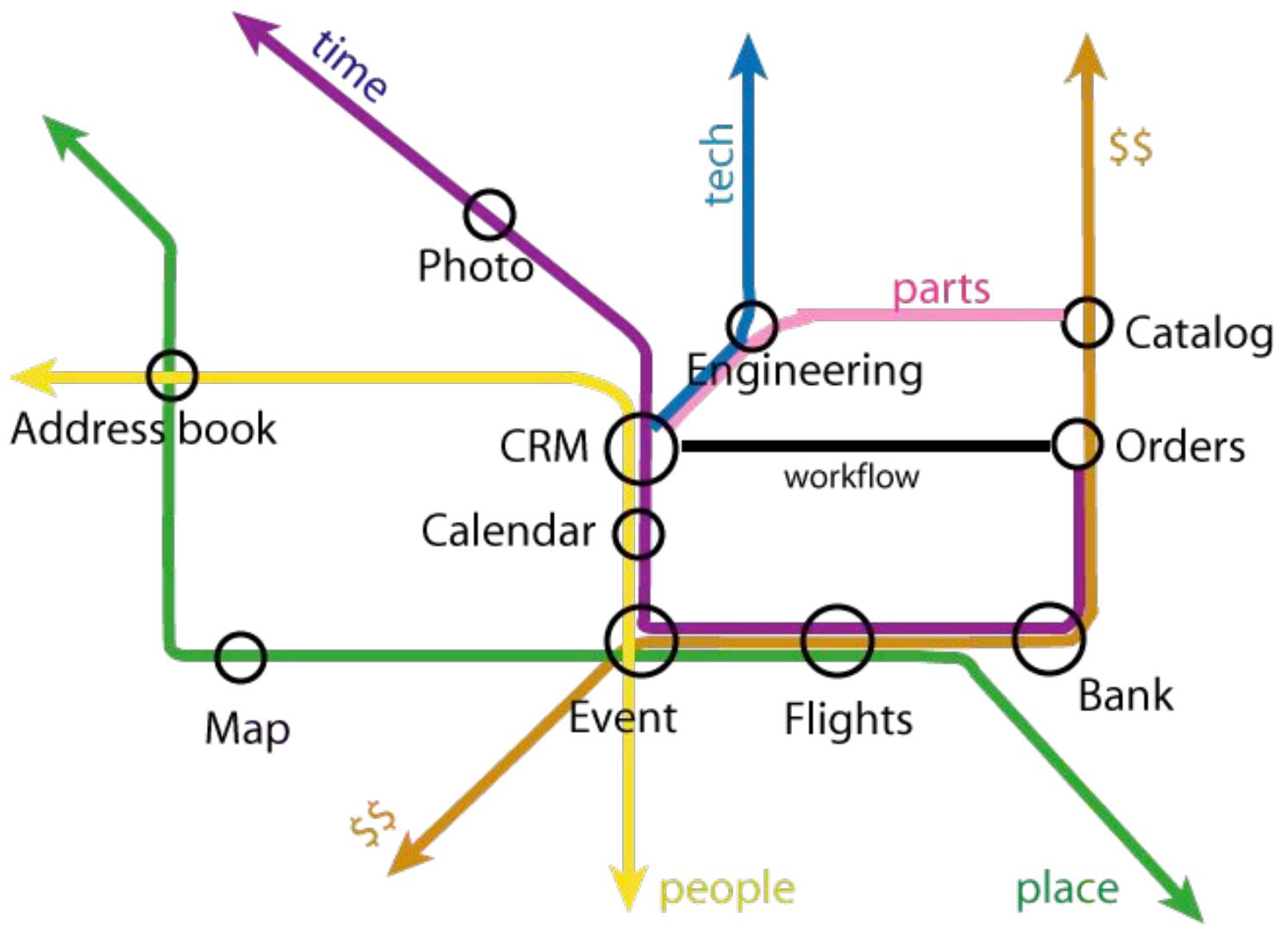
There are lots of comments on how the semantic web forces a top-down view.

That you have a big brother controlling everything!  
And that you will never get everyone to agree on the same description.

But that's the complete opposite of what it's doing – you define whatever data set you want

It's not top-down,  
it's bottom up!

RDF has been around for a long time – as long as XML  
It's just that it's a solution for a problem people weren't aware they needed!  
Solves the problem of describing data over the web... (stefano?)  
e.g. XML is everywhere: describe data in a



<http://www.w3.org/DesignIssues/diagrams/SemWebAppMetro.png>

This image shows how the semantic web might be used in an application.

It's drawn like the London tube – metro – map.

Each train line represents an ontology or domain.

Each station shows how an ontology can be used in the application.

For example, with place, you can show a map.

But you get intersections: place and people intersect for dealing with address information.

So this shows that the semantic web is not a single view of the world. It's lots of different models all playing together!



Home | What's on | What's Joost? | Joost beta | Share | Forum | Support

Try out Joost beta – Now For Friends

[How to get an invite](#)

### The new way of watching TV

All the things you love about TV, fused with the interactive power of the internet – just the way you want it. Enjoy the ride!



Countdown to Joost™

Be first to know when Joost goes live



#### What's on

Showtime! See what's playing at Joost.



#### What's Joost?

In short: it's TV, just the way you want it. See why inside.



#### The Joost blog

What are these guys thinking?  
Find out here.

People are beginning to launch serious applications backed by semantic web technologies too.

Joost is a new service that offers TV over the internet.

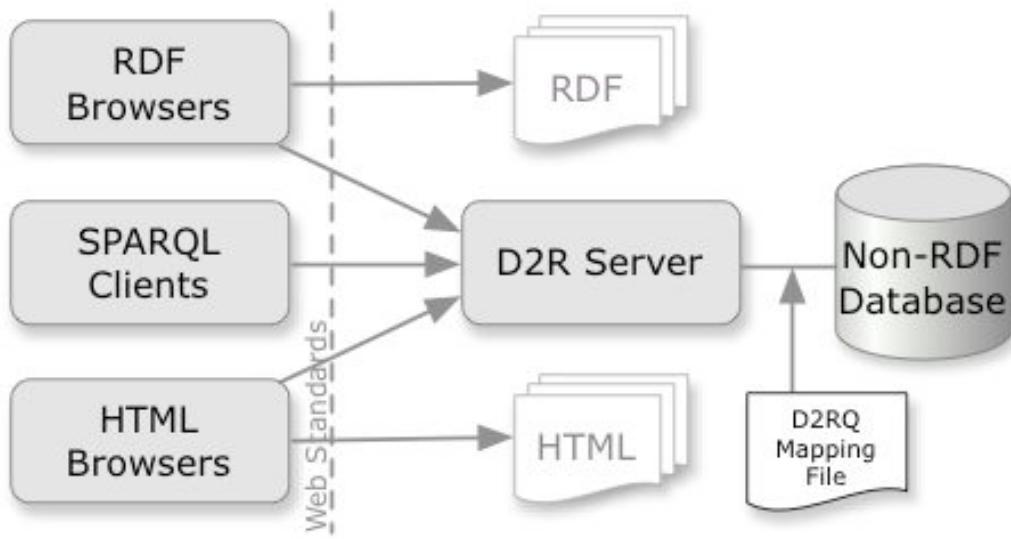
They're using semantic web tools to integrate information about the TV shows they offer.

This lets them deal with things like different regulations imposed by different companies and countries in a very flexible way.

# Scarcity of semantically annotated information

There is lots of great progress being made, but still the majority of information is not available in a semantic web format.

# D2R Server: mapping databases to semantic web



<http://sites.wiwiss.fu-berlin.de/suhl/bizer/d2r-server/>

Things are getting better though.

D2R Server is a really great bit of technology.  
It allows you to map a relational database to RDF.

So people are beginning to use this to make their existing data sets available on the semantic web.

# dbpedia.org

The DBpedia project is extracting structured information from the Wikipedia project and making it available in RDF.

Wikipedia is a huge source of information, and having some of that available in RDF will hopefully allow people to build some nice things!

"Who is wise? He that learns from everyone."

Benjamin Franklin



## Introducing Freebase.

Freebase is an open, shared database of the world's knowledge.



"Somersault." by netjer-lelahell flickr

### Free + Database = Freebase

It's about film, sports, politics, music, science and everything else all connected together. Our contributors are collecting data from all over the internet to build a massive, collaboratively-edited database of cross-linked data. It's a big job and we're just getting started.

### Share, reuse, remix.

We want to make it possible for you to add high quality structured information to your websites, mashups and applications without worrying about restrictive licenses. All data is licensed Creative Commons Attribution. We only ask that you link back to us.



There are other people trying to do the same thing as the semantic web.

Freebase is a startup company that is trying to create an open source knowledge base, but with a focus on more structured data.

They aren't using semantic web technology to do this, but you can map to and from the different formats quite easily.

Google and Microsoft also have similar initiatives.

# Performance and Scalability

Compared to optimised relational databases, RDF stores are slower and have trouble handling extremely large data sets.

This is because a relational database explicitly defines the structure of the data. So you can optimise the data structures using techniques like indexing.

RDF stores are improving, and over the few months I've seen lots of great progress. For example, Franz Allegrograph can deal with billions of triples.

But this only applies to storing and basic querying.  
If you want to do clever reasoning stuff over complex ontologies then performance is a big problem.

# Provenance

Provenance is the origin or source or where something comes from.

For performing experiments with semantic web data, it is important to know where knowledge has come from and how it has been processed.

It's useful to know who believes something and why.

There's been loads of great work in this area, including various research projects at Southampton. But there is no standard way to represent this information.

So the existing Semantic Web tools don't really support provenance information.

# Ontology Mapping

The big challenge is in ontology mapping.

As I've shown, the Semantic Web techniques allow us to share data in a standard way.

The assumption is that every data set and ontology will interoperate.

But in reality different representations of the same data exist.

With many of the data integration examples I've shown, it's been based on taking existing data and converting it to RDF.

When you have completely different types of data, with well established connections between the two, this approach can work very well.

But this relies on there being either:

- a shared vocabulary
- mappings between the different ontologies

Mapping between ontologies is hard work!

You have to work out which terms match in the different ontologies

This is not straightforward because different datasets describe the same things from a slightly different perspective!

# The End!

# Four Simple Rules

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 information.
4. Include links to other URIs so that they can discover more things.

# Resources

- W3C Health Care and Life Sciences  
<http://www.w3.org/2001/sw/hcls/>
- Semantic Web FAQ  
<http://www.w3.org/2001/sw/SW-FAQ>
- RDF Primer  
<http://www.w3.org/TR/rdf-primer/>

# a little semantics goes a long way!



<http://www.metade.org/blog>