

The Semantic Web and Linked Data

Globally Unique Identifiers

SPARQL - how to explore Linked Data

A basic beginners guide

Mark Wilkinson
CBGP UPM-INIA
Universidad Politécnica de Madrid
Spain

Resource Description Framework

RDF is a “framework” - a way of conceptualizing how to represent data and knowledge. There are a variety of ways to implement this conceptualization.

It is ~NOT a language (per se...)

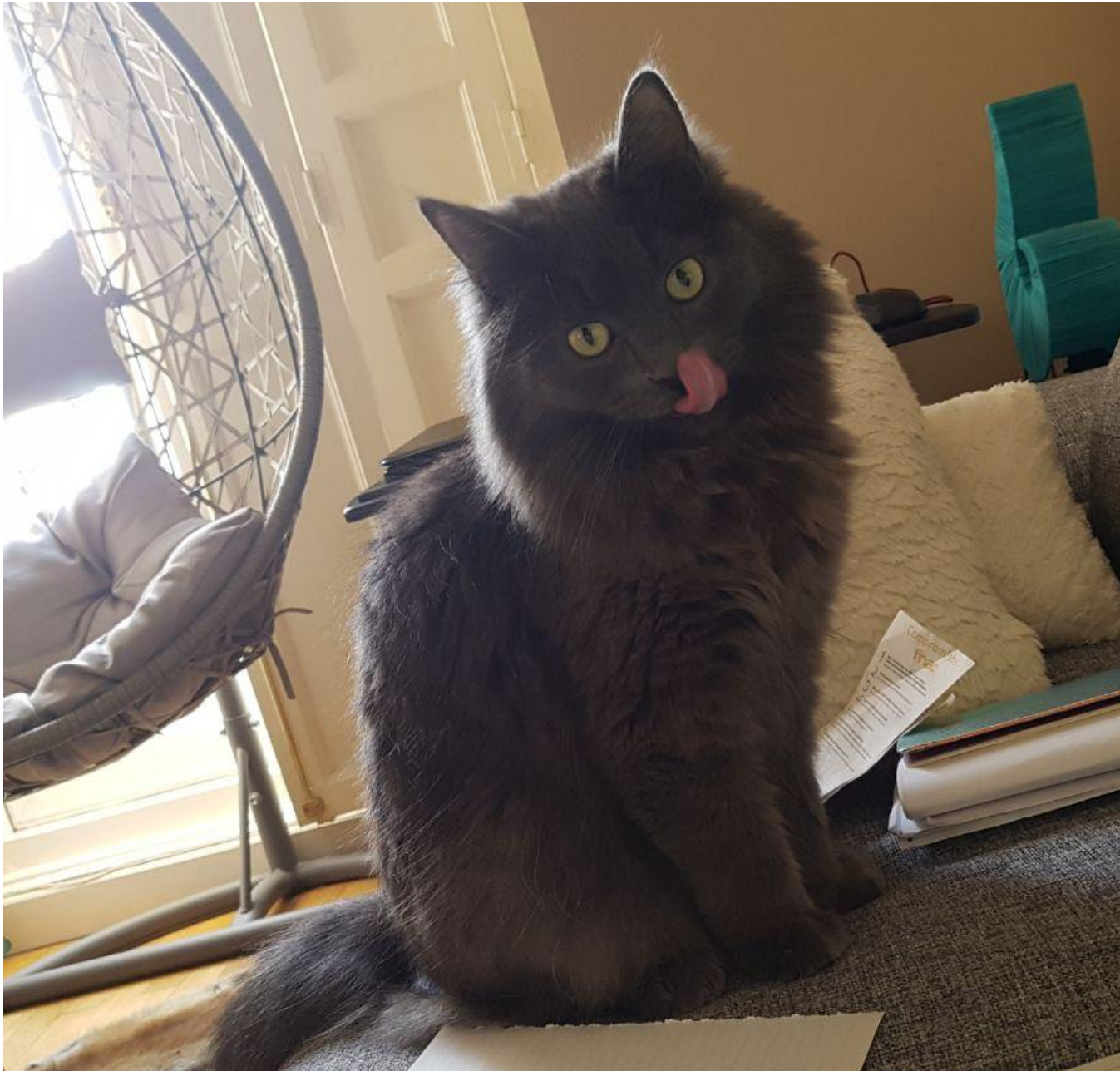
It is NOT a file-format or data syntax (though its implementations generally are...)

Resource Description Framework

RDF says that all data and knowledge can be represented through a combination of very simple statements called “Triples”

Much like human language, the statements take the form:

Subject Predicate Object



“Triples”

S P O

Mark owns Cat

“Triples”

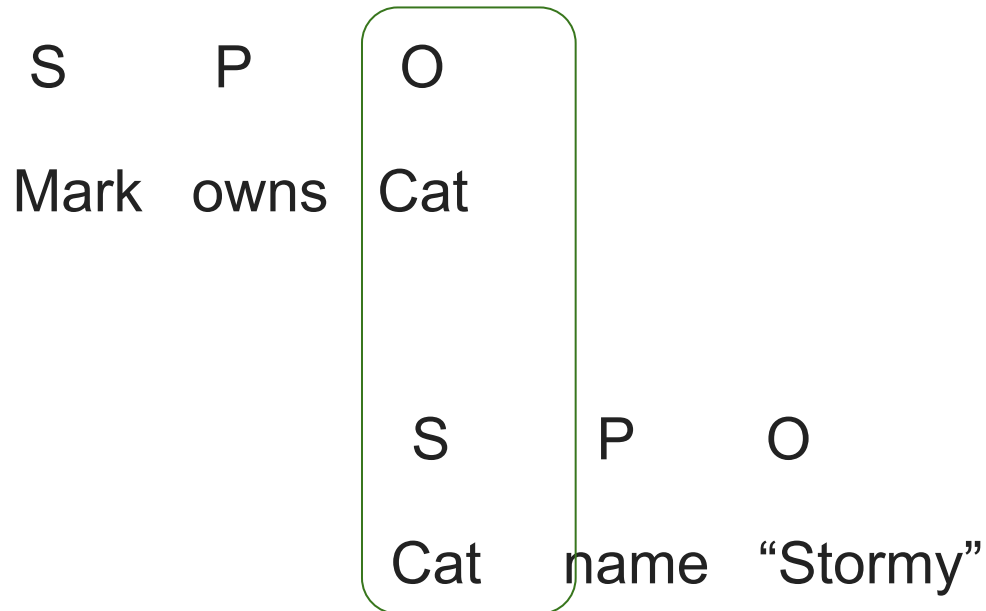
S P O

Mark owns Cat

Cat name “Stormy”

Resource Description Framework

“Triples”



The “Object” of one triple becomes the “Subject” of another triple

Resource Description Framework: “Triples”

Mark owns Cat

Cat name “Stormy”

Cat breed “British Long”

residence Madrid

Madrid population “3,1 Mil”

Resource Description Framework

“Graph”

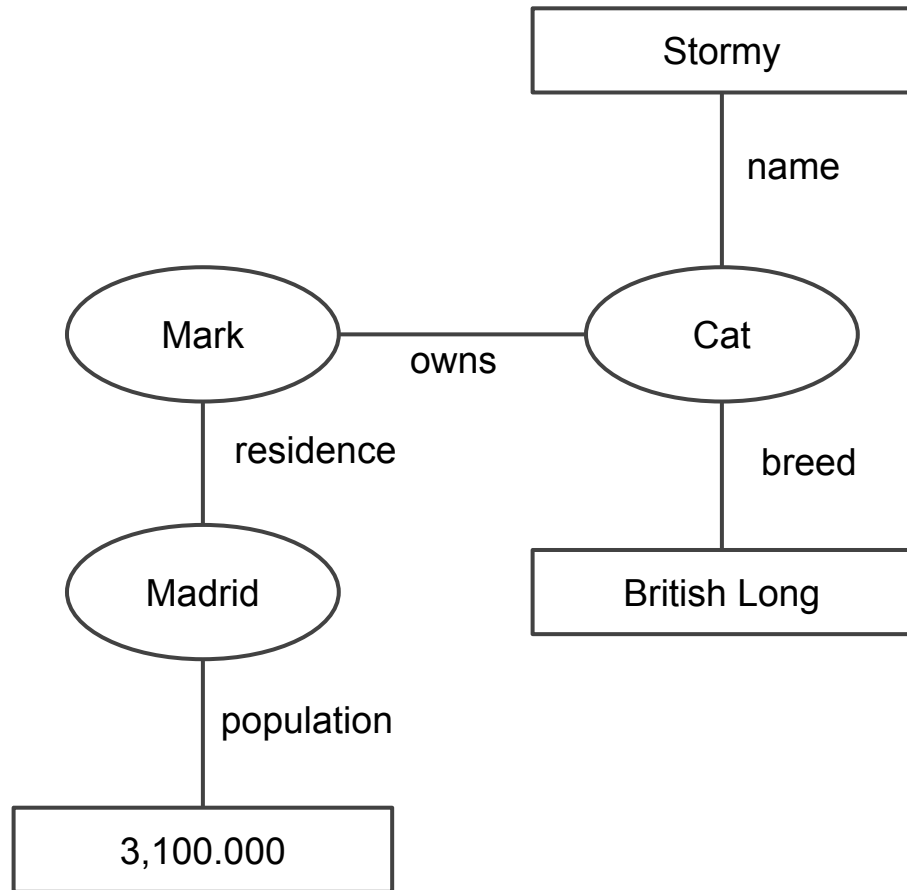
Mark owns

Cat name “Stormy”

Cat breed “British Long”

residence

Madrid population “3,100,000”



Resource Description Framework

“Resources” “Literals” “Predicates”

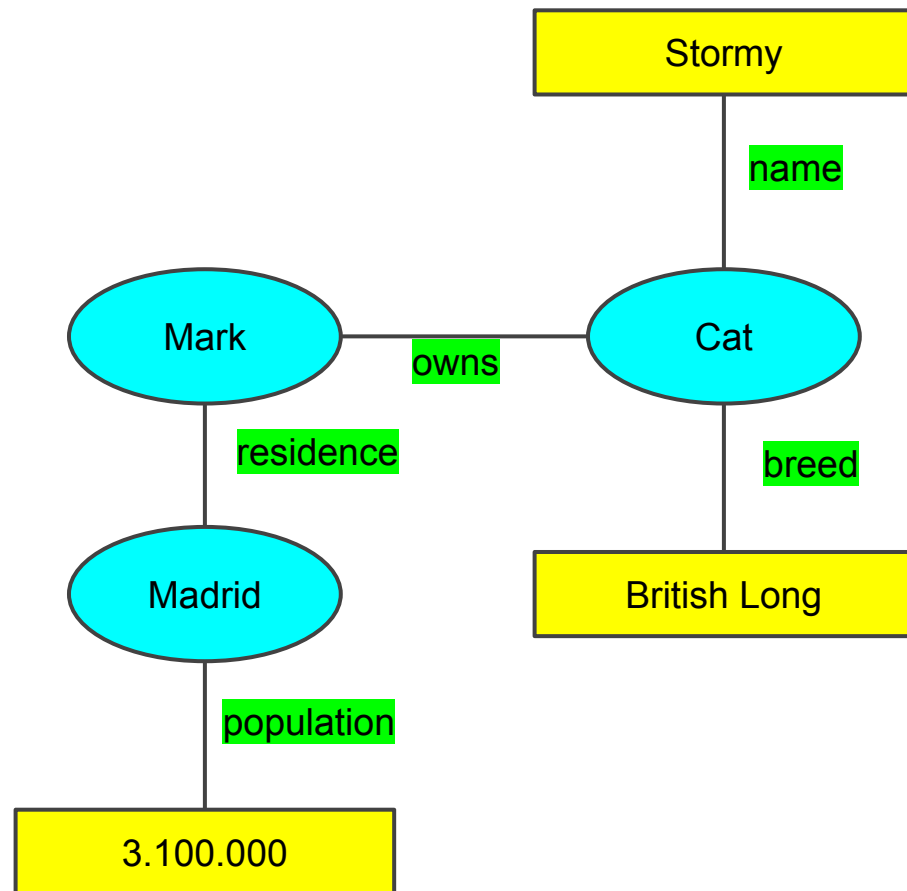
Mark owns

Cat name “Stormy”

Cat breed “British Long”

residence

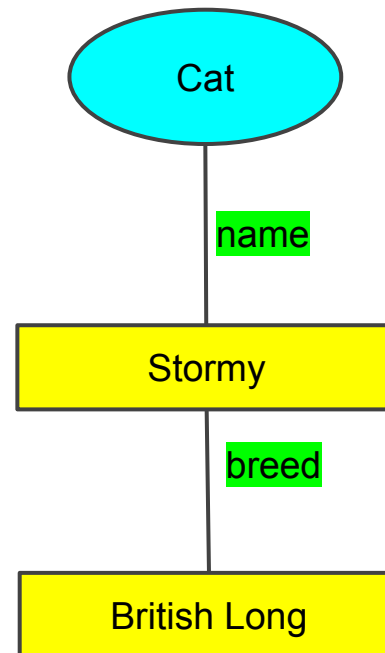
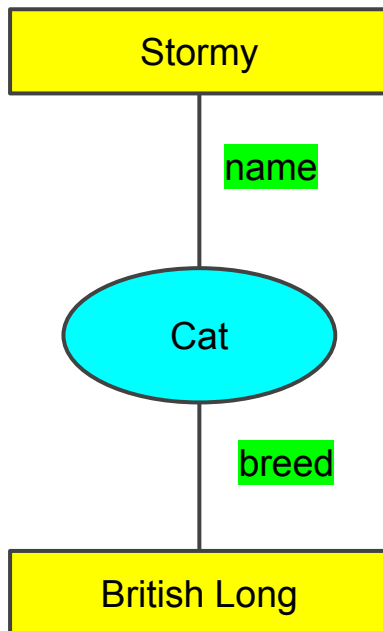
Madrid population “3.100,000”



RDF Rules

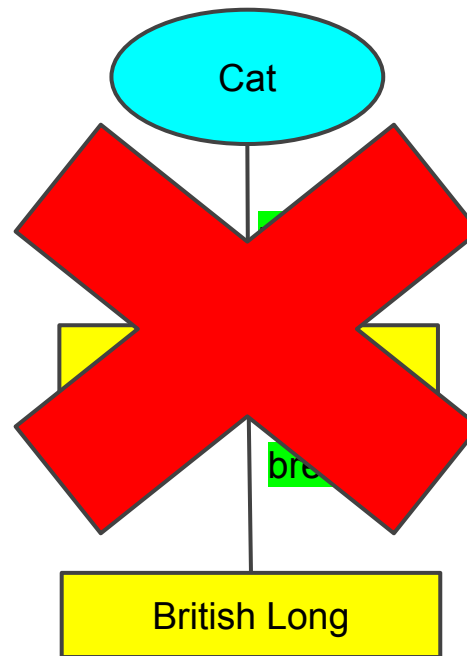
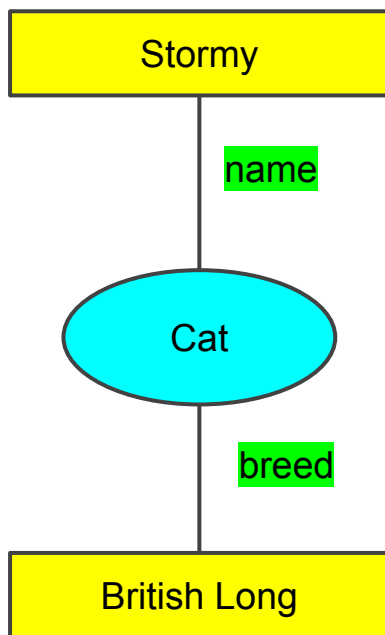
There is nothing smaller than a Triple

Only a **Resource** can be the subject of a triple



There is nothing smaller than a Triple

Only a **Resource** can be the subject of a triple





Globally Unique Identifiers (GUIDs)

There are two “flavours” of GUID

- 1) Those that are globally unique based on statistical (un)likelihood of a collision
- 2) Those that are globally unique based on convention/agreement

Official UUID definition

Network Working Group
Request for Comments: 4122
Category: Standards Track

P. Leach
Microsoft
M. Mealling
Refactored Networks, LLC
R. Salz
DataPower Technology, Inc.
July 2005

A Universally Unique Identifier (UUID) URN Namespace

Status of This Memo

This document specifies an Internet standards track protocol for the Internet community, and requests discussion and suggestions for improvements. Please refer to the current edition of the "Internet Official Protocol Standards" (STD 1) for the standardization state and status of this protocol. Distribution of this memo is unlimited.

Copyright Notice

Copyright (C) The Internet Society (2005).

Grains of sand on earth: 75,000,000,000,000,000,000

Number of possible GUIDs: 340,282,366,920,938,463,463,374,607,431,770,000,000

The Point:

If you use a GUID to identify something, there is an infinitesimally small (~0%) chance that someone else has used that same GUID to name something else

THEREFORE! (the point!)

You can safely merge your dataset with someone else's dataset, and not "collide"

There are two “flavours” of GUID

- 1) Those that are globally unique based on statistical (un)likelihood of a collision
- 2) Those that are globally unique based on convention/agreement

“Conventions” can be used to ensure that a set of characters is globally unique

For example, DOIs and Handles

A governing agency assigns a prefix (e.g. 10.xxxx for DOIs) to an identifier-generating authority

That prefix can only be (meaningfully) used to generate identifiers within that authority.

Unless the authority is mischievous, they can ensure that they never re-use an identifier, and nobody else “is allowed to” create an identifier with that prefix.

“Conventions” can be used to ensure that a set of characters is globally unique

URLs are another example

<http://www.wilkinsonlab.info>

I own that!

I create identifiers within the ***wilkinsonlab.info***
“namespace”

I would be crazy to create the same address for two
different pages!

(and in fact, *I cannot*, as you intuitively know)

“Conventions” can be used to ensure that a set of characters is globally unique

The only time I *could* experience a “collision” with someone else’s name is if they are using the string of characters

“http://wilkinsonlab.info”

to mean something other than the identifier for my website (which would be crazy)

So... Web addresses are pretty good GUIDs!

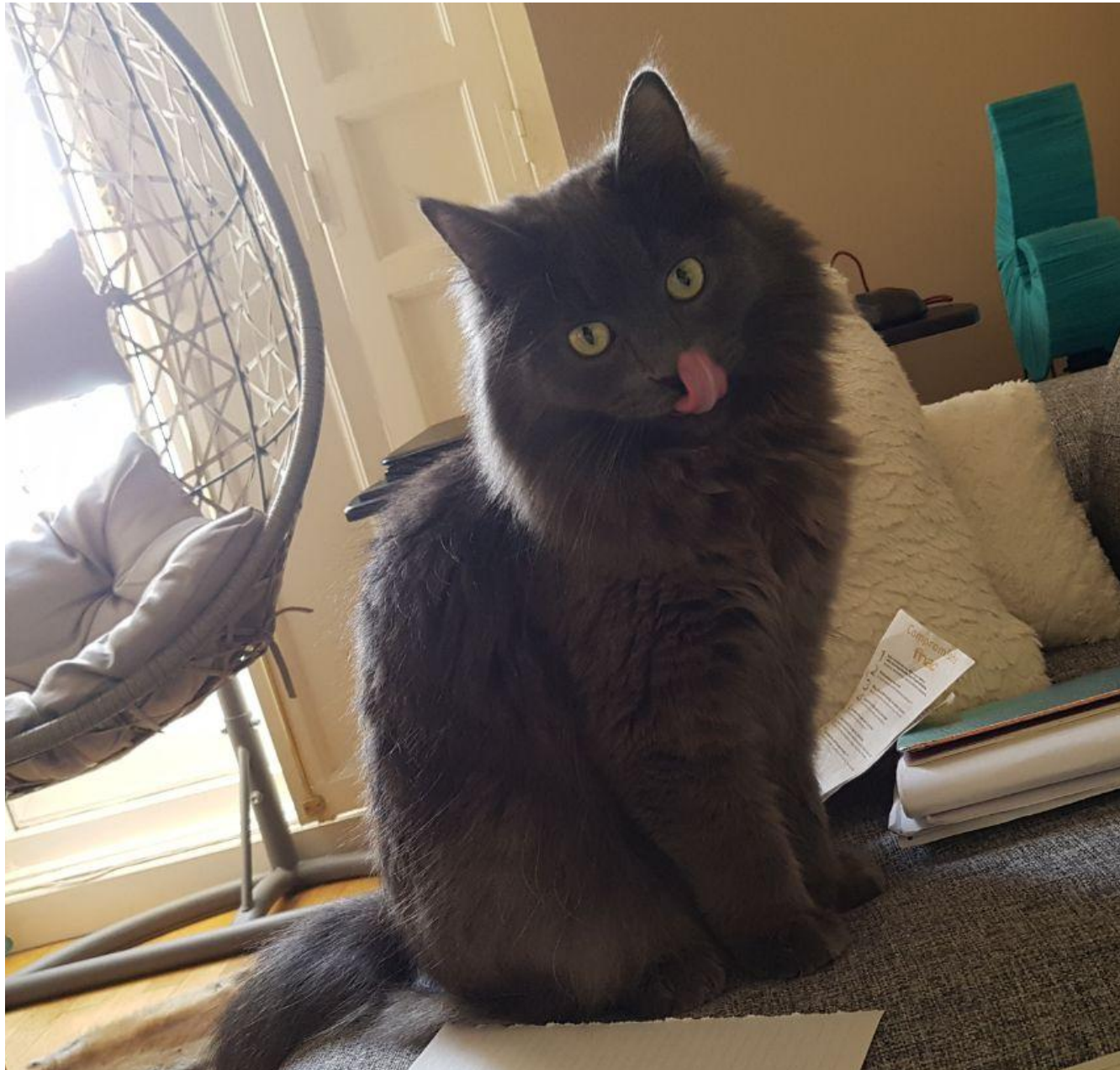
URLs (Web Addresses) are also resolvable!

That is, if I have a URL, there is a standard approach to finding the “thing” that URL is identifying.

The protocol is embedded in the URL itself
(HTTP → the Web!)

This is distinct from other kinds of GUIDs like DOIs, where the identifier is not associated with a protocol (and therefore might be resolved in a wide variety of ways)

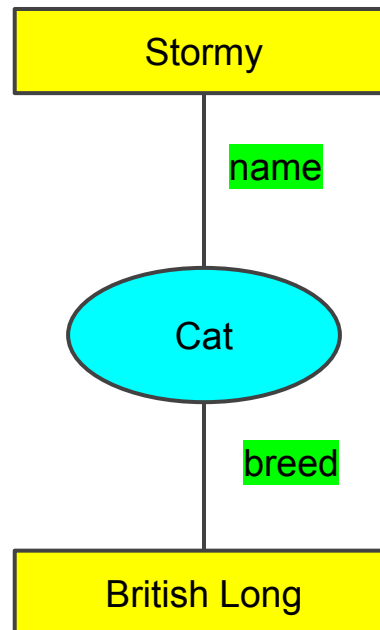




Before the detour we were talking about my cat

(we were also talking about RDF... but mostly about my cat ;-)

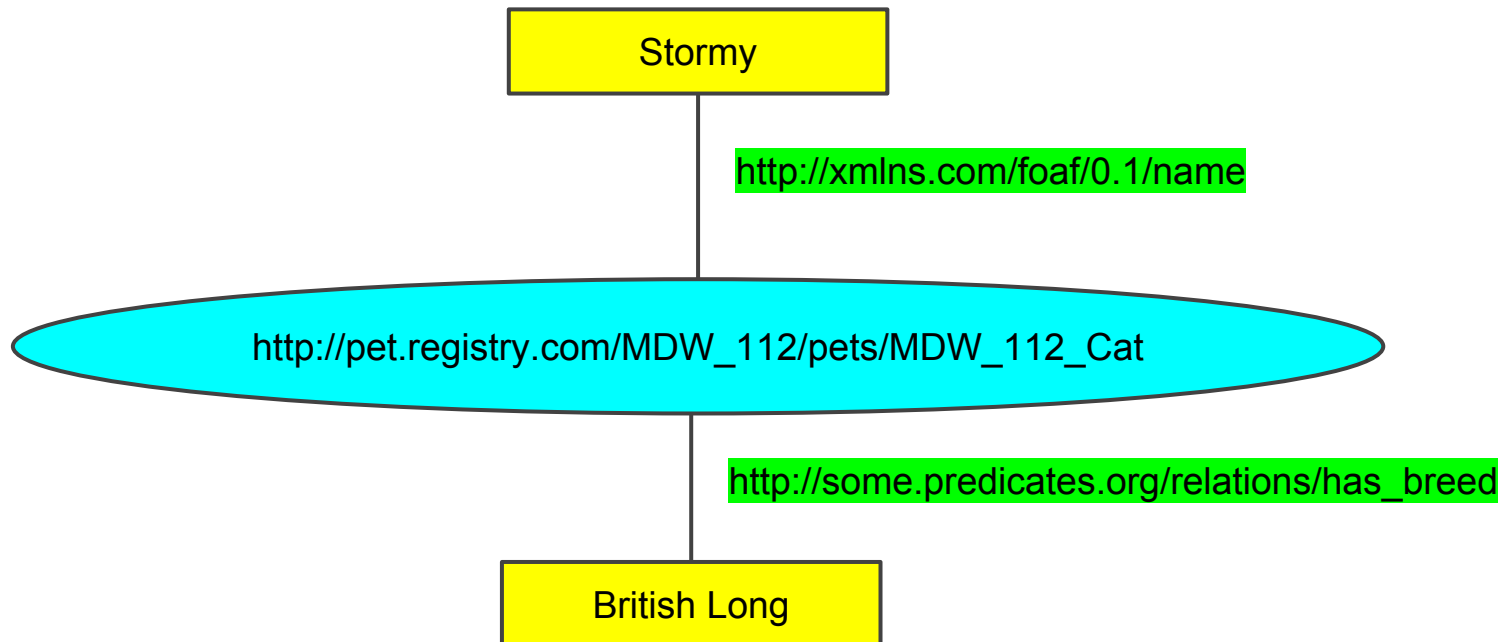
What is the connection between RDF and GUIDs?



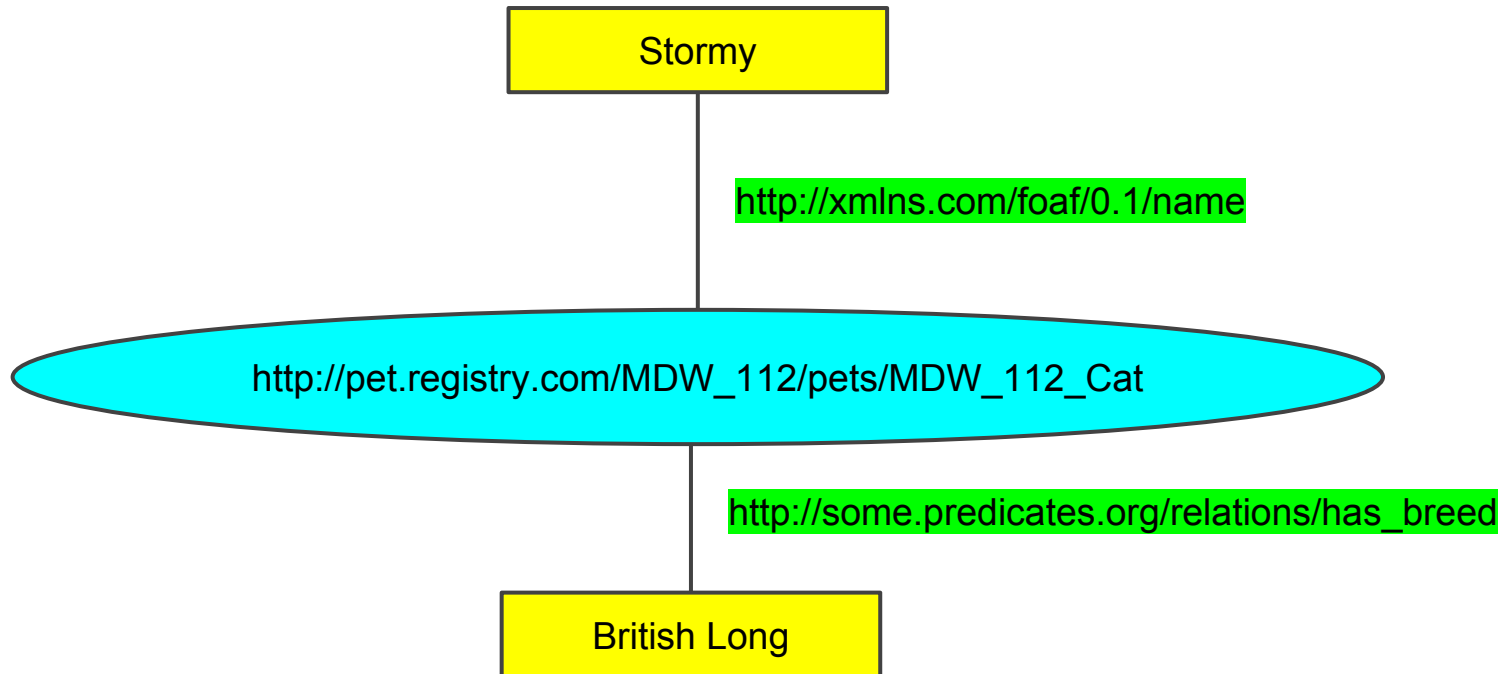
RDF is used on the Web as follows:

Resources are assigned URLs

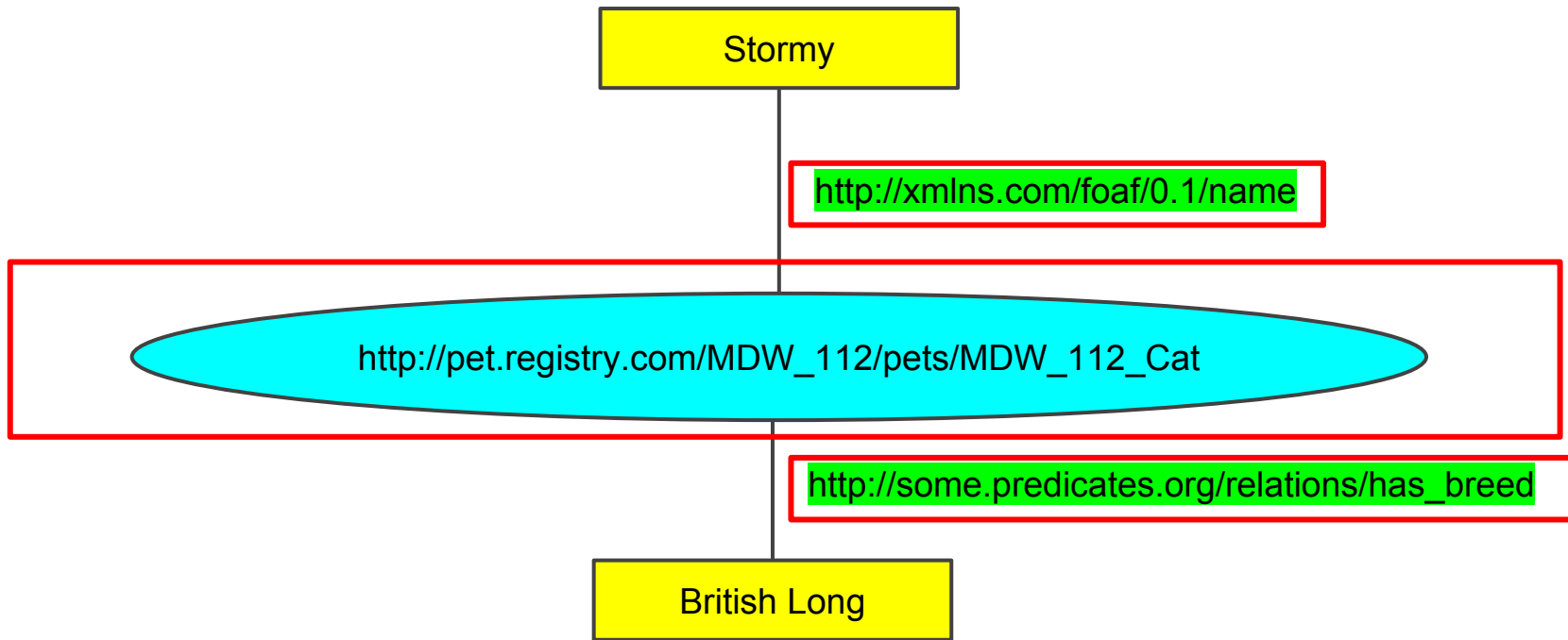
Predicates are assigned URLs



Benefits of RDF that uses URLs



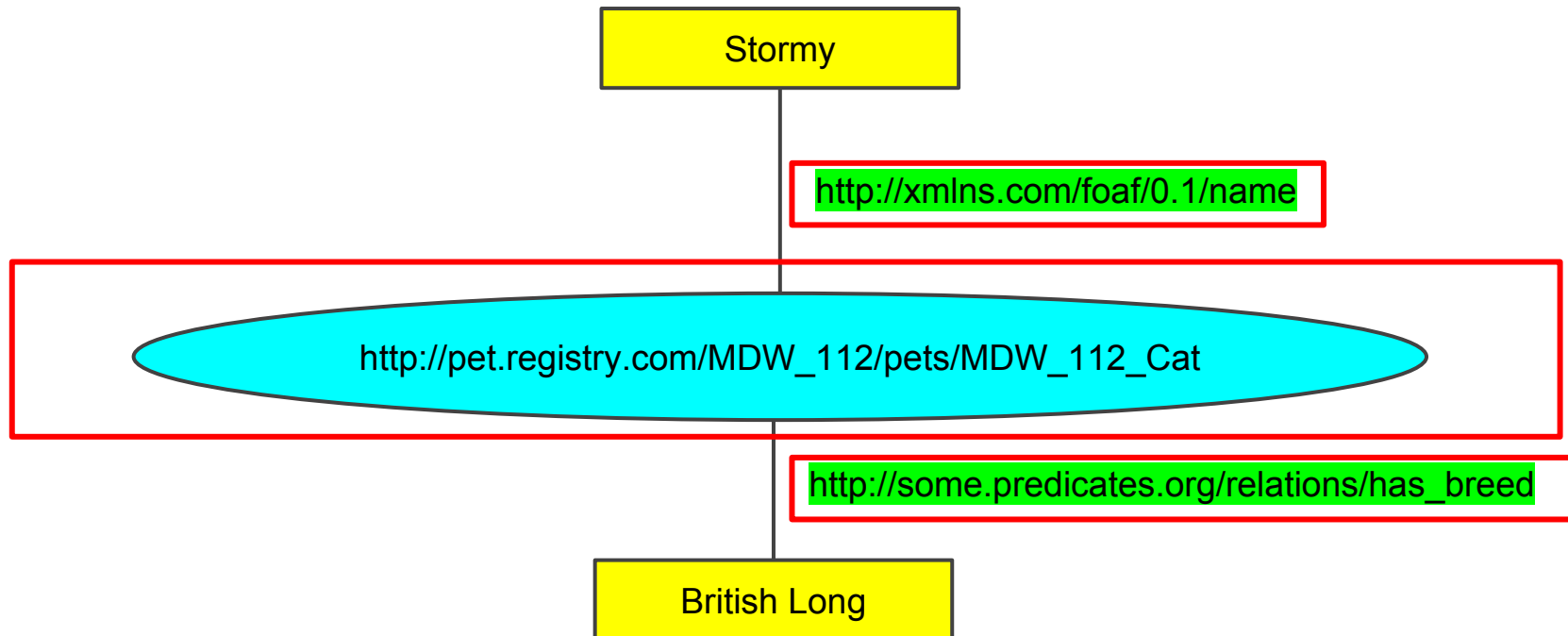
Benefits of RDF that uses URLs



All of these are **globally unique, SHARED*** and **unambiguous**

*meaning anyone, anywhere, can use them... BETTER TO NOT THINK OF THEM AS "LOCATIONS" → NAMES!

Benefits of RDF that uses URLs

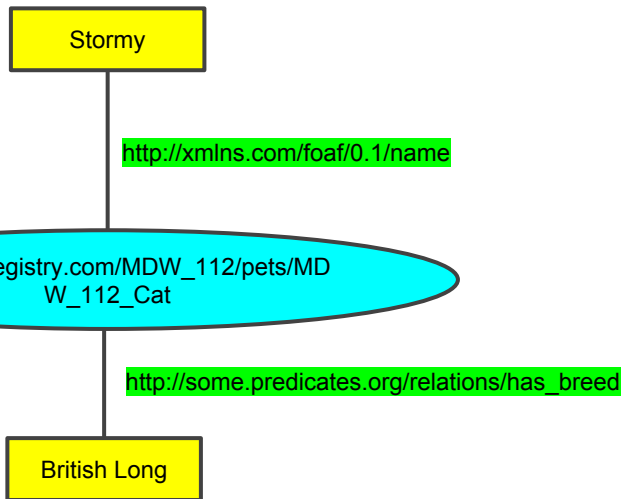


All of these can be **automatically looked-up by a machine!**

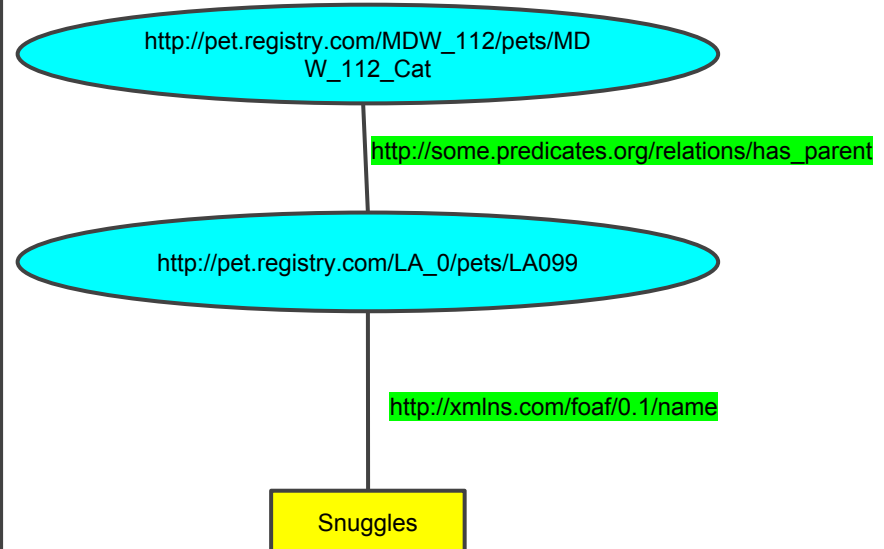
Get the page at "http://pet.registry.com/MDW_112/pets/MDW_112_Cat"

Benefits of RDF that uses URLs

My Graph

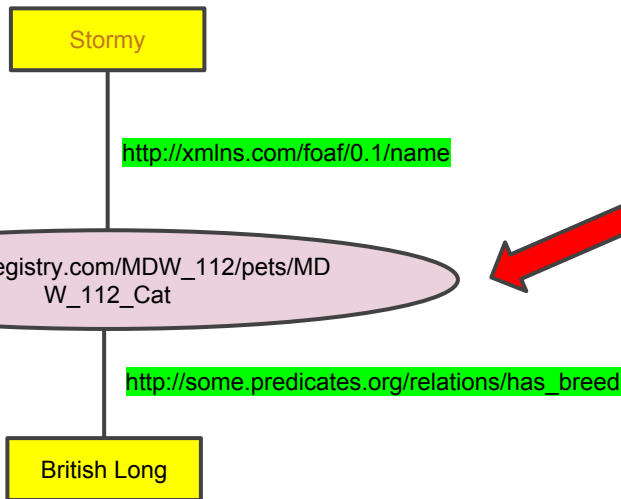


Pet Shop's Graph

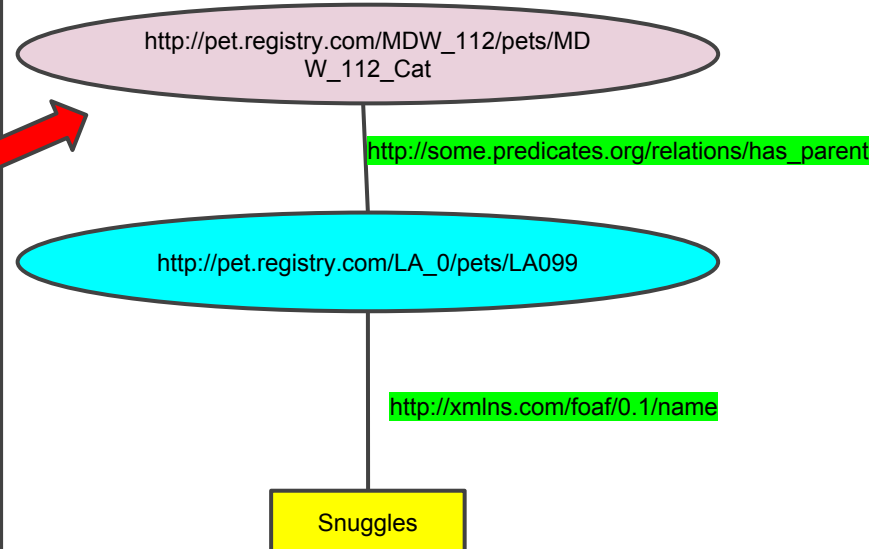


Benefits of RDF

My Graph



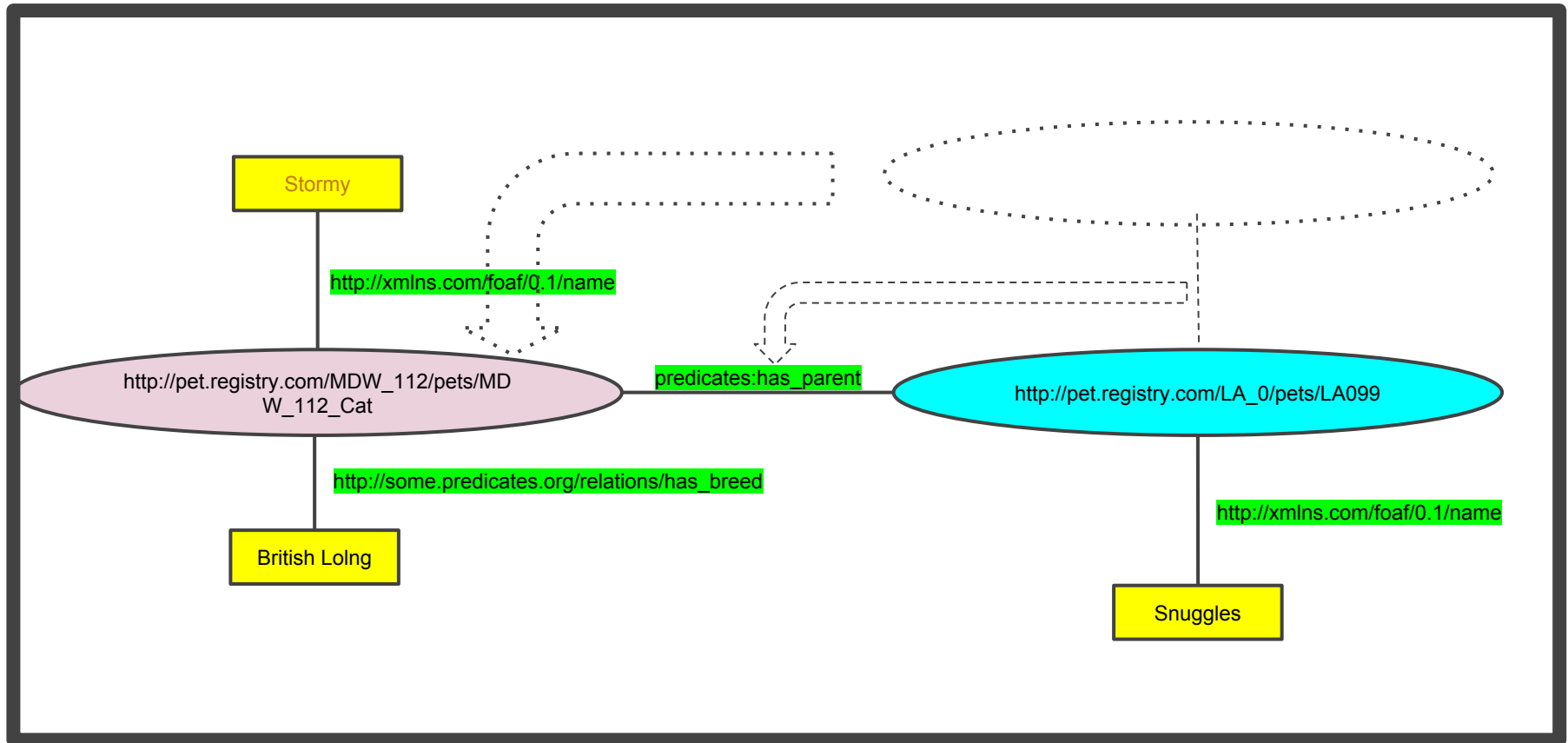
Pet Shop's Graph



URLs are globally Unique! → Two sites with statements ***about the same thing...***

Benefits of RDF

Unified Graph



A machine can do this data integration automatically

No special technology was required
to do that integration!

Because we use GUIDs + RDF, the integration
“just happened”

You simply download and concatenate the files!

(of course, there are more powerful ways to do this
than downloading the data)

The Relationship between RDF and the Web

Gene - Detail	
search list download trait classification	
Detail of Gene	
Basic Information	
CGSNL Gene Symbol	AGR1
Gene Symbol Synonym	agr1*, agr1
CGSNL Gene Name	AGRAVITROPISM 1
Gene Name Synonym	agravitropism 1, agravitropism 1, agravitropism-1
Protein Name	
Allele	
Chromosome No.	
Explanation	PO:0009009; embryo ; PO:0009005; root ; PO:0009025; leaf. GRO:0007047; 02-seedling ; GRO:0007043; 01-germination..
Trait Class	Seed - Morphological traits - Embryo
Expression	
Sequence/Locus	
Accession No.	-
RAP ID	(IRGSP 1.0 / Build5)
INSD Accession List (Test version)	-
Map	
Locate(cM)	
References	
Hong, S.K., T, Aoki, H. Kitano, H. Satoh and Y. Nagato (1995) Dev Genet. Phenotypic diversity of 188 rice embryo mutants.	
TextPresso Search	Search textpresso for AGR1 (Recent references may be retrievable, but without any warranty)
DB Reference	
Gramene ID	GR:0060019
Ontologies	
Gene Ontology	multicellular organismal development(GO:0007275) gravitropism(GO:0009630) anatomical structure morphogenesis(GO:0009653) embryonic development(GO:0009790)
Trait Ontology	root anatomy and morphology trait(TO:0000043) leaf anatomy and morphology trait(TO:0000419)

[Click here f](#)

Hyperlinks to Functional Annotations

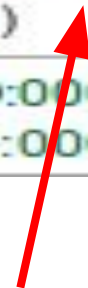


Hyperlinks carry no meaning

multicellular organismal development([GO:0007275](#))
gravitropism([GO:0009630](#))
anatomical structure morphogenesis([GO:0009653](#))
embryonic development([GO:0009790](#))

root anatomy and morphology trait([TO:0000043](#))
leaf anatomy and morphology trait([TO:0000419](#))

What will happen if I click this link?
Will I see a movie? Will I download a file?



Humans design nice Web pages

so that other Humans

can make good guesses

about what they want to click

Humans design nice Web pages

so that other Humans

can make good guesses

about what they want to click

That is the only reason that the Web is so successful!

Human Intuition!

The Relationship between RDF and the Web

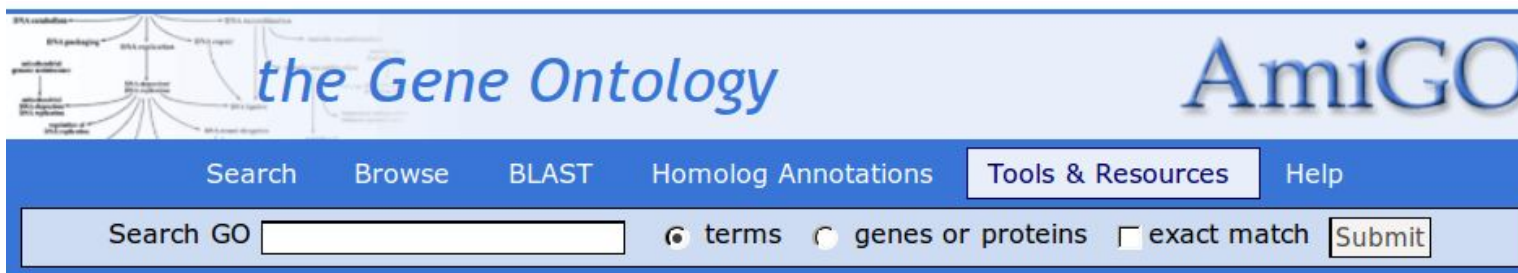
Gene - Detail	
search list download trait classification	
Detail of Gene	
Basic Information	
CGSNL Gene Symbol	AGR1
Gene Symbol Synonym	agr1*, agr1
CGSNL Gene Name	AGRAVITROPISM 1
Gene Name Synonym	agravitropism 1, agravitropism 1, agravitropism-1
Protein Name	
Allele	
Chromosome No.	
Explanation	PO:0009009; embryo ; PO:0009005; root ; PO:0009025; leaf. GRO:0007047; 02-seedling ; GRO:0007043; 01-germination..
Trait Class	Seed - Morphological traits - Embryo
Expression	
Sequence/Locus	
Accession No.	-
RAP ID	(IRGSP 1.0 / Build5)
INSD Accession List (Test version)	-
Map	
Locate(cM)	
References	
Hong, S.K., T, Aoki, H. Kitano, H. Satoh and Y. Nagato (1995) Dev Genet. Phenotypic diversity of 188 rice embryo mutants.	
TextPresso Search	Search textpresso for AGR1 (Recent references may be retrievable, but without any warranty)
DB Reference	
Gramene ID	GR:0060019
Ontologies	
Gene Ontology	multicellular organismal development(GO:0007275) gravitropism(GO:0009630) anatomical structure morphogenesis(GO:0009653) embryonic development(GO:0009790)
Trait Ontology	root anatomy and morphology trait(TO:0000043) leaf anatomy and morphology trait(TO:0000419)

Click here f

Hyperlinks to Functional Annotations



The Relationship between RDF and the Web



the Gene Ontology

AmiGO

Search Browse BLAST Homolog Annotations Tools & Resources Help

Search GO terms genes or proteins exact match

gravitropism

[Term information](#) ↓ [Term neighborhood](#) ↓ [External references](#) ↓ [157 gene product associations](#) →

Term Information

Accession	GO:0009630
Ontology	Biological Process
Synonyms	exact: geotropism
Definition	The orientation of plant parts under the stimulation of gravity. <i>Source:</i> ISBN:0198547684
Comment	None
Subset	None
Community	Add usage comments for this term on the GONUTS wiki.

The Relationship between RDF and the Web

<http://www.shigen.nig.ac.jp/rice/oryzabaseV4/gene/detail/24>

<http://amigo.geneontology.org/amigo/term=GO%3A0009630>

Oryzabase sponsored by NBRP
Integrated Rice Science Database

Home About Strains distribute/deposit Education Genome/Dev Genes Tools Download
Last updated: Mar 6, 2013

Gene - Detail

search list download trait classification

Detail of Gene

Click here for feedback

Basic Information	
CGSNL Gene Symbol	AGR1
Gene Symbol Synonym	agr1*, agr1
CGSNL Gene Name	AGRAVITROPISM 1
Gene Name Synonym	agravitropism 1, agravitropism 1, agravitropism-1
Protein Name	
Allele	
Chromosome No.	
Explanation	PO:0009009; embryo ; PO:0009005; root ; PO:0009025; leaf; GRO:0007047; 02-seedling ; GRO:0007043; 01-germination.
Trait Class	Seed - Morphological traits - Embryo
Expression	
Sequence/Locus	
Accession No.	-
RAP ID	(IRGSP 1.0 / Build5)
INSD Accession List (Test version)	-
Map	
Locate (cM)	
References	
Hong, S.K., T. Aoki, H. Kitano, H. Satoh and Y. Nagata (1995) Dev Genet. Phenotypic diversity of 188 rice embryo mutants.	
Textpresso Search	Search textpresso for AGR1 (Recent references may be retrievable, but without any warranty)
DB Reference	
Gramene ID	GR:0060019
Ontologies	
Gene Ontology	multicellular organismal development(GO:0007275) gravitropism(GO:0009630) anatomical structure morphogenesis(GO:0009653) embryonic development(GO:0009750)
Trait Ontology	root anatomy and morphology trait(TO:0000043) leaf anatomy and morphology trait(TO:0000419)

the Gene Ontology
AmiGO

Search Browse BLAST Homolog Annotations Tools & Resources Help

Search GO terms genes or proteins exact match Submit

gravitropism

Term information Term neighborhood External references 157 gene product associations

Term Information

Accession	GO:0009630
Ontology	Biological Process
Synonyms	exact: geotropism
Definition	The orientation of plant parts under the stimulation of gravity. <i>Source:</i> ISBN:0198547684
Comment	None
Subset	None
Community	Add usage comments for this term on the GONUTS wiki.

The Relationship between RDF and the Web

A human knows what this “link” means - it means that the rice gene is involved in gravitropism

But there are ~ 1.3 Trillion Gigabytes of data on the Web!
Too much for a human to ever explore...

We need a Web that can be explored by machines!!

<http://www.shigen.nig.ac.jp/rice/oryzabaseV4/gene/detail/24>

<http://amigo.geneontology.org/amigo/term=GO%3A0009630>

The screenshot shows the Oryzabase website with a search bar and navigation menu. The main content area is titled "Gene - Detail" and shows information for the gene AGR1. A red arrow points from the "Gene Ontology" section of this page to the Amigo page.

Basic Information	
CGSNL Gene Symbol	AGR1
Gene Symbol Synonym	agr1*, agr1
CGSNL Gene Name	AGRAVITROPISM 1
Gene Name Synonym	agravitropism 1, agravitropism 1, agravitropism-1
Protein Name	
Allele	
Chromosome No.	
Explanation	PO:0009009; embryo; PO:0009005; root; PO:0009025; leaf; GRO:0007047; 02-seedling; GRO:0007043; 01-germination.
Trait Class	Seed - Morphological traits - Embryo
Expression	
Sequence/Locus	
Accession No.	-
RAP ID	(IRGSP 1.0 / Build5)
INSD Accession List (Test version)	-
Map	
Locate (cM)	
References	
Hong, S.K., T. Aoki, H. Kitano, H. Satoh and Y. Nagata (1995) Dev Genet. Phenotypic diversity of 188 rice embryo mutants.	
Textpresso Search	Search textpresso for AGR1 (Recent references may be retrievable, but without any warranty)
DB Reference	
Gramene ID	GR:0060019
Ontologies	
Gene Ontology	multicellular organismal development(GO:0007275) gravitropism(GO:0009630) anatomical structure morphogenesis(GO:0009653) embryonic development(GO:0009750)
Trait Ontology	root anatomy and morphology trait(TO:0000043) leaf anatomy and morphology trait(TO:0000419)

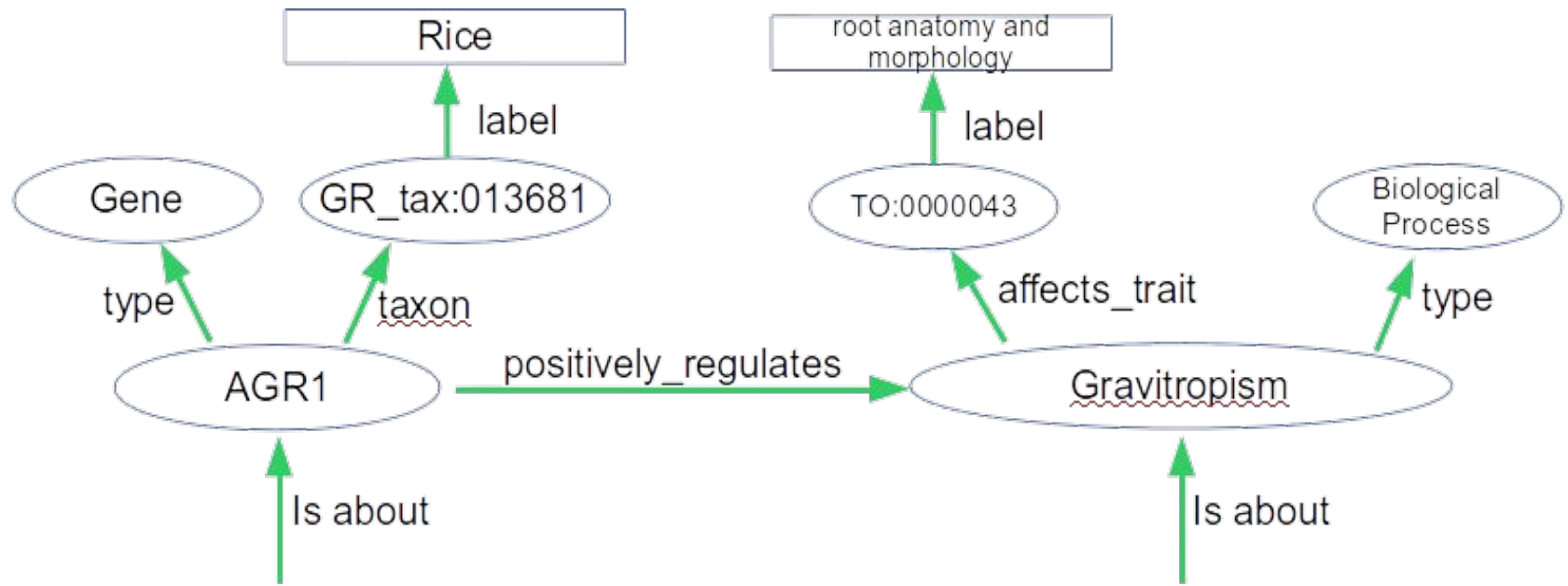
The screenshot shows the Amigo Gene Ontology search interface. It includes a search bar with the text "Search GO" and a "Submit" button. There are also options for "terms", "genes or proteins", and "exact match".

gravitropism

The screenshot shows the Amigo Term Information page for the term "gravitropism". It includes a navigation bar with "Term information", "Term neighborhood", "External references", and "157 gene product associations". The main content area is titled "Term Information" and lists various details about the term.

Accession	GO:0009630
Ontology	Biological Process
Synonyms	exact: geotropism
Definition	The orientation of plant parts under the stimulation of gravity. Source: ISBN:0198547684
Comment	None
Subset	None
Community	Add usage comments for this term on the GONUTS wiki.

The Relationship between RDF and the Web



<http://www.shigen.nig.ac.jp/rice/oryzabaseV4/gene/detail/24>

<http://amigo.geneontology.org/amigo/term=GO%3A0009630>

Oryzabase powered by ANBP
 International Rice Genome Database

Home About News Distribution Educator Genealogy Data Tools Download

Gene - Detail

Detail of Gene

Basic Information	
GDDB Gene Symbol	AGR1
Gene Symbol Synonym	agr1, agr1
GDDB Gene Name	AGR1/ATR1/GRM1
Gene Name Synonym	agr1tropism1, agr1tropism1, agr1tropism-1
Protein Name	
Alias	
Chromosome No.	
Explanation	PO-0000006 embryo; PO-0000005 root; PO-0000006 leaf; QRC0000047; QZ-0000000; QRC0000043; OT-gammarion; Seed - Morphological trait - Embryo
Trist Class	
Extension	
Sequence Name	
Accession No.	
PLAF ID	C (IRGSP 4.0 / GRM1)
RefSeq Accession List (Gene names)	
Map	
Location	
References	Wang, S. R., Y. Aoki, H. Hirata, H. Saito and Y. Nagata (1995) Rice (Oryza sativa) gravitropism: cloning of the rice agr1 gene. Plant Cell Physiol 36: 1001-1008.
Search	Search (help) for AGR1 (Please) references may be retrieved, but without any warranty.
DB Reference	
Gene ID	GR-0000019
Orthologs	<ul style="list-style-type: none"> Arabidopsis thaliana orthologous development GO:0007275 gravitropism GO:0009630 embryonic development morphogenesis GO:0009630 embryonic development GO:0009630 root anatomy and morphology trait TO:0000043 leaf anatomy and morphology trait TO:0000043
Gene Ontology	
Gene Ontology	

the Gene Ontology
 AmiGO

Search GO

Search GO [] terms genes or proteins exact match Submit

gravitropism

Term information | Term neighborhood | External references | 157 gene product associations

Term Information

Accession GO:0009630

Ontology Biological Process

Synonyms exact: geotropism

Definition The orientation of plant parts under the stimulation of gravity. Source: ISBN:0198547664

Comment None

Subset None

Community Add usage comments for this term on the GORUTS wiki.

Term Neighborhood for gravitropism (GO:0009630)

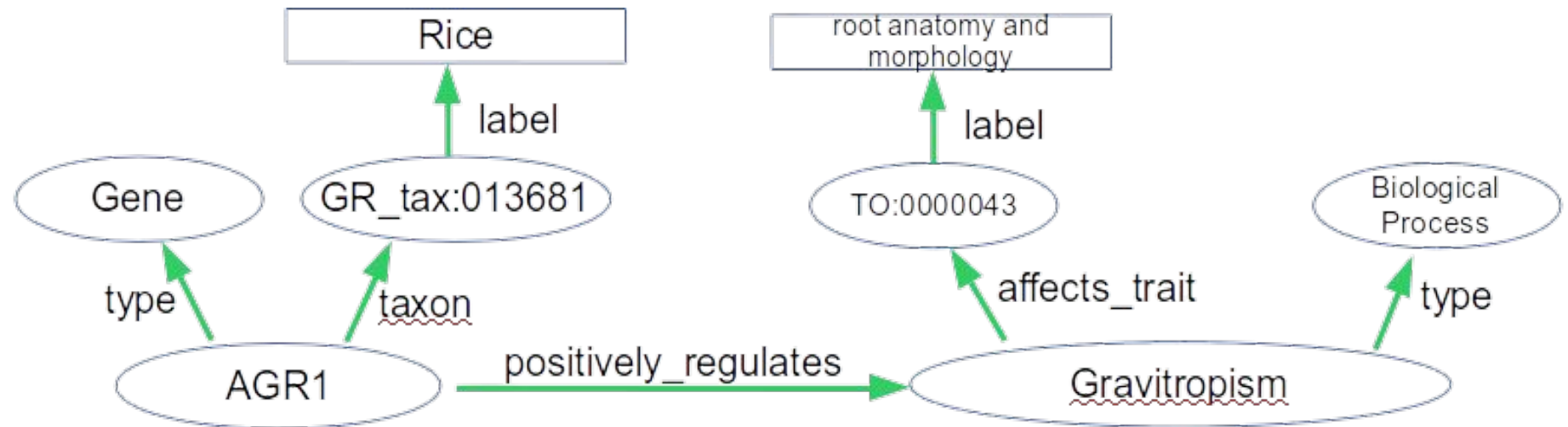
Filter lineage gene product counts

Data source Species

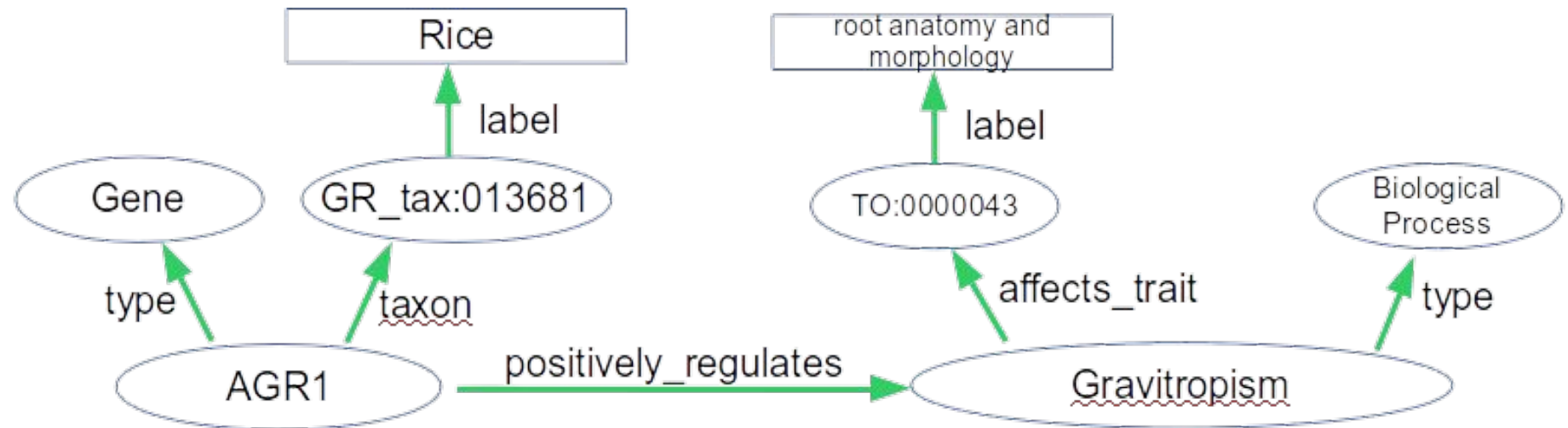
GRM1 2000000000

GRM1 2000000000

...and in my perfect world...
we'd just eliminate the human-readable Web entirely! ;-)



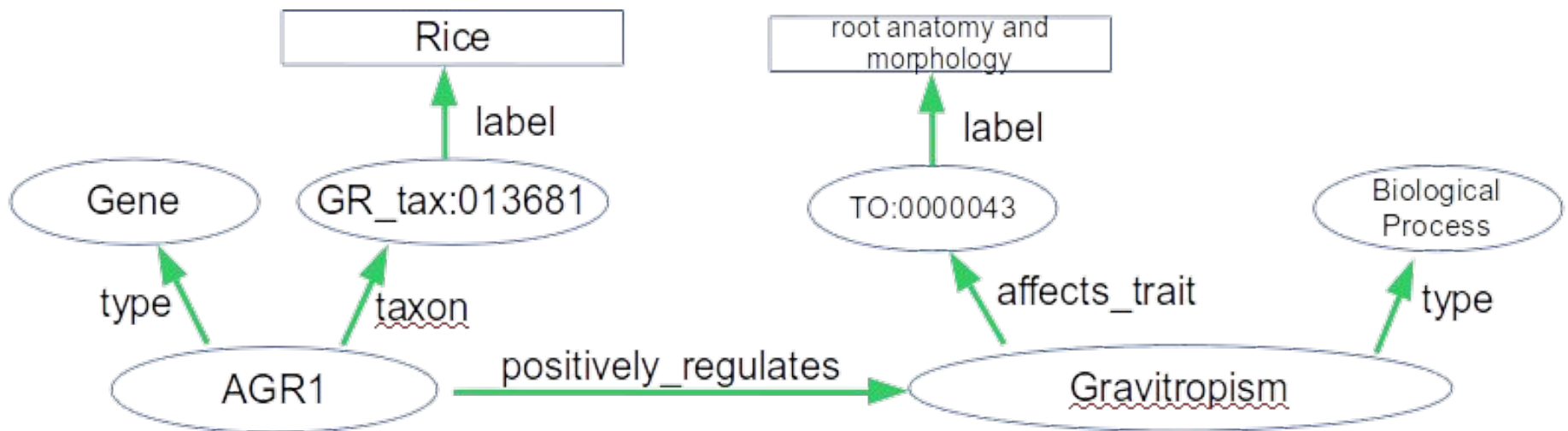
...and in my perfect world...
we'd just eliminate the human-readable Web entirely! ;-)



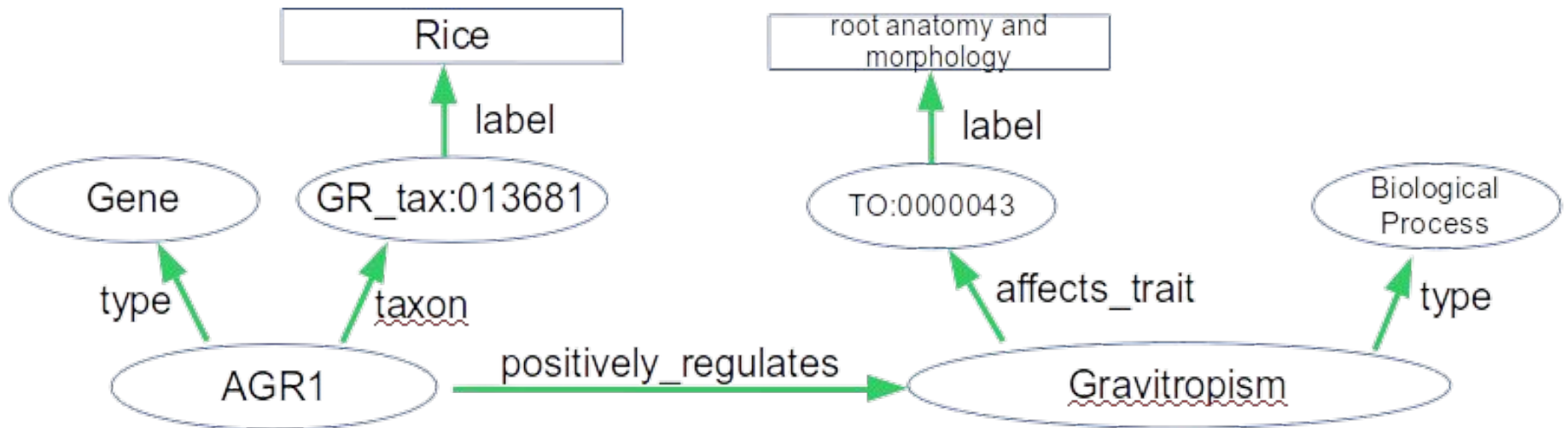
(There are ways of embedding RDF into HTML, so the two technologies are not mutually exclusive! ...but I won't be teaching you how to do that.

If you want to learn how to do it, search for "RDFa")

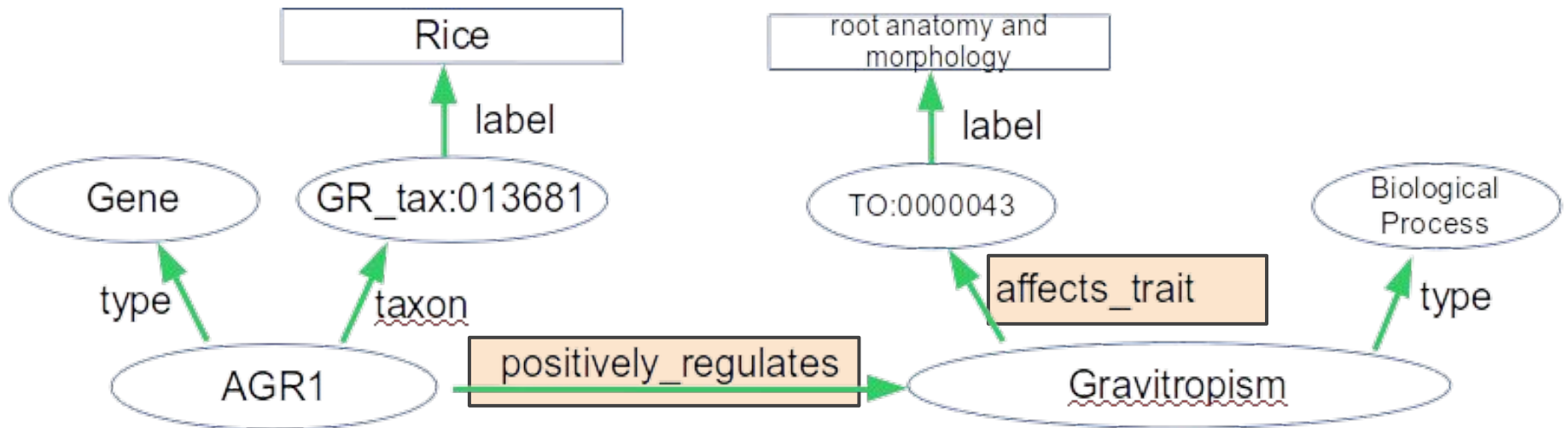
This is called “The Semantic Web”



“Semantics” implies “meaning”
Where is the “meaning” in this picture?

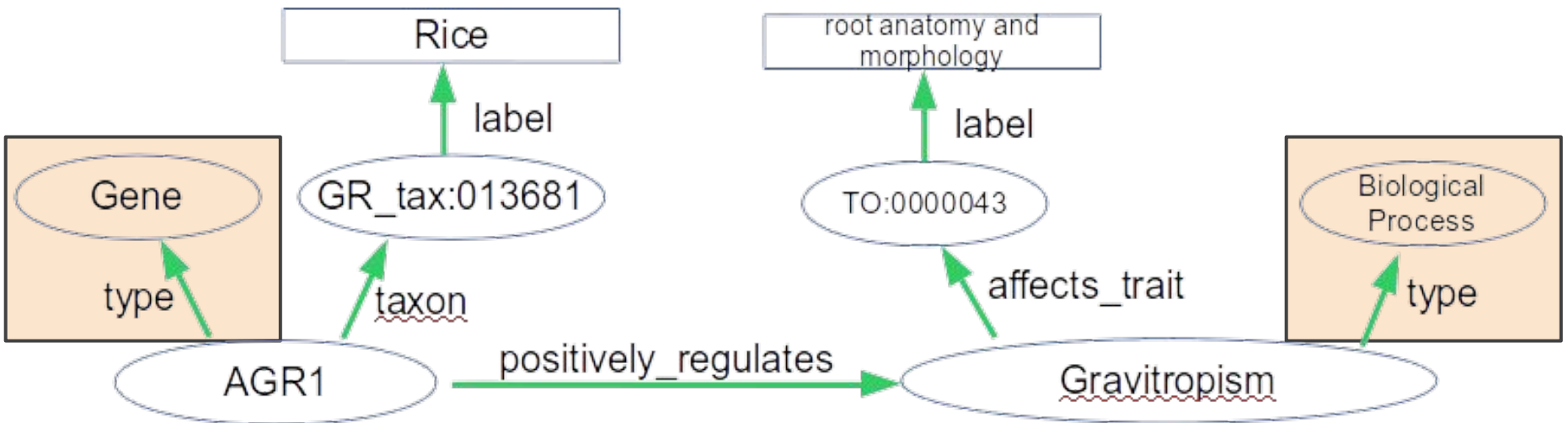


“Semantics” implies “meaning”
Where is the “meaning” in this picture?
Two parts of a Graph are able to carry meaning



Predicates...

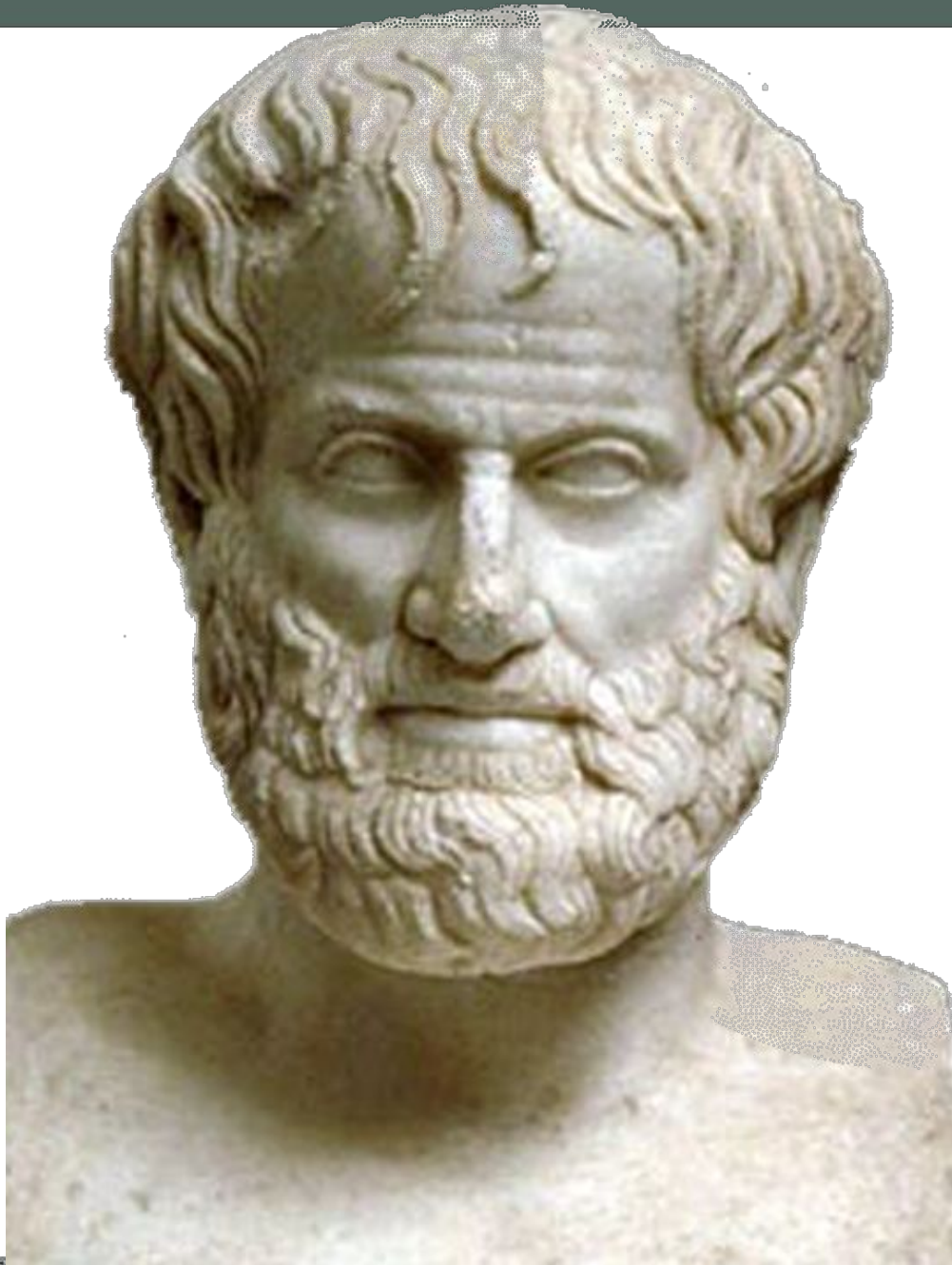
“Semantics” implies “meaning”
Where is the “meaning” in this picture?
Two parts of a Graph are able to carry meaning



and “rdf:type” nodes

(<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>)

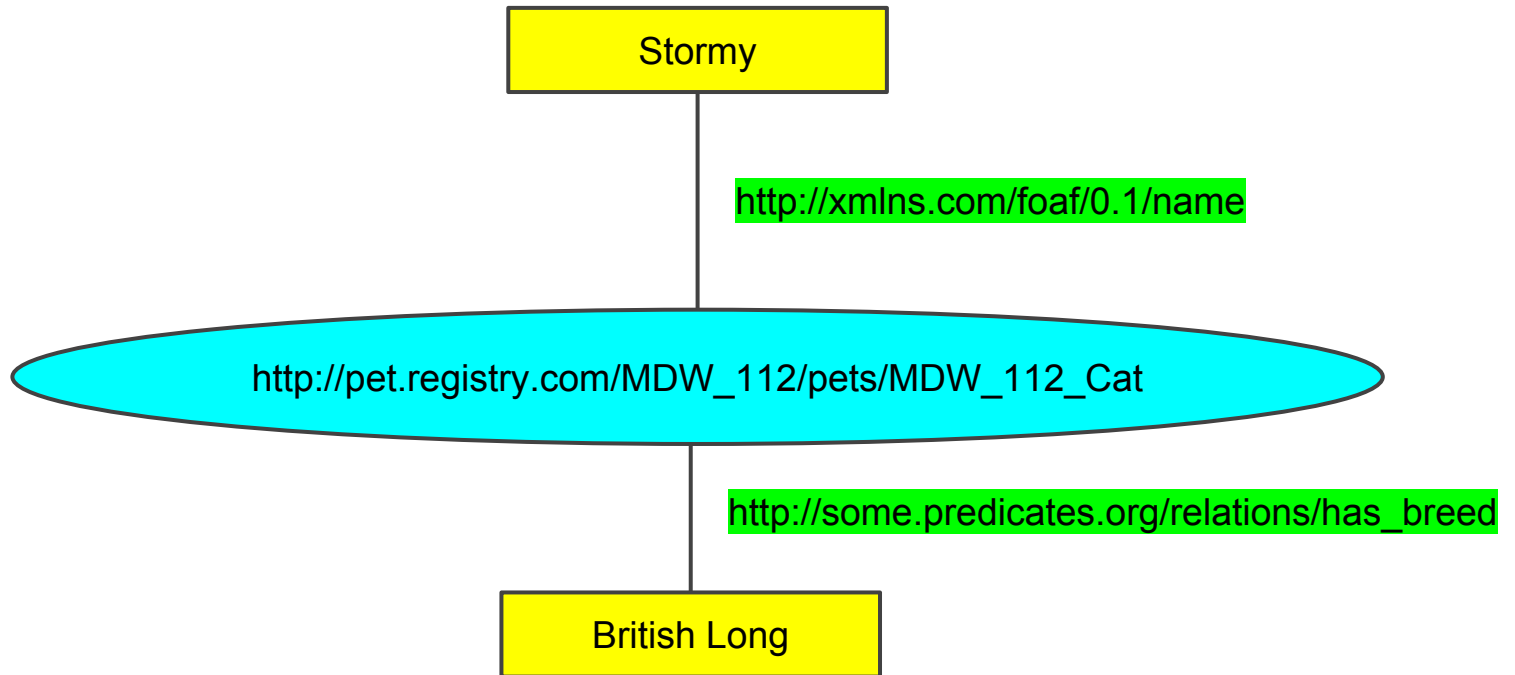
Where does “meaning” come from?



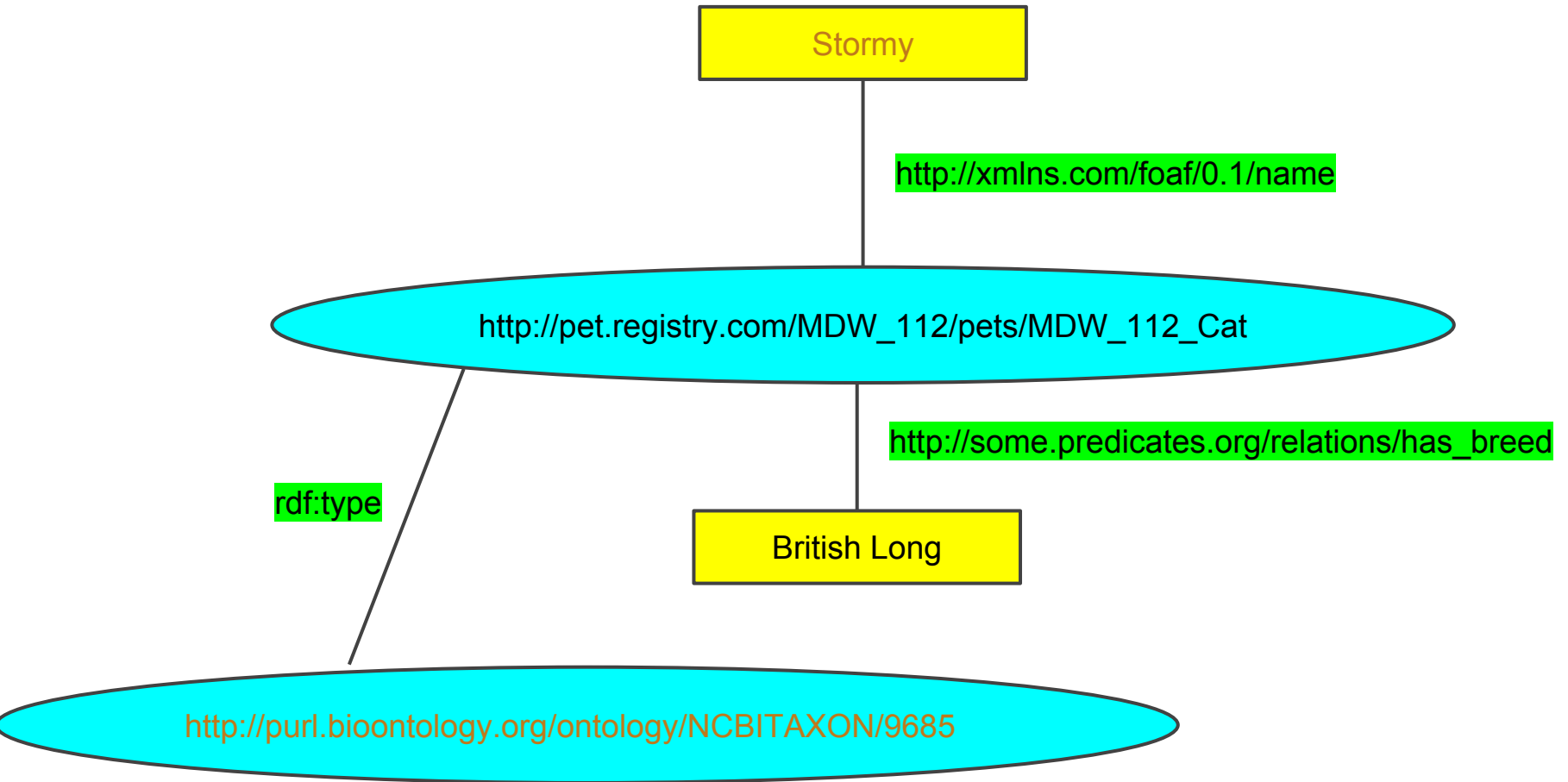
Ontologies

the branch of metaphysics dealing with the nature of being.

Let's talk about my cat again...



Let's talk about my cat again...




<http://purl.bioontology.org/ontology/NCBITAXON/9685>

There is global agreement on what "taxon:9685" means

Details	Visualization	Notes (0)	Class Mappings (17) 
Preferred Name	Felis catus		
Synonyms	Felis domesticus cats cat Felis silvestris catus		
ID	http://purl.bioontology.org/ontology/NCBITAXON/9685		
altLabel	Felis domesticus cats cat Felis silvestris catus		
cui	C0007450		
DIV	Mammals		
notation	9685		
prefLabel	Felis catus		
RANK	species		
tui	T015		
subClassOf	Felis		

<http://purl.bioontology.org/ontology/NCBITAXON/9685>

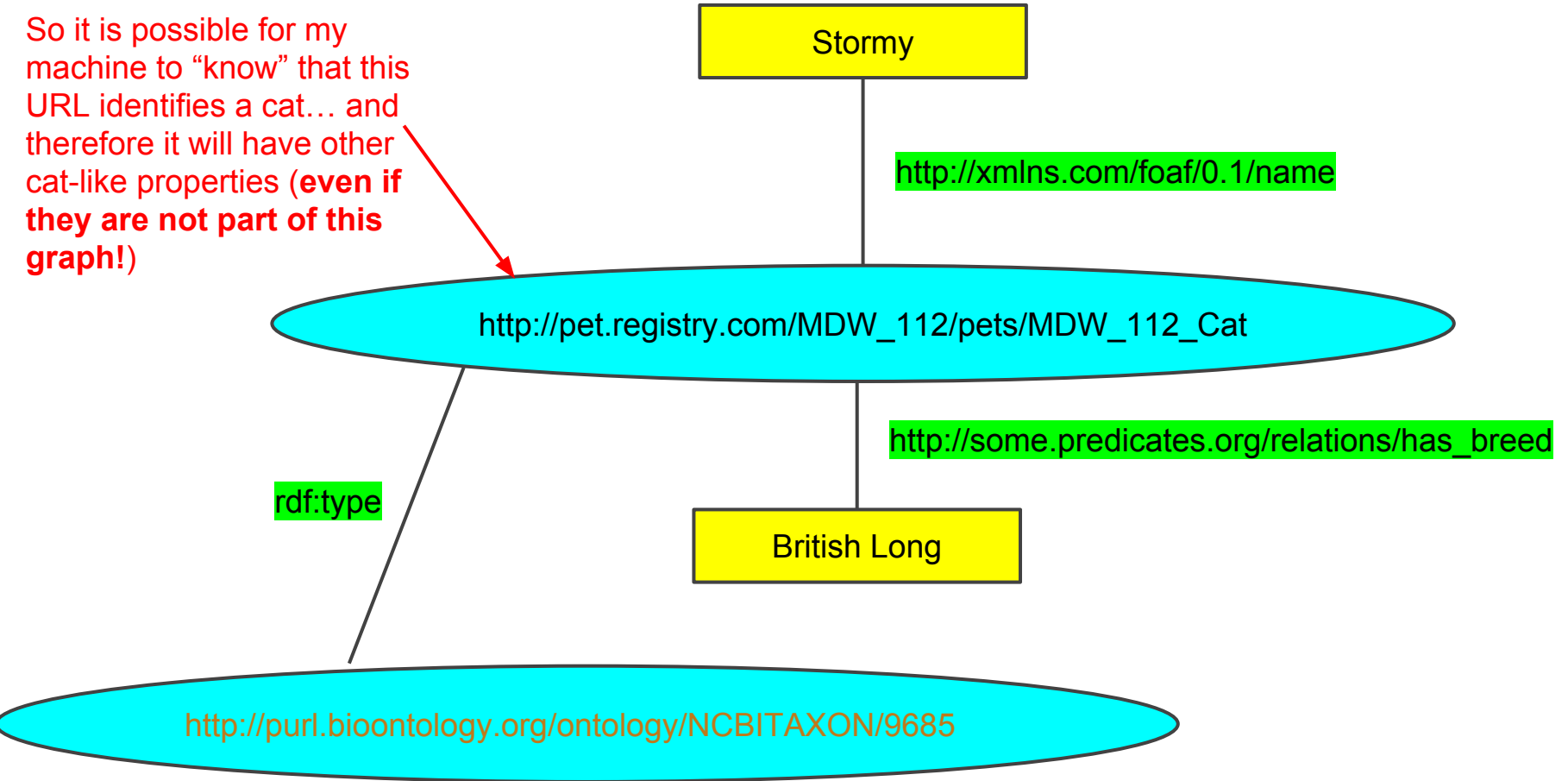
There is global agreement on what "taxon:9685" means
(and... you can also get the same information in 100%
machine-readable RDF! Not just the Web page)

Details	Visualization	Notes (0)	Class Mappings (17) 
Preferred Name	Felis catus		
Synonyms	Felis domesticus cats cat Felis silvestris catus		
ID	http://purl.bioontology.org/ontology/NCBITAXON/9685		
altLabel	Felis domesticus cats cat Felis silvestris catus		
cui	C0007450		
DIV	Mammals		
notation	9685		
prefLabel	Felis catus		
RANK	species		
tui	T015		
subClassOf	Felis		

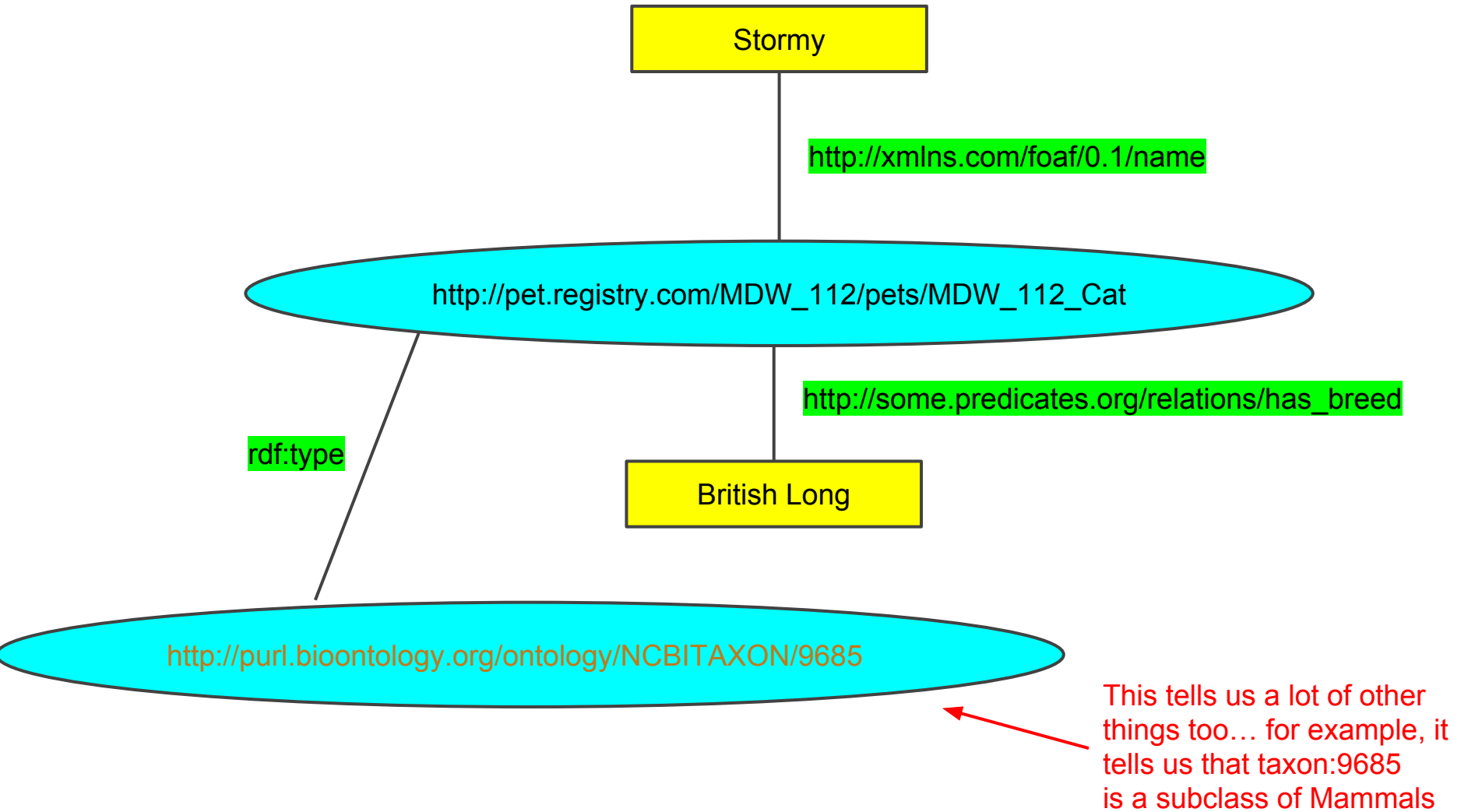
NCBITAXON

Let's talk about my cat again...

So it is possible for my machine to “know” that this URL identifies a cat... and therefore it will have other cat-like properties (**even if they are not part of this graph!**)

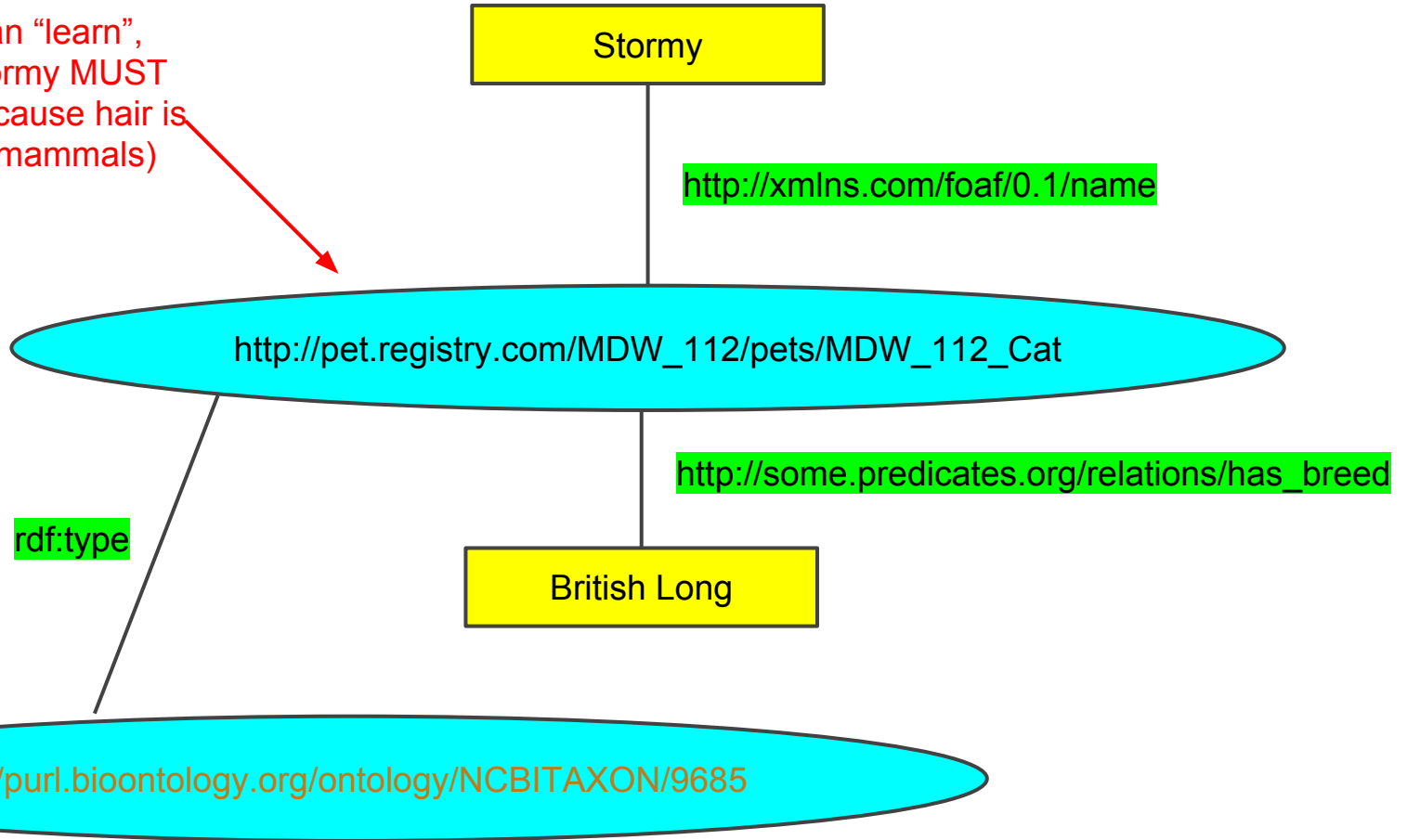


Let's talk about my cat again...

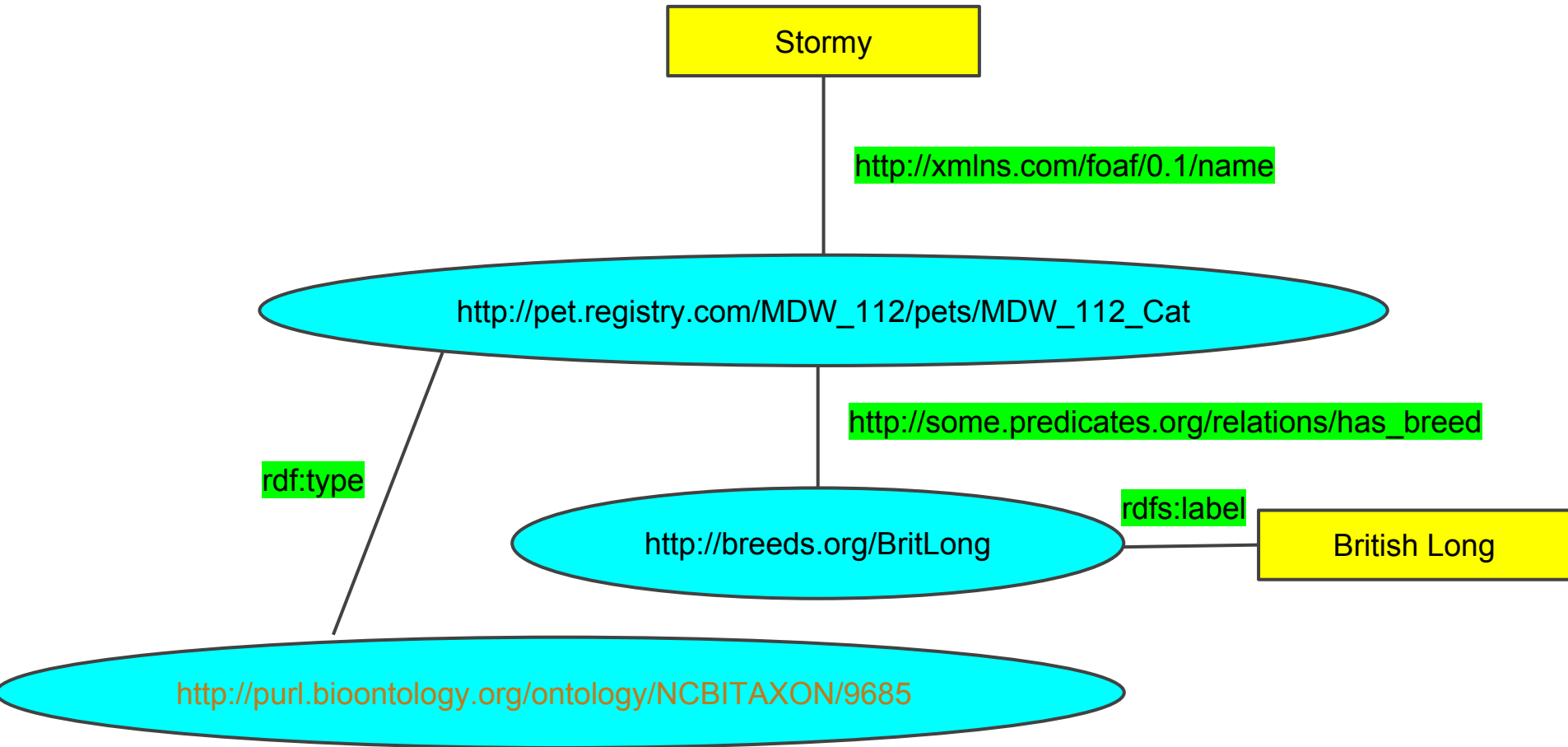


Let's talk about my cat again...

So a machine can “learn”, by itself, that Stormy MUST HAVE HAIR (because hair is a property of all mammals)

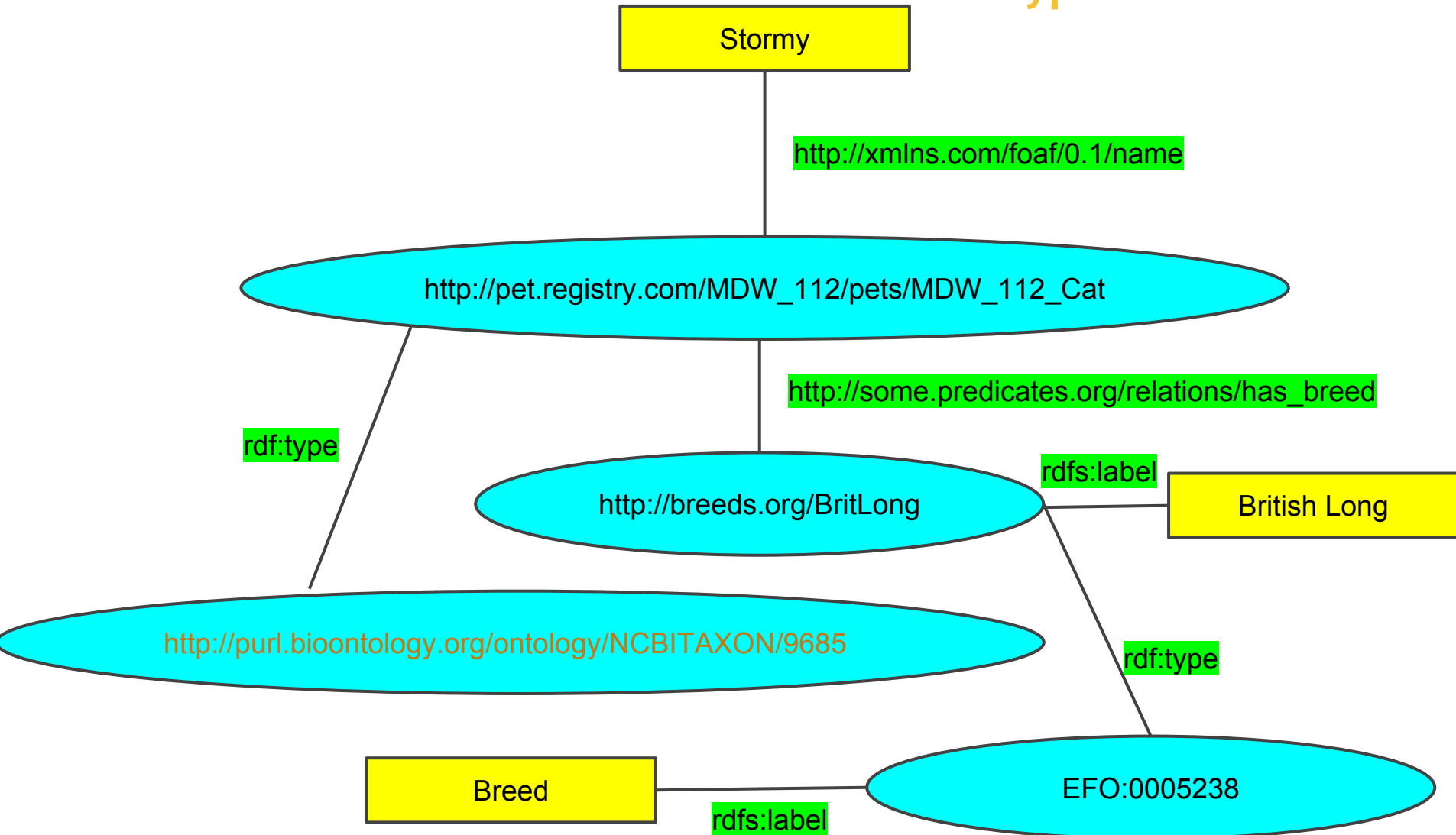


Ontologizing like crazy!



Now a machine can learn that Stormy must be a Grey cat
(british long's are always grey)

Note ...*all* Resources are now rdf:type'd!





Experimental Factor Ontology

[Home](#)

[Browse EFO](#)

[Submit Term](#)

[EBI RDF Platform](#)

breed

http://www.ebi.ac.uk/efo/EFO_0005238

A breed is a specific group of domestic animals or plants having homogeneous appearance, homogeneous behavior, and other characteristics that distinguish it from other animals or plants of the same species and that were arrived at through selective breeding.

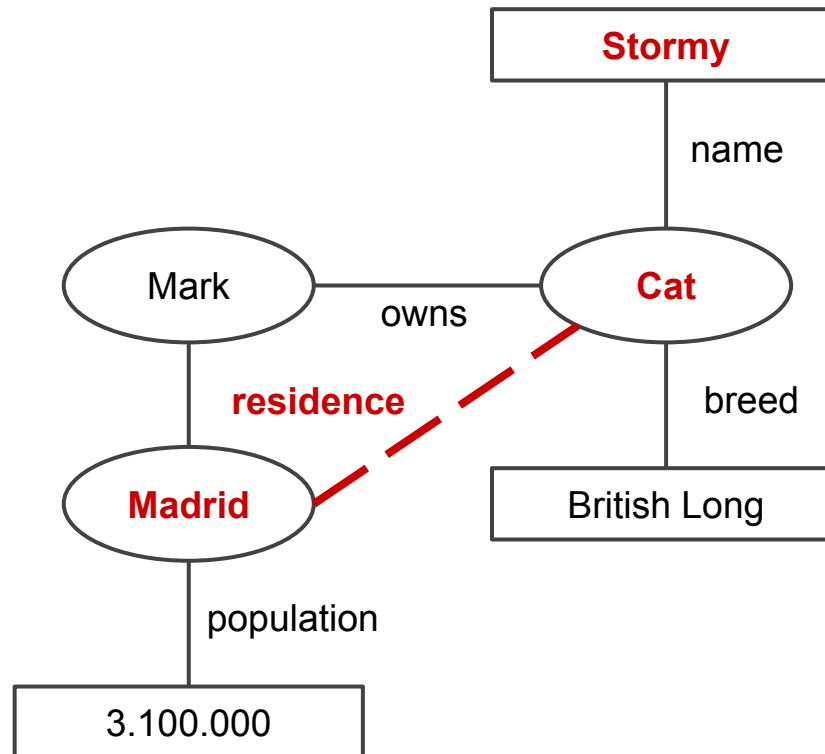
[parents](#)

[population](#)

We hypothesise that a LOT of
new discoveries

will come by automatic “inference”

on this Global Graph of all human knowledge



Link The World!

Ontologize The World!

:-)

Imagine if 1.3 Trillion Gigabytes of data were represented in this format, and linked into ontologies... what the machines could discover for us!

What does RDF “look like”

Remember, RDF is a “framework” - a way of thinking about things

To be USEFUL we have to find a way to represent it and pass it from machine-to-machine

There are several choices of “serialization”:

- XML (RDF/XML)
- n3 (N-triples)
- Turtle
- RDFa
-



Consider the BRCA1 gene

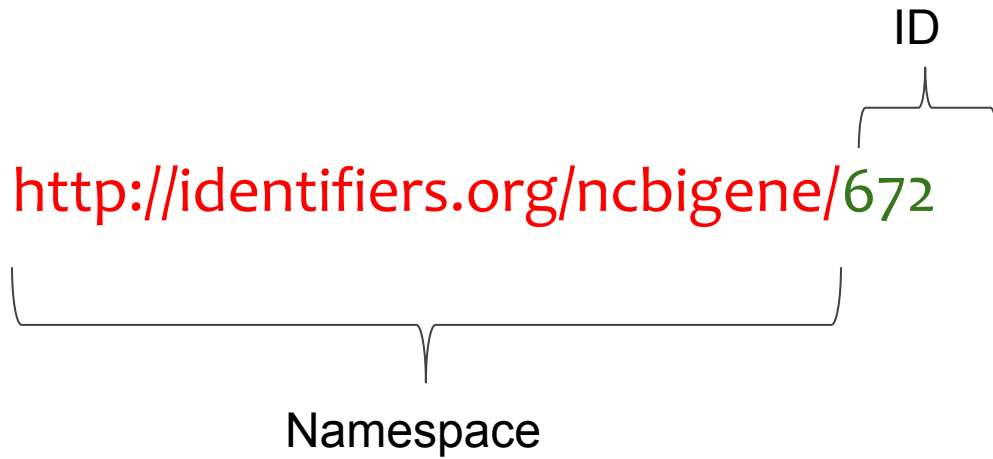
Its URL is: <http://identifiers.org/ncbigene/672>

It would be tedious (and error-prone) to type this over and over again.

So we use “Namespaces” to make it easier

<http://identifiers.org/ncbigene/672>

Namespaces



<http://identifiers.org/ncbigene/>

672

PREFIX ns: <<http://identifiers.org/ncbigene/>>

672

Namespaces

The prefix can be any series of characters;
you can choose whatever you wish

PREFIX ncbi: <<http://identifiers.org/ncbigene/>>

672

PREFIX ncbi: <<http://identifiers.org/ncbigene/>>

ncbi:672

PREFIX ncbi: <<http://identifiers.org/ncbigene/>>

ncbi:672

And for the rest of the document
you refer to BRCA1 as ncbi:672

easy!



Turtle - very easy for humans

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix prov: <http://www.w3.org/ns/prov#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .

<https://linkeddata.systems/markdwilkinson.foaf>
  a                prov:Person , foaf:Person ;
  rdfs:label       "Mark D. Wilkinson" ;
  foaf:account

      <https://orcid.org/0000-0001-6960-357X#orcid-id> ;
  foaf:based_near [

      a gn:Feature ;
      gn:countryCode    "ES" ;
      gn:parentCountry <http://sws.geonames.org/2510769/ > ;
      foaf:familyName   "Wilkinson" ;
      foaf:givenName    "Mark" ;
      foaf:name         "Mark D. Wilkinson" ;
```

N-Triples: ~~hard for humans!

```
<http://example.org/bob#me> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://xmlns.com/foaf/0.1/Person> .
<http://example.org/bob#me> <http://xmlns.com/foaf/0.1/knows> <http://example.org/alice#me> .
<http://example.org/bob#me> <http://schema.org/birthDate>
"1990-07-04"^^<http://www.w3.org/2001/XMLSchema#date> .
<http://example.org/bob#me> <http://xmlns.com/foaf/0.1/topic_interest>
<http://www.wikidata.org/entity/Q12418> .
<http://www.wikidata.org/entity/Q12418> <http://purl.org/dc/terms/title> "Mona Lisa" .
<http://www.wikidata.org/entity/Q12418> <http://purl.org/dc/terms/creator>
<http://dbpedia.org/resource/Leonardo_da_Vinci> .
<http://data.europeana.eu/item/04802/243FA8618938F4117025F17A8B813C5F9AA4D619>
<http://purl.org/dc/terms/subject> <http://www.wikidata.org/entity/Q12418> .
```

```
<?xml version="1.0" encoding="utf-8"?>
  <rdf:RDF
    xmlns:dcterms="http://purl.org/dc/terms/"
    xmlns:foaf="http://xmlns.com/foaf/0.1/"
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:schema="http://schema.org/">
    <rdf:Description rdf:about="http://example.org/bob#me">
      <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
      <schema:birthDate
rdf:datatype="http://www.w3.org/2001/XMLSchema#date">1990-07-04</schema:birthDate>
      <foaf:knows rdf:resource="http://example.org/alice#me"/>
      <foaf:topic_interest rdf:resource="http://www.wikidata.org/entity/Q12418"/>
    </rdf:Description>
    <rdf:Description rdf:about="http://www.wikidata.org/entity/Q12418">
      <dcterms:title>Mona Lisa</dcterms:title>
      <dcterms:creator rdf:resource="http://dbpedia.org/resource/Leonardo_da_Vinci"/>
    </rdf:Description>
    <rdf:Description
rdf:about="http://data.europeana.eu/item/04802/243FA8618938F4117025F17A8B813C5F9AA4D619">
      <dcterms:subject rdf:resource="http://www.wikidata.org/entity/Q12418"/>
    </rdf:Description>
  </rdf:RDF>
```

Namespaces

```
<?xml version="1.0" encoding="utf-8"?>
<rdf:RDF
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">

<rdf:Description rdf:about="http://example.org/bob#me">
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
  <foaf:knows rdf:resource="http://example.org/alice#me"/>
```

Interpreted as:

foaf:knows ⇒ <http://xmlns.com/foaf/0.1/knows>

rdf:Description ⇒

<http://www.w3.org/1999/02/22-rdf-syntax-ns#Description>

Same thing in Turtle

```
@prefix contact: <http://www.w3.org/2000/10/swap/pim/contact#>.

<http://linkeddata.systems/markdwilkinson.foaf>
  a contact:Person;
  contact:fullName "Eric Miller";
  contact:mailbox <mailto:em@w3.org>;
  contact:personalTitle "Dr.".
```

Makes it easier to read (for humans).

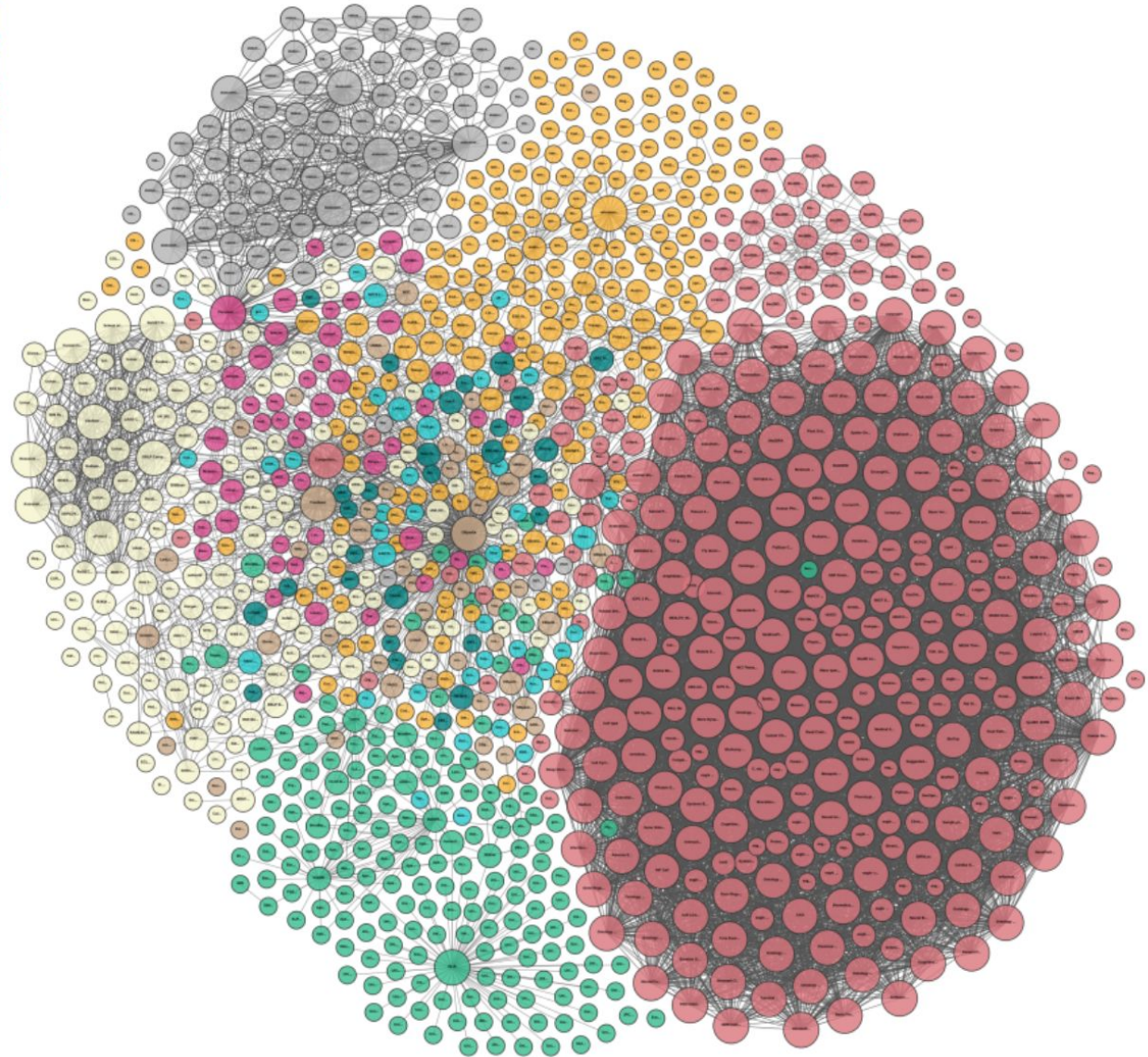
(but readability ISN'T the reason that namespaces exist!! ...that's another topic...)

How much data in the world is represented in
RDF?

(hint... a lot!)

LOD Cloud Today

Last updated: 2017-02-20



How do I get to the data?

SPARQL - the query language for RDF

SPARQL - briefly and simply

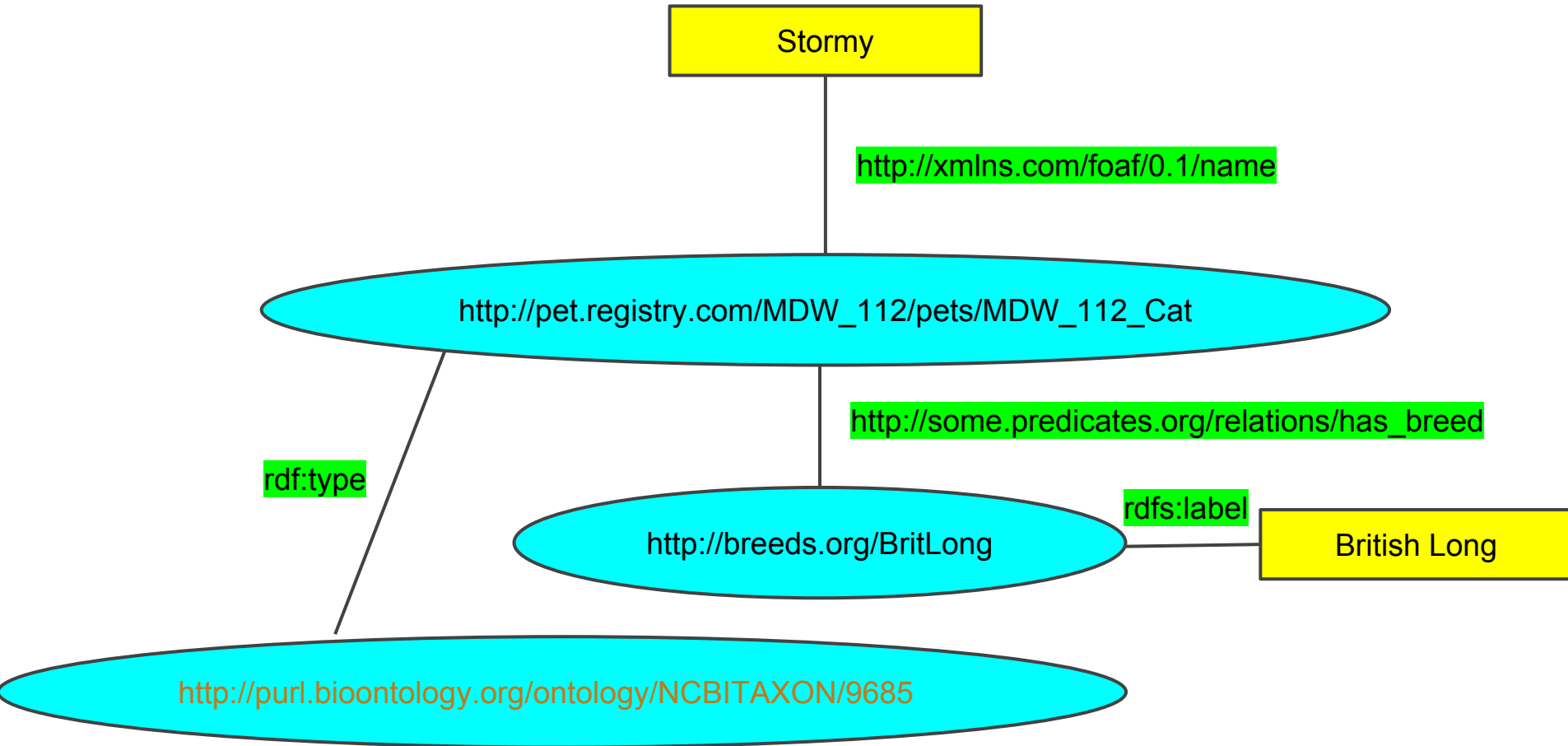
You can query RDF Databases (RDF databases are called “Triplestores”)

Many Triplestores include a “SPARQL Endpoint” where you can send queries in the SPARQL language.

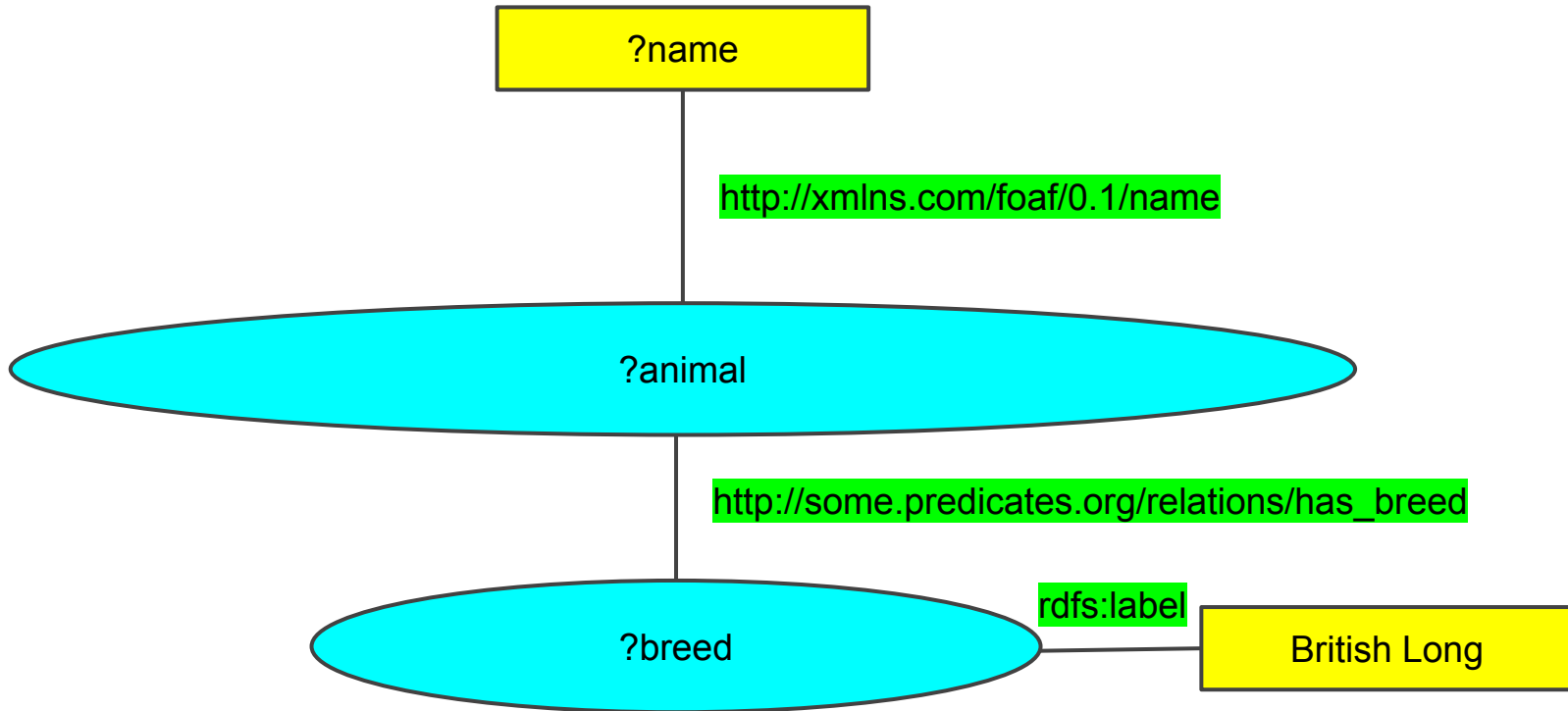
With SPARQL you create a “graph pattern” and ask the Triplestore to look for everything that matches the pattern

SPARQL looks quite a bit like RDF!

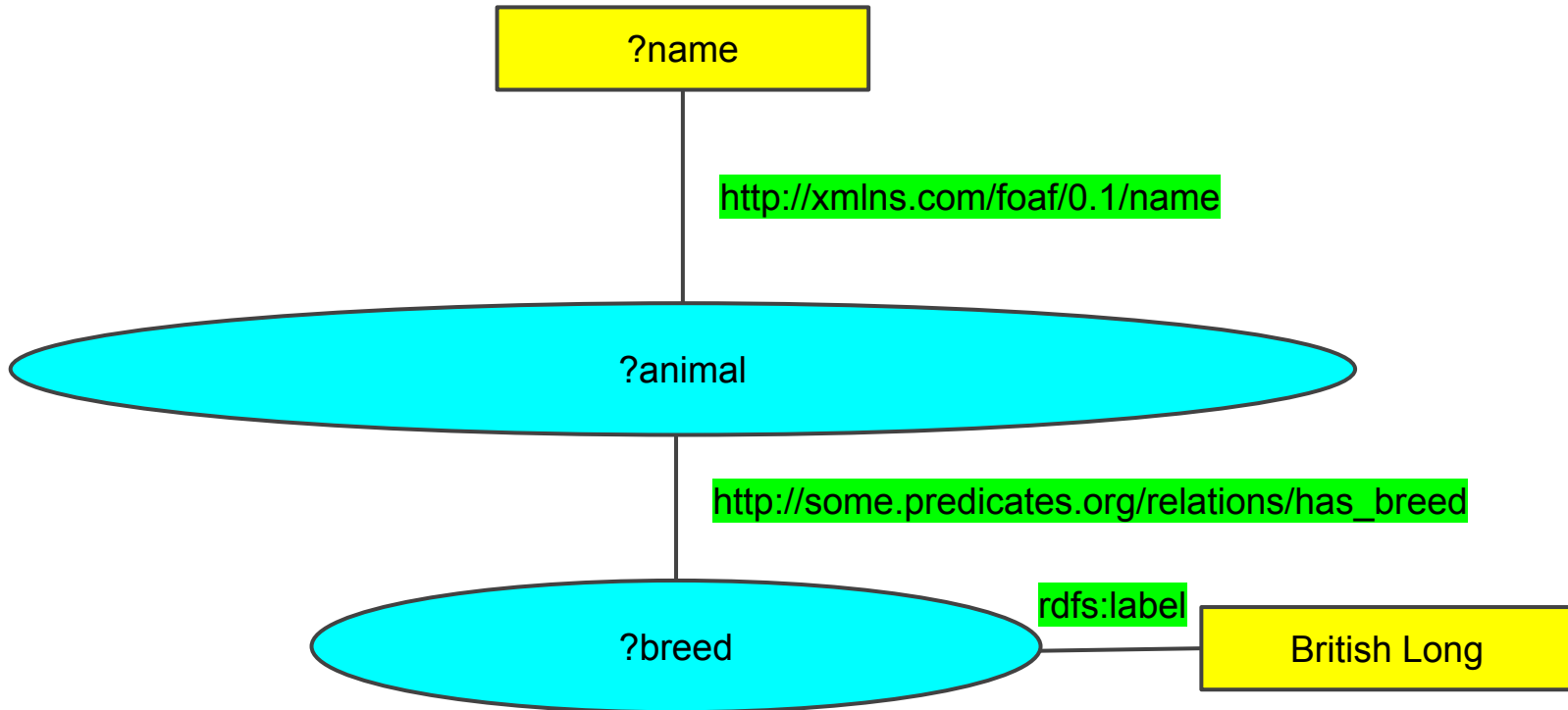
What are the names of all British Longs?



What are the names of all British Longs?



What are the names of all British Longs?



```
PREFIX pred: <http://some.predicates.org/relations/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
```

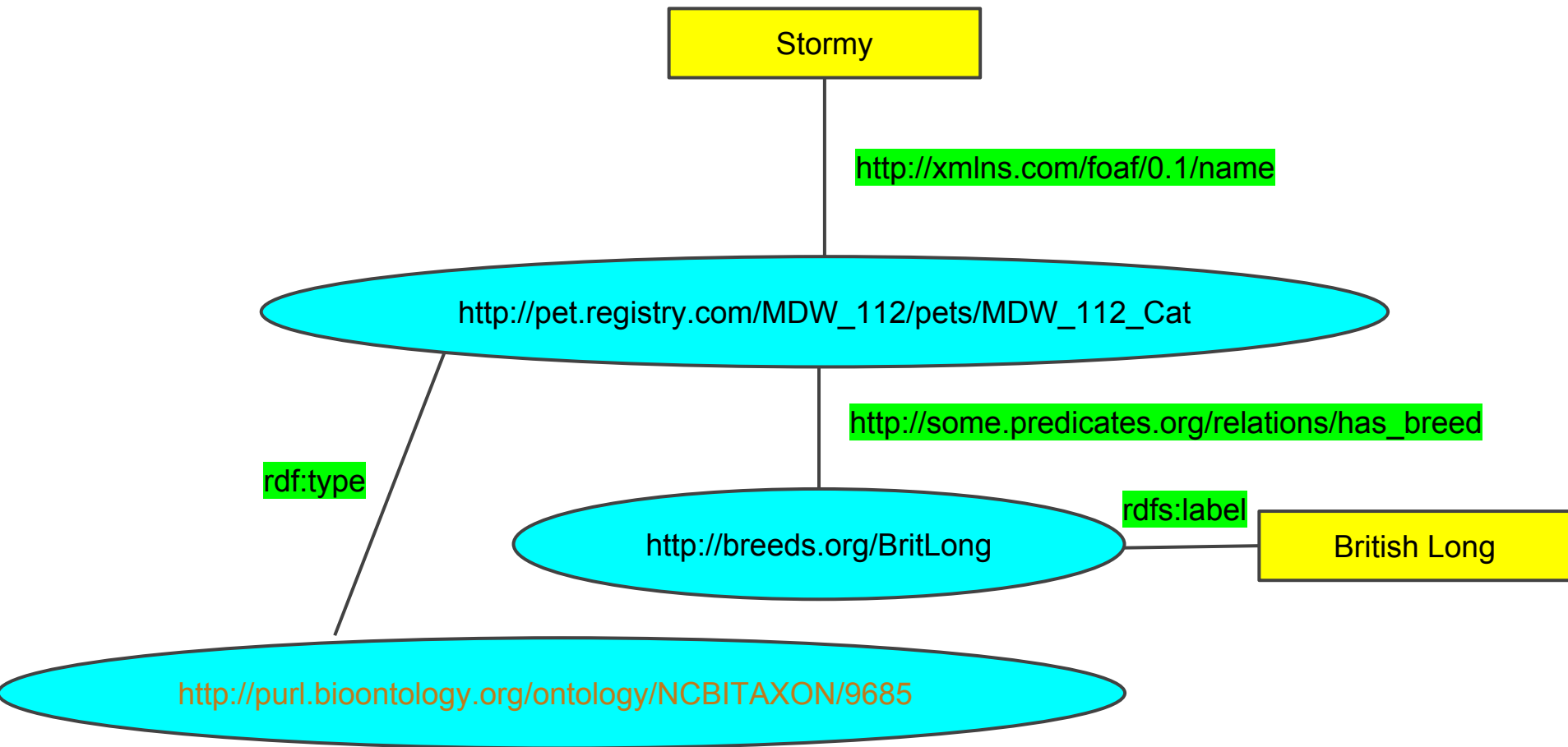
```
SELECT ?name
WHERE {
    ?breed rdfs:label "British Long" .
    ?animal pred:has_breed ?breed .
    ?animal foaf:name ?name
}
```

Now you.... Please create SPARQL for

“what breed is Stormy?”

“what species is Stormy?”

(try to solve both in one query!)




```
PREFIX pred: <http://some.predicates.org/relations/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
```

```
SELECT ?species ?breedname
WHERE {
    ?animal foaf:name "Stormy" .
    ?animal pred:has_breed ?breed .
    ?breed rdfs:label ?breedname .
    ?animal rdf:type ?species
}
```

`rdf:type` is so commonly used on the Semantic Web that it has its own special SPARQL abbreviation: “`a`”

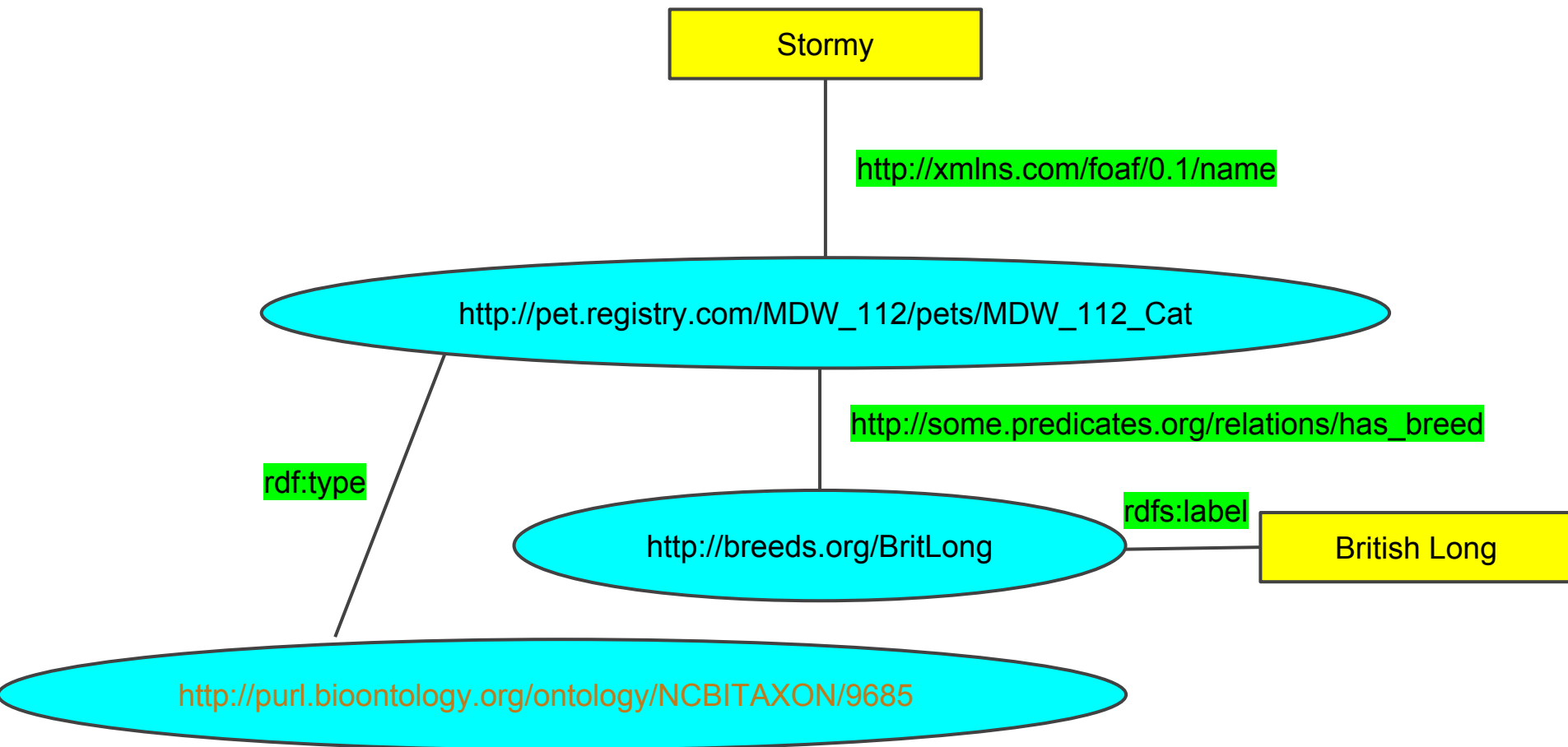
```
PREFIX pred: <http://some.predicates.org/relations/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
```

```
SELECT ?species ?breedname
WHERE {
    ?animal foaf:name "Stormy" .
    ?animal pred:has_breed ?breed .
    ?breed rdfs:label ?breedname .
    ?animal a rdf:type ?species
}
```

Stormy a Cat ⇒ Pumpkin rdf:type Cat

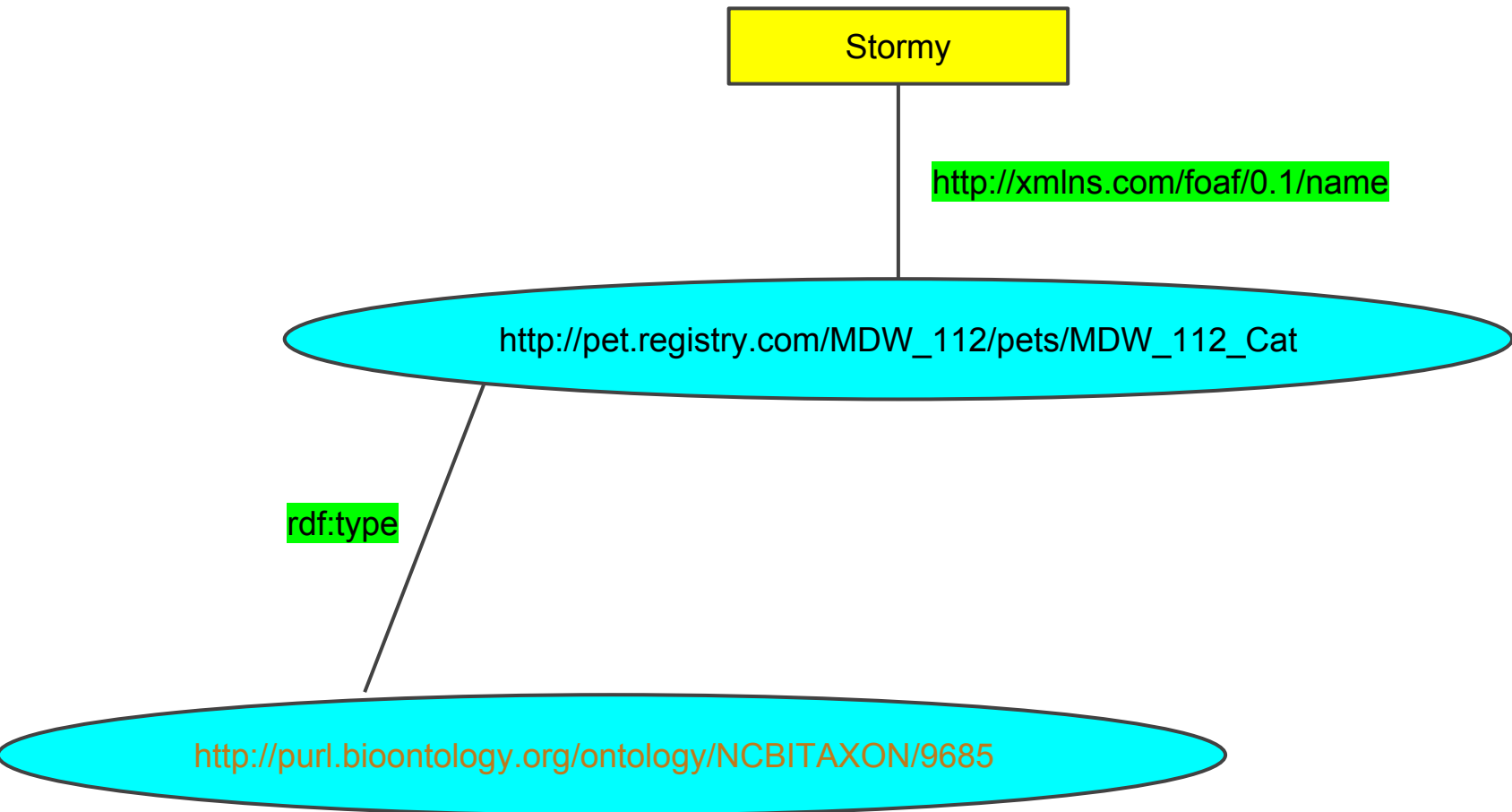
Now you.... Create the SPARQL for

“what predicate connects a cat and the cat’s name?”



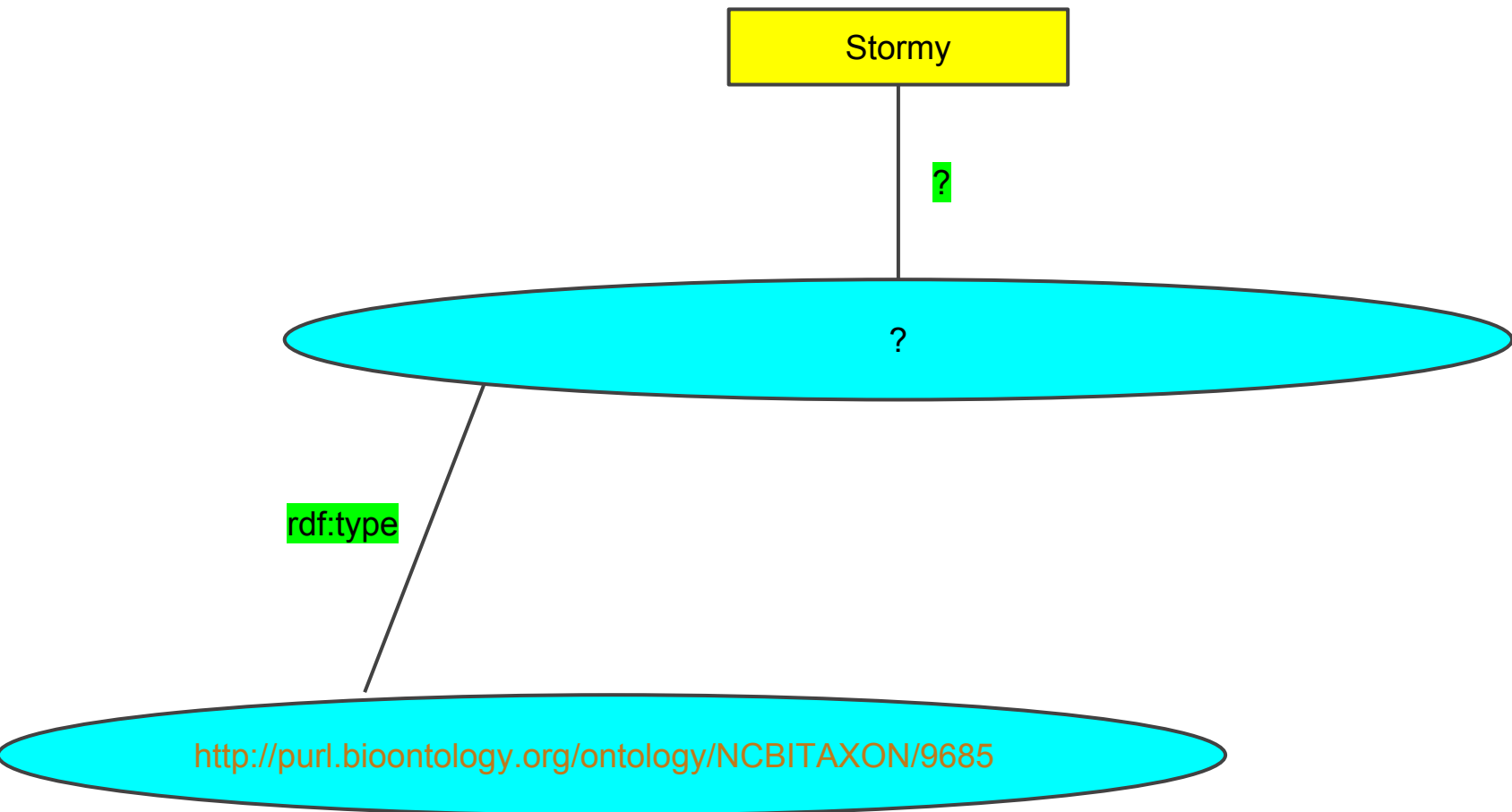
Now you.... Create the SPARQL for

“what predicate connects a cat and the cat’s name?”



Now you.... Create the SPARQL for

“what predicate connects a cat and the cat’s name?”



```
PREFIX pred: <http://some.predicates.org/relations/>  
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>  
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
```

```
SELECT ?predicate  
WHERE {  
    ?animal a <http://purl.bioontology.org/ontology/NCBITAXON/9685>  
    .  
    ?animal ?predicate "Stormy"  
}
```

- You can use a URL explicitly in your query
- Predicates can also be variables that you SELECT

```
PREFIX pred: <http://some.predicates.org/relations/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX taxon: <http://purl.bioontology.org/ontology/NCBITAXON/>
```

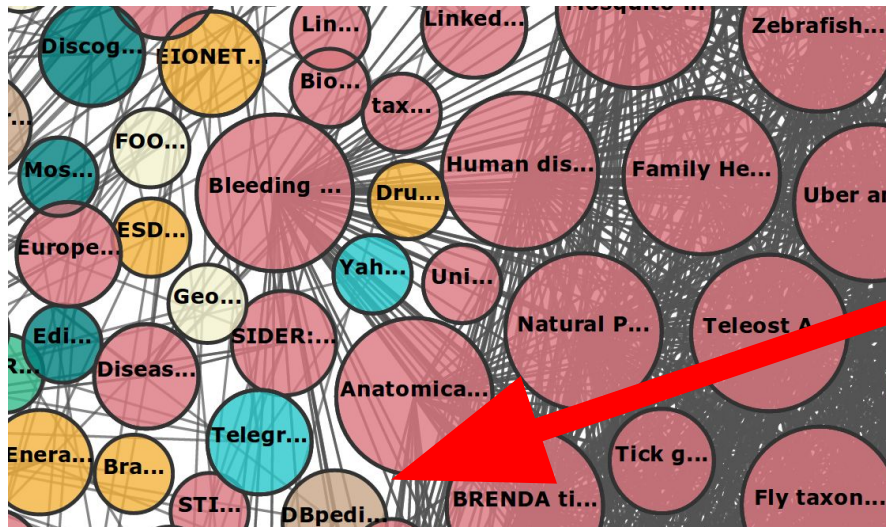
```
SELECT ?predicate
WHERE {
    ?animal a taxon:9685 .
    ?animal ?predicate "Stormy"
}
```

- You can use a URL explicitly in your query
- Predicates can also be variables that you SELECT

How do you find SPARQL endpoints

Google is your fastest option, but:

<http://lod-cloud.net/> that diagram is clickable



<http://sparql.uniprot.org>

Mark teaches SPARQL

Mark teaches SPARQL

(but only from April 15, 2016
until April 16, 2016)

Mark teaches SPARQL

(but only from April 15, 2016
until April 16, 2016)

What does that look like in RDF?

Some things are hard to do in RDF

Mark teaches SPARQL

(but only from April 15, 2016
until April 16, 2016)

What does that look like in RDF?

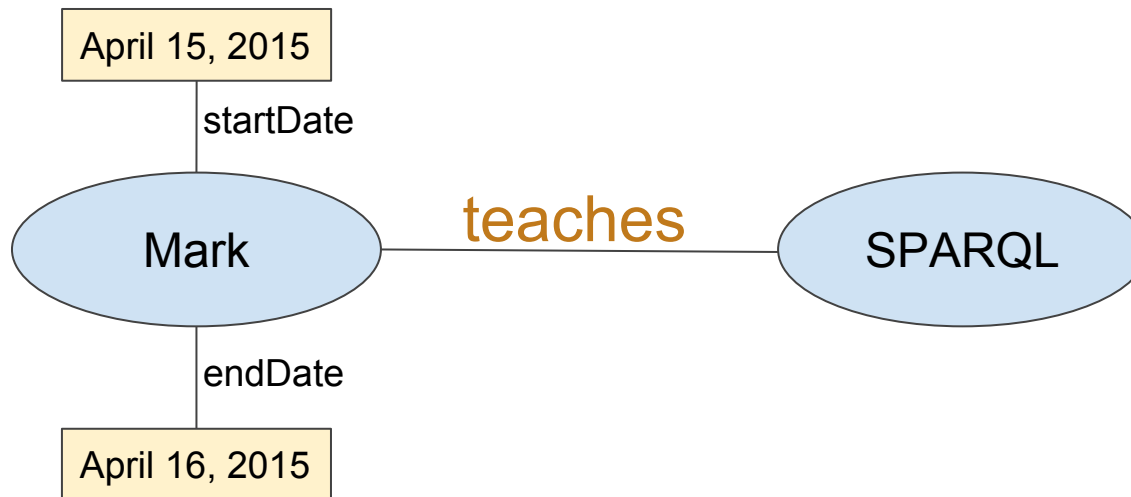


Some things are hard to do in RDF

Mark teaches SPARQL

(but only from April 15, 2016
until April 16, 2016)

What does that look like in RDF?

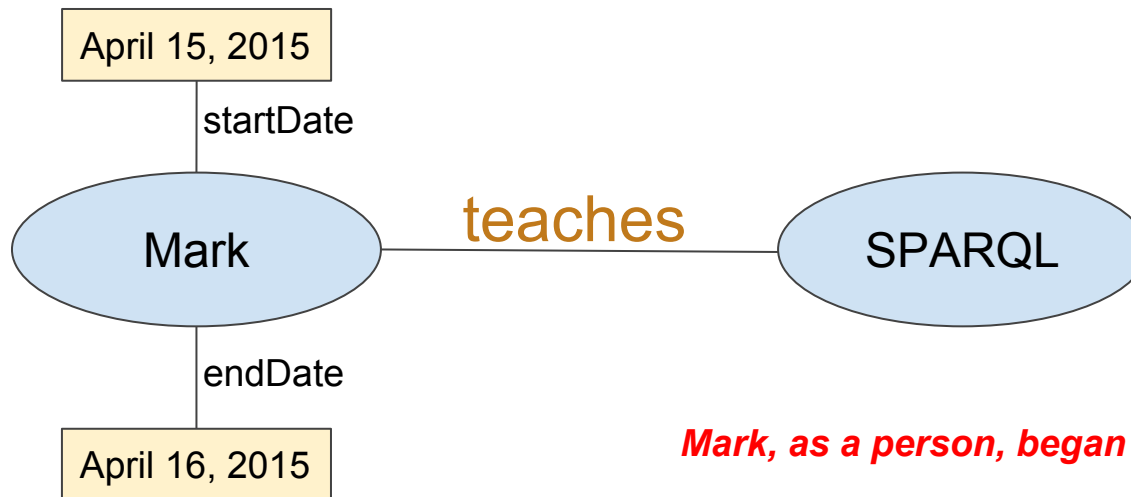


Some things are hard to do in RDF

Mark teaches SPARQL

(but only from April 15, 2016
until April 16, 2016)

What does that look like in RDF?



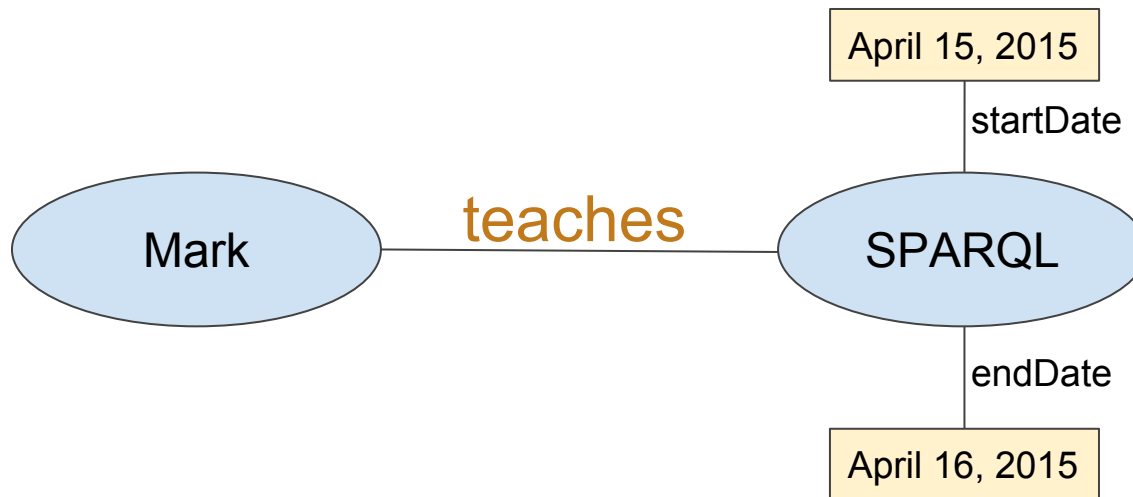
Mark, as a person, began April 15, 2015

Some things are hard to do in RDF

Mark teaches SPARQL

(but only from April 15, 2016
until April 16, 2016)

What does that look like in RDF?

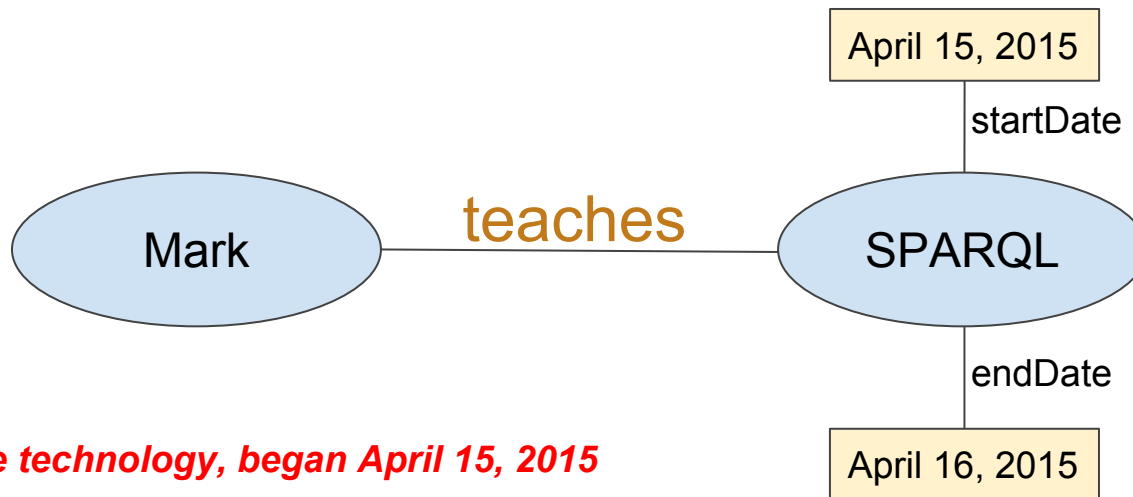


Some things are hard to do in RDF

Mark teaches SPARQL

(but only from April 15, 2016
until April 16, 2016)

What does that look like in RDF?



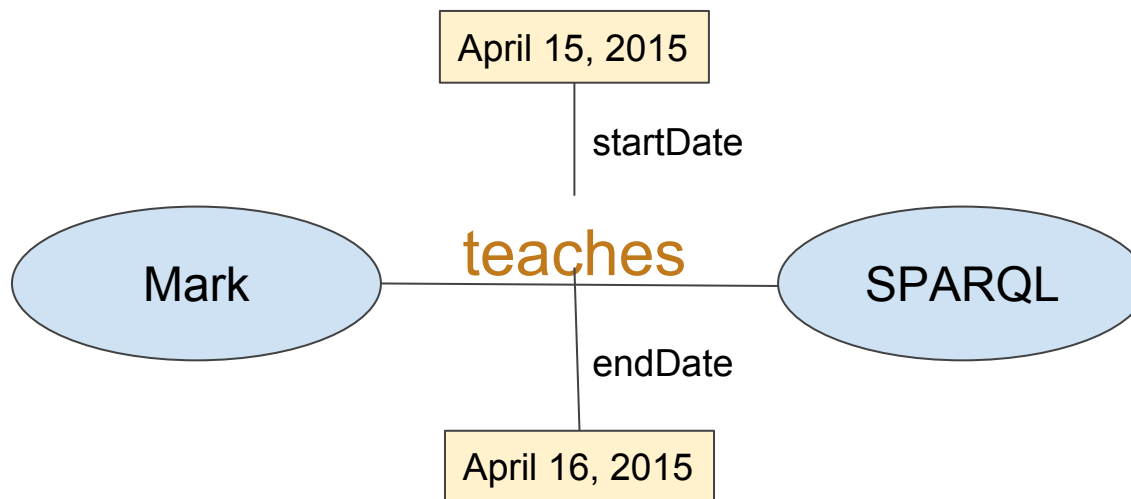
SPARQL, the technology, began April 15, 2015

Some things are hard to do in RDF

Mark teaches SPARQL

(but only from April 15, 2016
until April 16, 2016)

What does that look like in RDF?

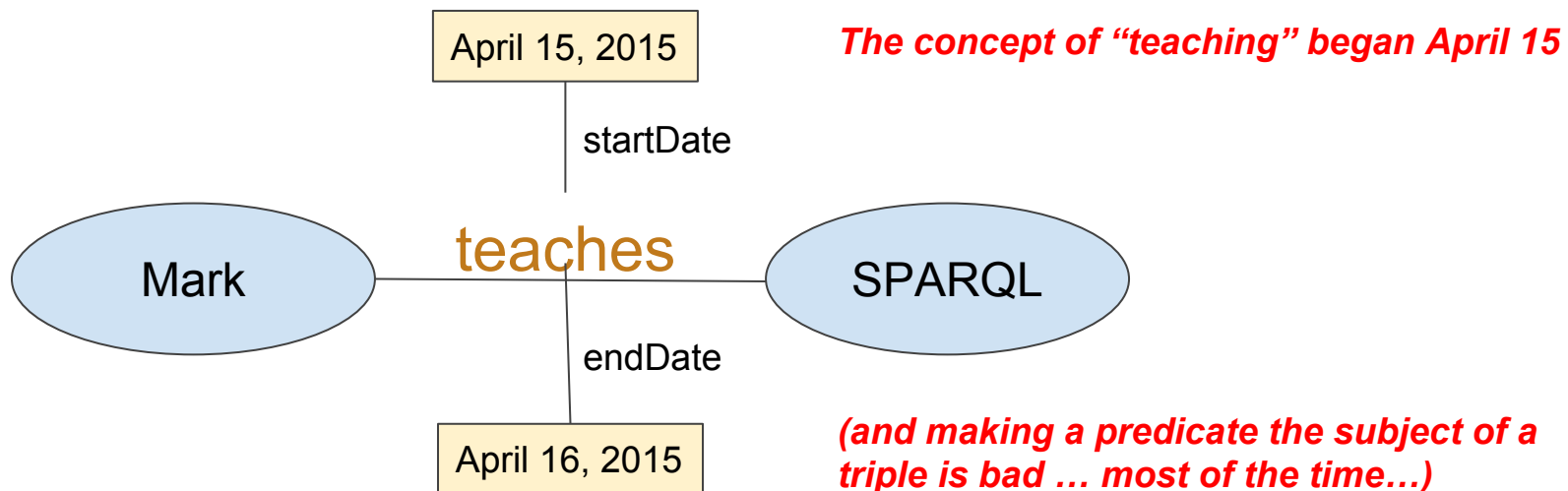


Some things are hard to do in RDF

Mark teaches SPARQL

(but only from April 15, 2016
until April 16, 2016)

What does that look like in RDF?

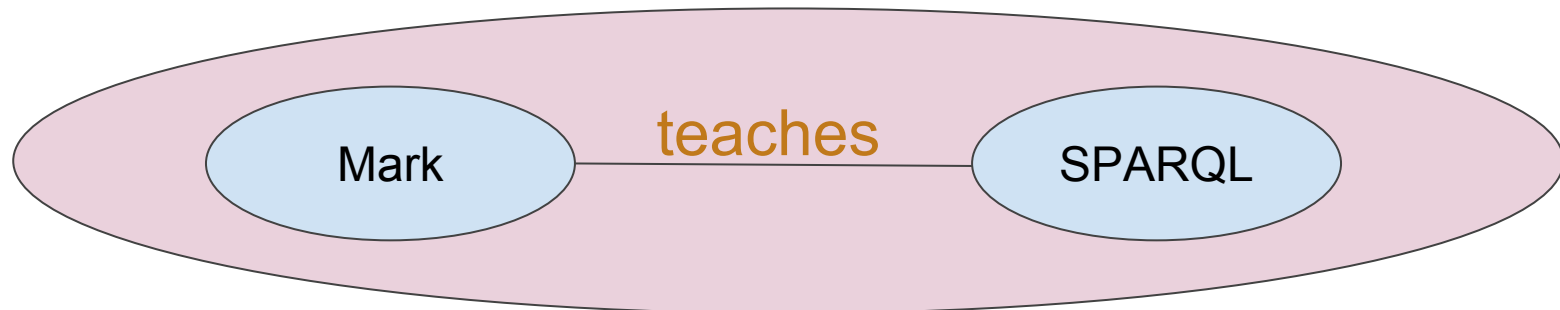


Some things are hard to do in RDF

Mark teaches SPARQL

(but only from April 15, 2016
until April 16, 2016)

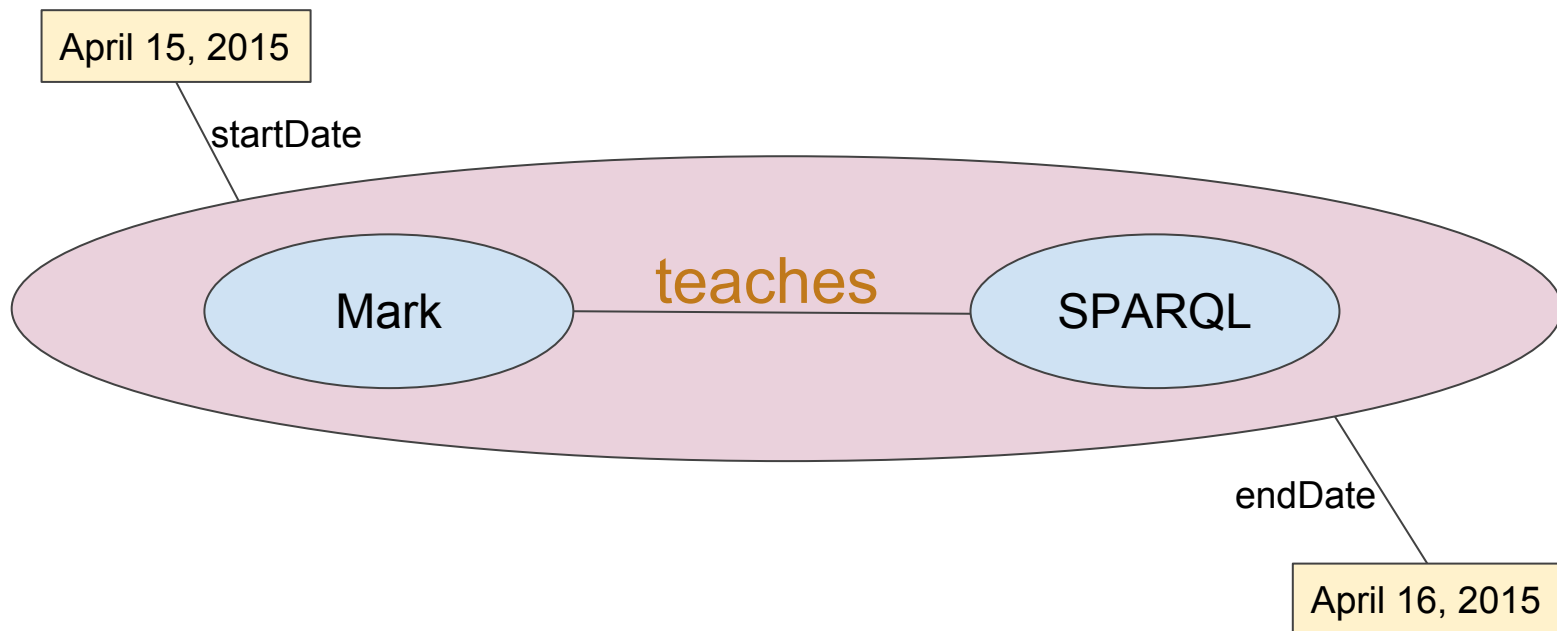
What does that look like in RDF?



You *really* want to say something about this entire triple

Named Graphs (a.k.a. “context”)

A way to group triples into a shared context
The **context** can become the subject of triples



How do you represent this “context”?

Context is just another URI...

E.g. <http://example.org/some/context/123456>

What does it look like in RDF?

Instead of Triples, you use Quads

S: `http://people.org/person/Mark`

P: `http://activity.org/teaches`

O: `http://technology.org/W3C/SPARQL`

C: `http://example.org/contexts/1234546`

S: `http://example.org/contexts/1234546`

P: `http://daterange.com/begins`

O: `"April 15, 2015"`

---- Null -----

What does it look like in RDF?

Instead of Triples, you use Quads

S: `http://people.org/person/Mark`

P: `http://activity.org/teaches`

O: `http://technology.org/W3C/SPARQL`

C: `http://example.org/contexts/1234546`

S: `http://example.org/contexts/1234546`

P: `http://daterange.com/begins`

O: `"April 15, 2015"`

The context URI =
the name
of the Named Graph

---- Null -----

What does it look like in RDF?

Instead of Triples, you use Quads

S: `http://people.org/person/Mark`

P: `http://activity.org/teaches`

O: `http://technology.org/W3C/SPARQL`

C: `http://example.org/contexts/1234546`

S: `http://example.org/contexts/1234546`

P: `http://daterange.com/begins`

O: `"April 15, 2015"`

The information about
that context

---- Null -----

What does it look like in RDF?

Instead of Triples, you use Quads

S: `http://people.org/person/Mark`

P: `http://activity.org/teaches`

O: `http://technology.org/W3C/SPARQL`

C: `http://example.org/contexts/1234546`

S: `http://people.org/person/Mark`

P: `http://activity.org/resides`

O: `http://hotels.com/NH_Utrecht`

C: `http://example.org/contexts/1234546`

The triples in this graph
share a context, and are in
the same “Named Graph”

What does this look like in SPARQL

Browse to: `http://training.fairdata.solutions/sparql`

linkeddata.systems:8890/sparql

Virtuoso SPARQL Query Editor

[About](#) | [Namespace Prefixes](#) | [Inference rules](#)

Default Data Set Name (Graph IRI)

`http://example.org/contexts/1234546`

Query Text

This is where you put the named graph URI/URL

What does this look like in SPARQL

Browse to: <http://training.fairdata.solutions/sparql>

linkeddata.systems:8890/sparql

Virtuoso SPARQL Query Editor

[About](#) | [Namespace Prefixes](#) | [Inference rules](#)

Default Data Set Name (Graph IRI)

Query Text

```
SELECT * WHERE {
  GRAPH <http://example.org/contexts/1234546> {
    ?s ?p ?o
  }
}
```

Or you can put the named graph in the SPARQL query

How do people use context?

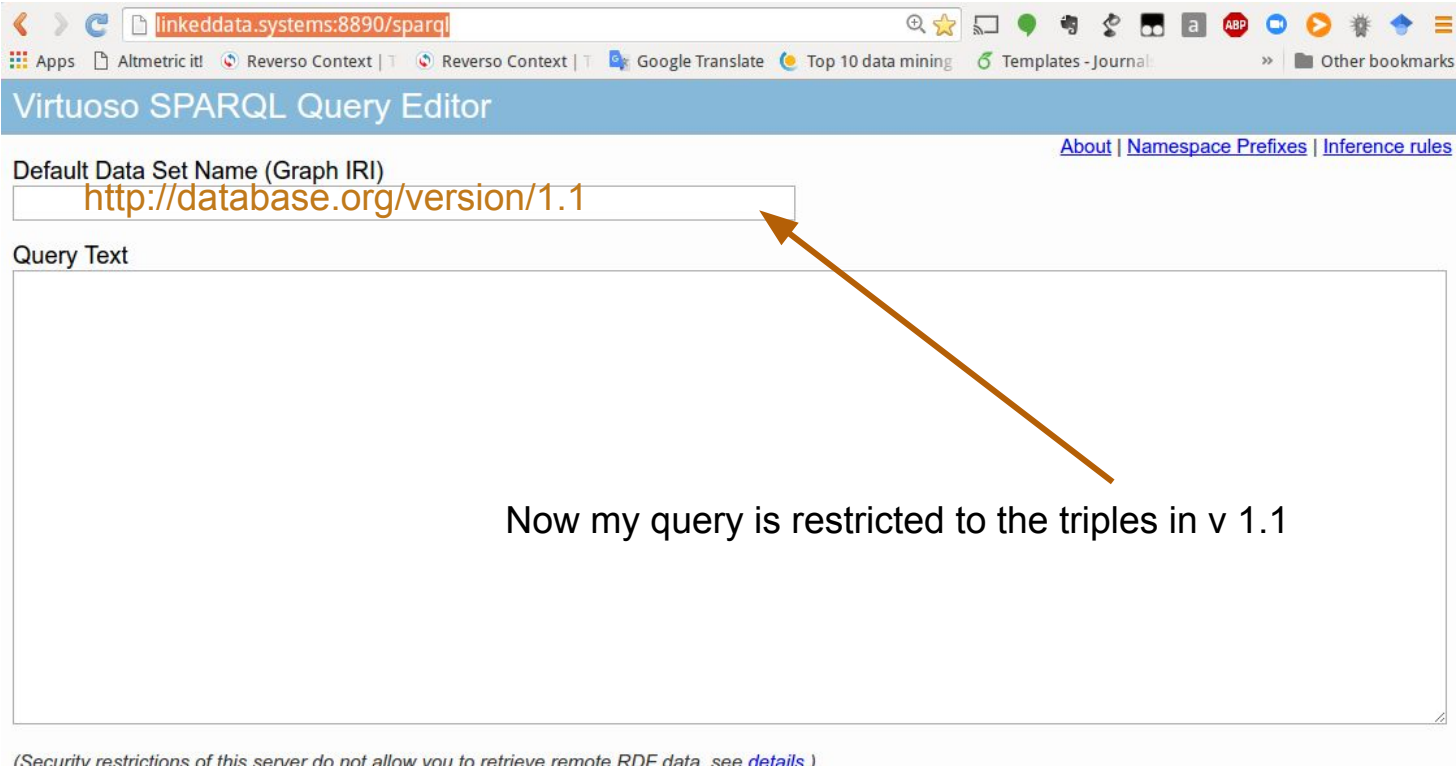
Typical uses include:

- Versioning
 - <http://database.org/version/1.1>
 - <http://database.org/version/1.2>
 - You query whichever version you need by putting the appropriate named graph in the box above the SPARQL query
- NanoPublications
 - Nanopubs use named graphs to attach citation/authorship information to scientific data and results. “This set of triples came from...”

What does this look like in SPARQL

Browse to: `http://ltraining.fairdata.solutions/sparql`

This is a “shortcut”, since most of the time you will always want to query the same graph. This allows you to set the desired graph as default.



Virtuoso SPARQL Query Editor

[About](#) | [Namespace Prefixes](#) | [Inference rules](#)

Default Data Set Name (Graph IRI)
`http://database.org/version/1.1`

Query Text

Now my query is restricted to the triples in v 1.1

(Security restrictions of this server do not allow you to retrieve remote RDF data, see [details](#).)

What does this look like in SPARQL

Browse to: `http://linkeddata.systems:8890/sparql`

This is a “shortcut”, since most of the time you will always want to query the same graph. This allows you to set the desired graph as default.

Virtuoso SPARQL Query Editor

[About](#) | [Namespace Prefixes](#) | [Inference rules](#)

Default Data Set Name (Graph IRI)

Query Text

Now my query is restricted to the triples in v 1.1

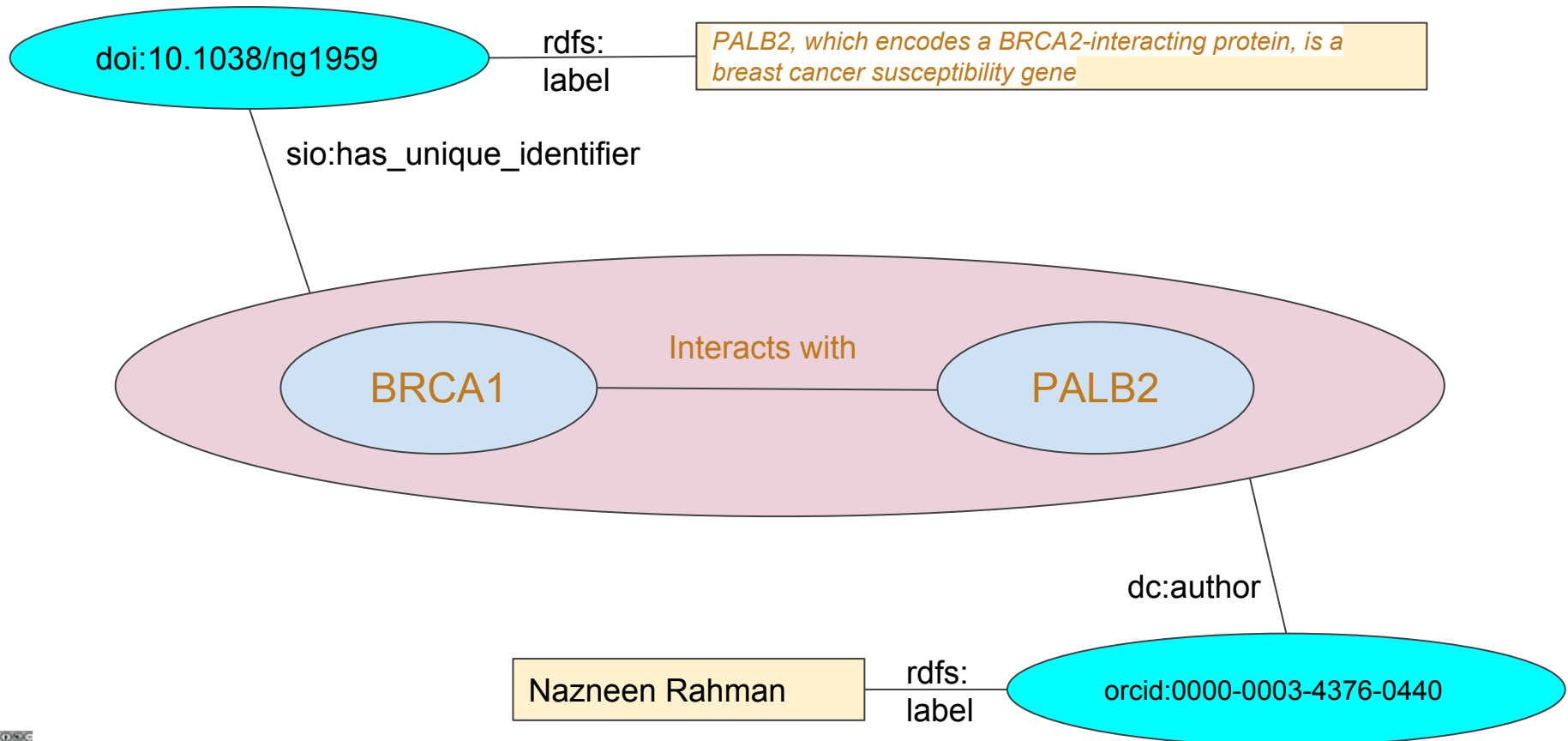
(Security restrictions of this server do not allow you to retrieve remote RDF data, see [details](#).)

How do people use context?

Typical uses include:

- Versioning
 - <http://database.org/version/1.1>
 - <http://database.org/version/1.2>
 - You query whichever version you need by putting the appropriate named graph in the box above the SPARQL query
- NanoPublications
 - Nanopubs use named graphs to attach citation/authorship information to scientific data and results. “This set of triples came from...”

A way to attach citation information to data



We will revisit Named Graphs later...

In a few hours we will discuss the **World Wide Web Consortium's** proposal for the “**Linked Data Platform**”, which uses Named Graphs to “simulate” a filesystem, with folders and files.

Important Point: We can do FAIR using ONLY W3C-approved standards.

FAIR invents NOTHING