

# Databases

## Graph Databases Motivation

André Santanchè e Patrícia Cavoto  
Institute of Computing - UNICAMP  
August 2015

# Graph Database

Graph as the basic data model of a database



# Why Graphs?

# Why Graphs?

- The Web Effect
  - Linked Data
  - Social Networks and Social Content
- Sharing and interconnecting
- Complex Networks



# Why Graphs? Linked Data

# Wikipedia

Firefox W Paris - Wikipedia, the free encyclopedia + en.wikipedia.org/wiki/Paris

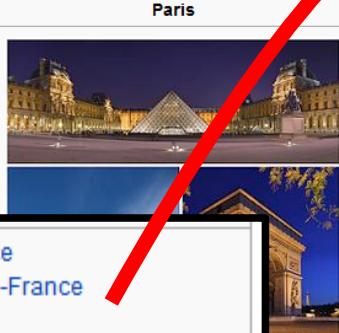
Article Talk Read View source View history Search

## Paris

From Wikipedia, the free encyclopedia

This article is about the capital of France. For other uses, see [Paris \(disambiguation\)](#).

**Paris** (English /pærɪs/, /pərɪs/; French: [paʁi] (listen)) is the capital and most populous city of France. It is situated on the River Seine, in the north of the country, at the heart of the Île-de-France region. Within its administrative limits (the 20 arrondissements), the city had 2,234,105 inhabitants in 2009 while its metropolitan area is one of the largest population centres in Europe with more than 12 million inhabitants.



Country	France
Region	Île-de-France
Department	Paris
Subdivisions	20 arrondissements
Government	• Mayor (2008–14) Bertrand Delanoë (PS)
Area <sup>[1]</sup>	• Urban (2010) 2,844.8 km <sup>2</sup> (1,098.4 sq mi) • Metro (2010) 17,174.4 km <sup>2</sup> (6,631.1 sq mi) • Land <sup>1</sup> 105.4 km <sup>2</sup> (40.7 sq mi)
Population (2010) <sup>[5]</sup>	• Rank 1st in France

Infobox

Firefox W Île-de-France - Wikipedia, the free encyclopedia + en.wikipedia.org/wiki/Île-de-France\_(region)

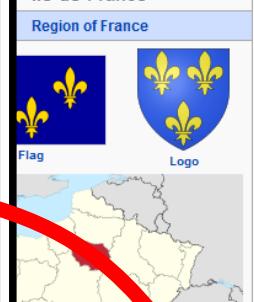
Article Talk Read Edit View history Search

## Île-de-France

From Wikipedia, the free encyclopedia (Redirected from [Île-de-France \(region\)](#))

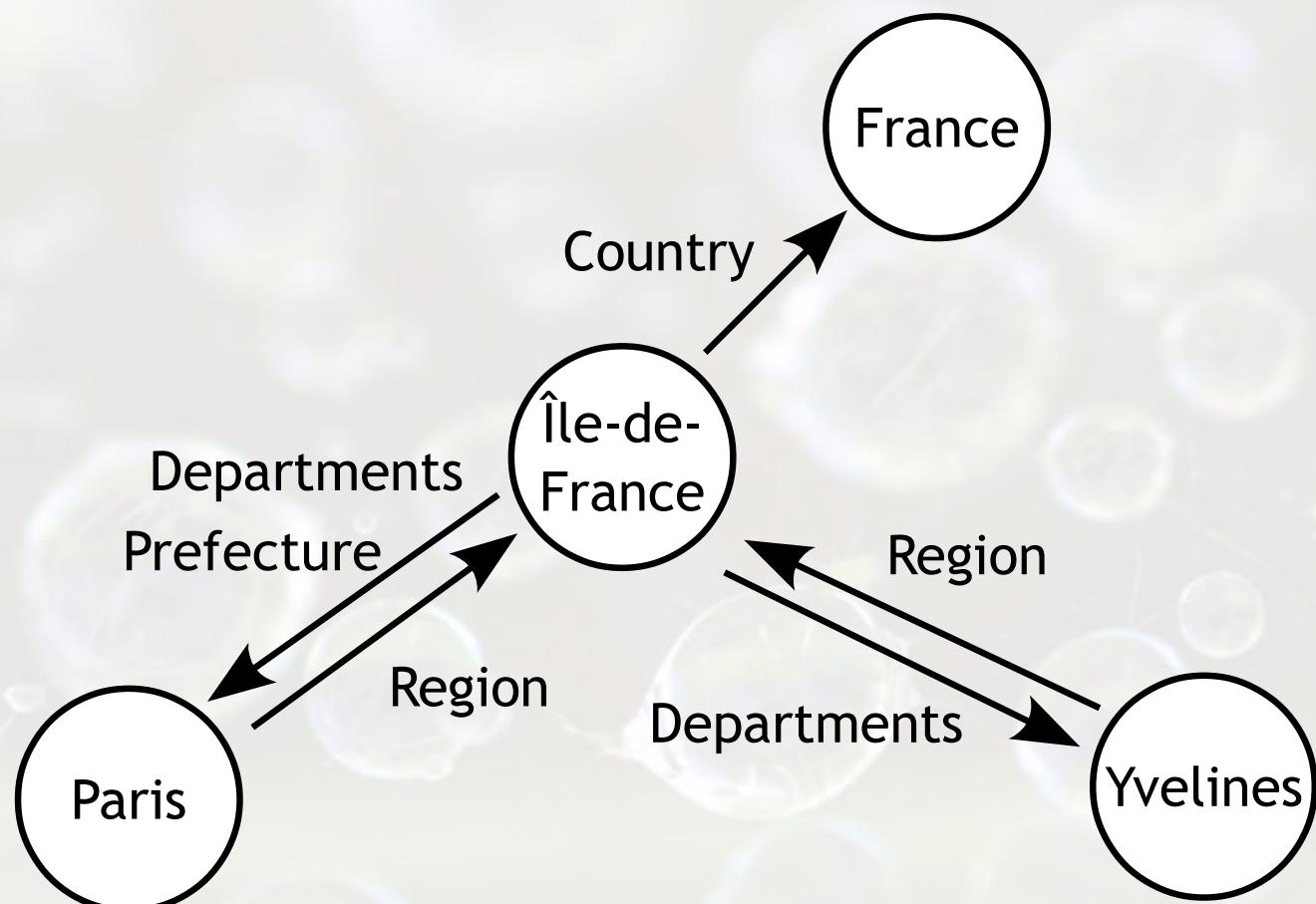
For other uses, see [Île-de-France \(disambiguation\)](#).

**Île-de-France** (French: [il də fʁɑ̃s]) is a region of France located in the north-central part of the country, bordering the English Channel to the west and the North Sea to the northwest. It is the second-most populous region in France, with a population of 12,012,000 inhabitants in 2012.

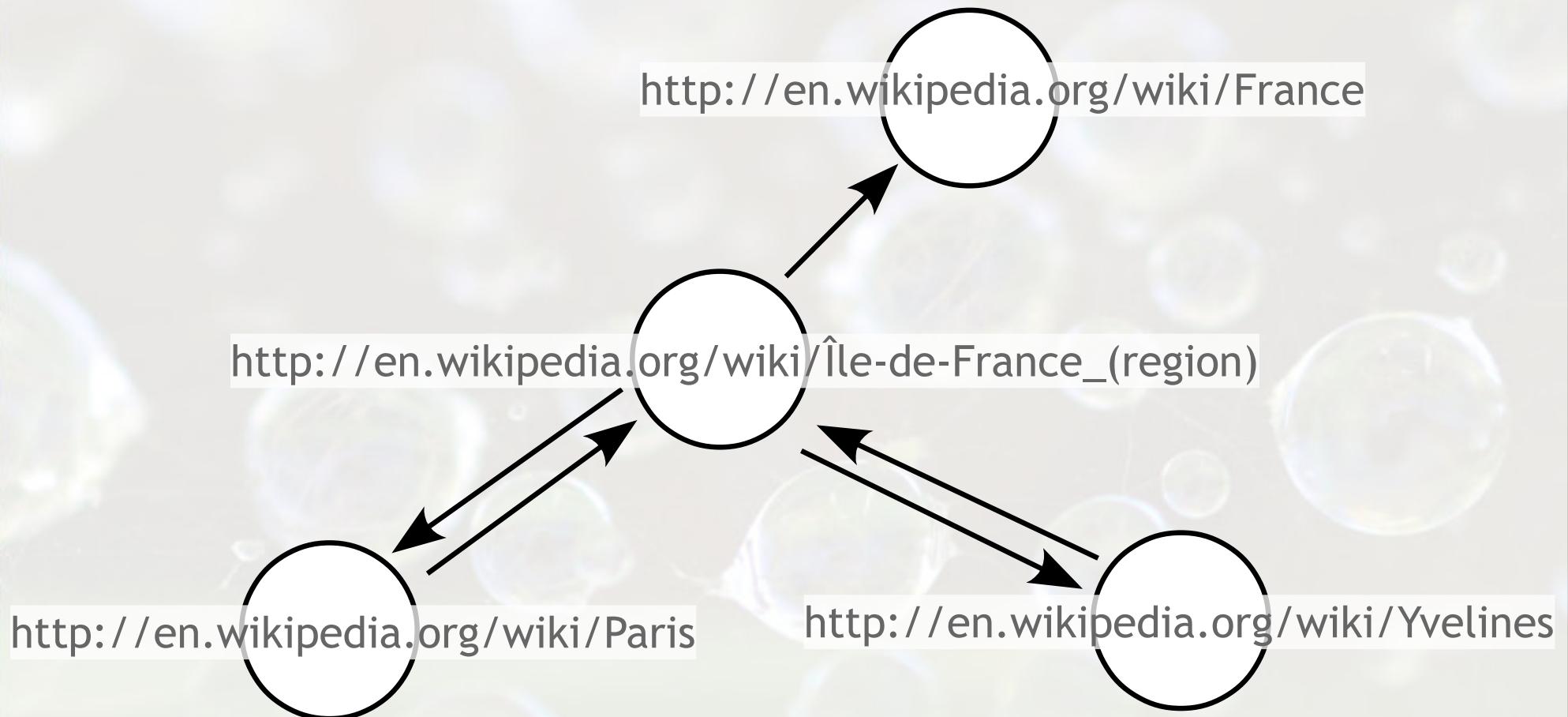


Country	France
Prefecture	Paris
Departments	8 Paris Essonne Hauts-de-Seine Seine-Saint-Denis Seine-et-Marne Val-de-Marne Val-d'Oise Yvelines
Government	• President Jean-Paul Huchon (PS)
Area	• Total 12,012 km <sup>2</sup> (4,638 sq mi)
Population (2012) <sup>[1]</sup>	• Total 11,914,812 • Density 990/km <sup>2</sup> (2,600/sq mi)

# DBpedia



# DBpedia (URIs)



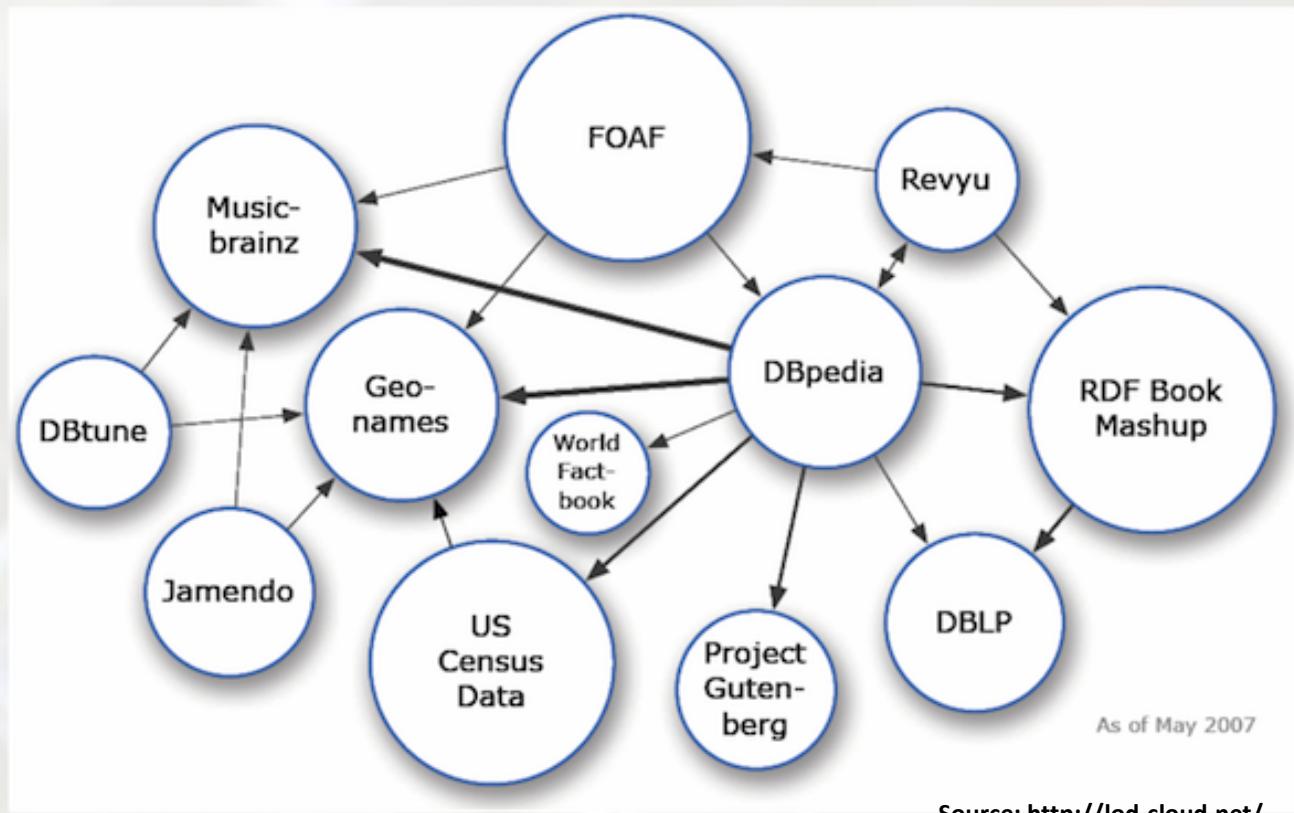
# DBpedia - English

- **4 million things**
- **3.22 million classified in a consistent ontology**
  - 832,000 persons
  - 639,000 places (427,000 populated)
  - 372,000 creative works
    - 116,000 music albums; 78,000 films; 18,500 video games
  - 209,000 organizations
  - 226,000 species
  - 5,600 diseases.

# DBpedia - International

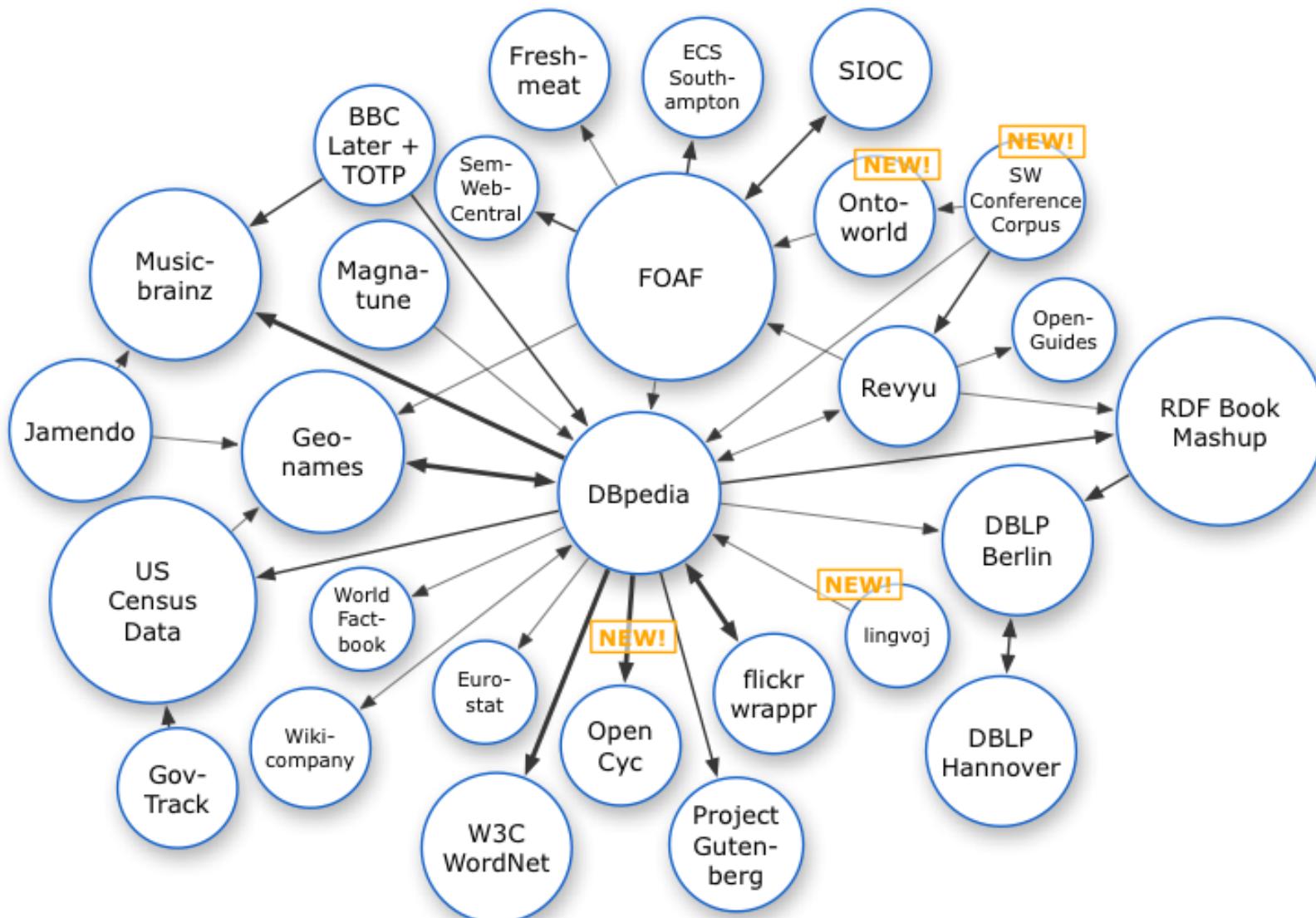
- 119 languages
- 24.9 million things
- 16.8 million interlinked with English
- 12.6 million unique things

# Linked Data



Datasets published following Linked Data ‘format’: **05/2007**

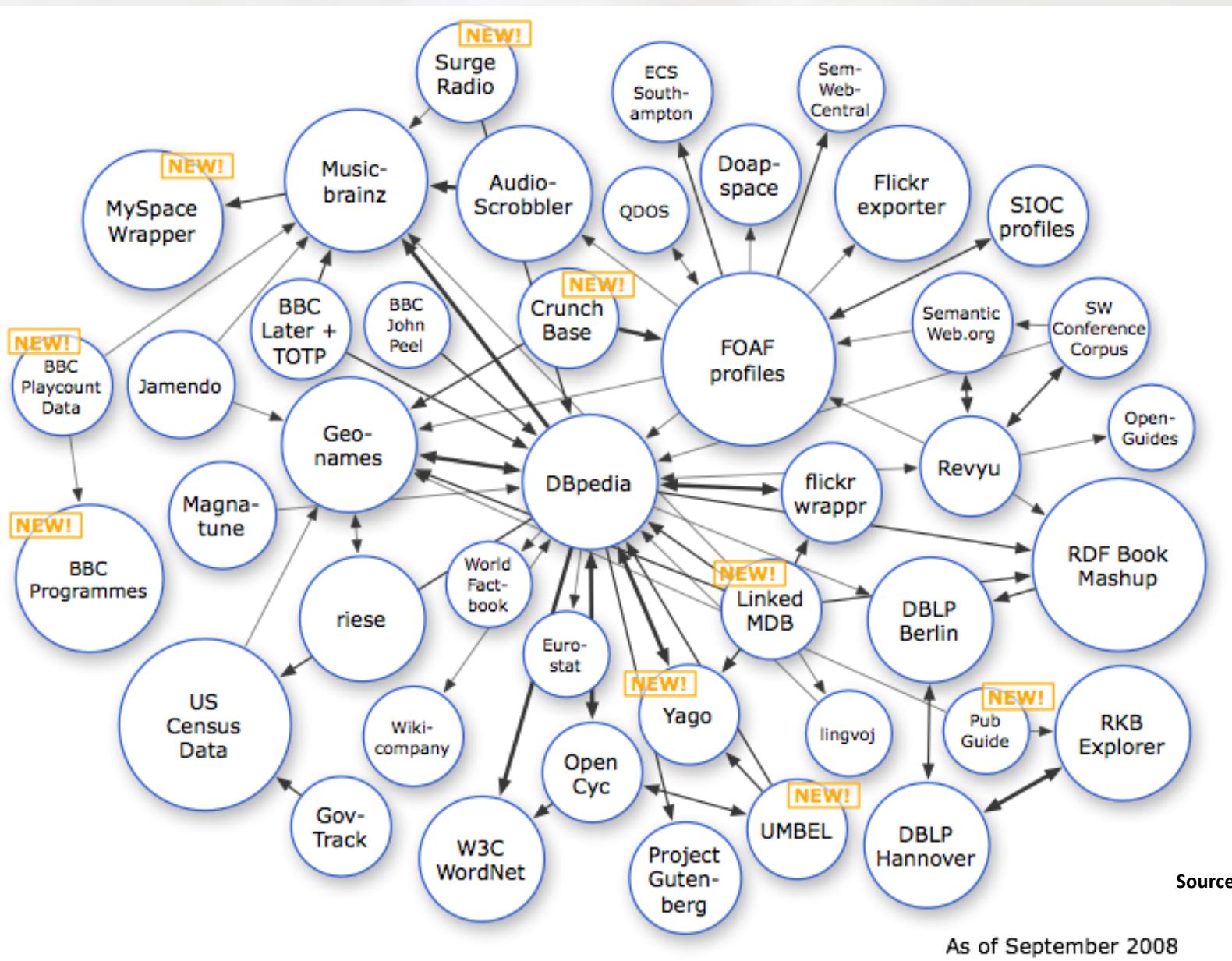
# Linked Data



Source: <http://lod-cloud.net/>

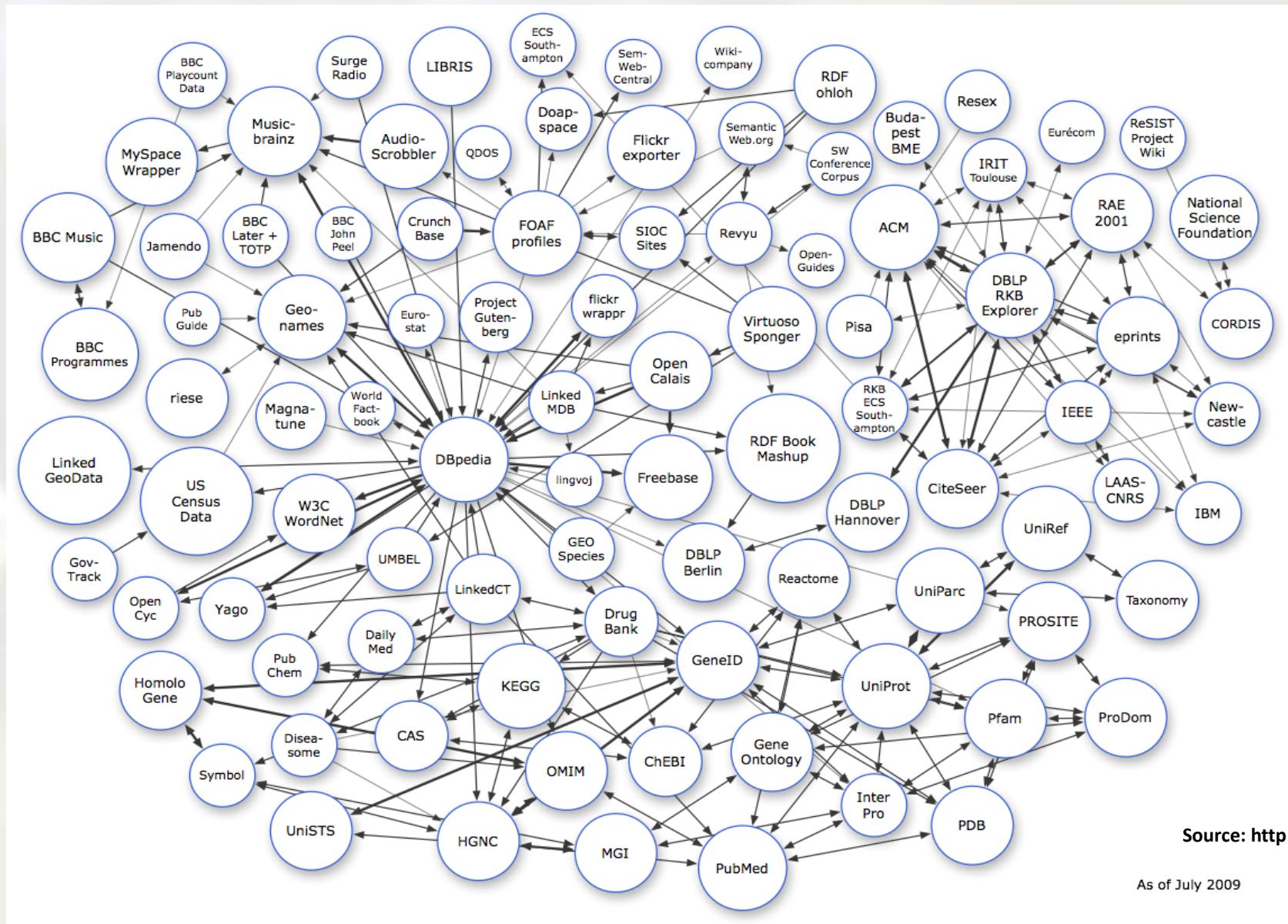
Datasets published following Linked Data 'format': **11/2007**

# Linked Data



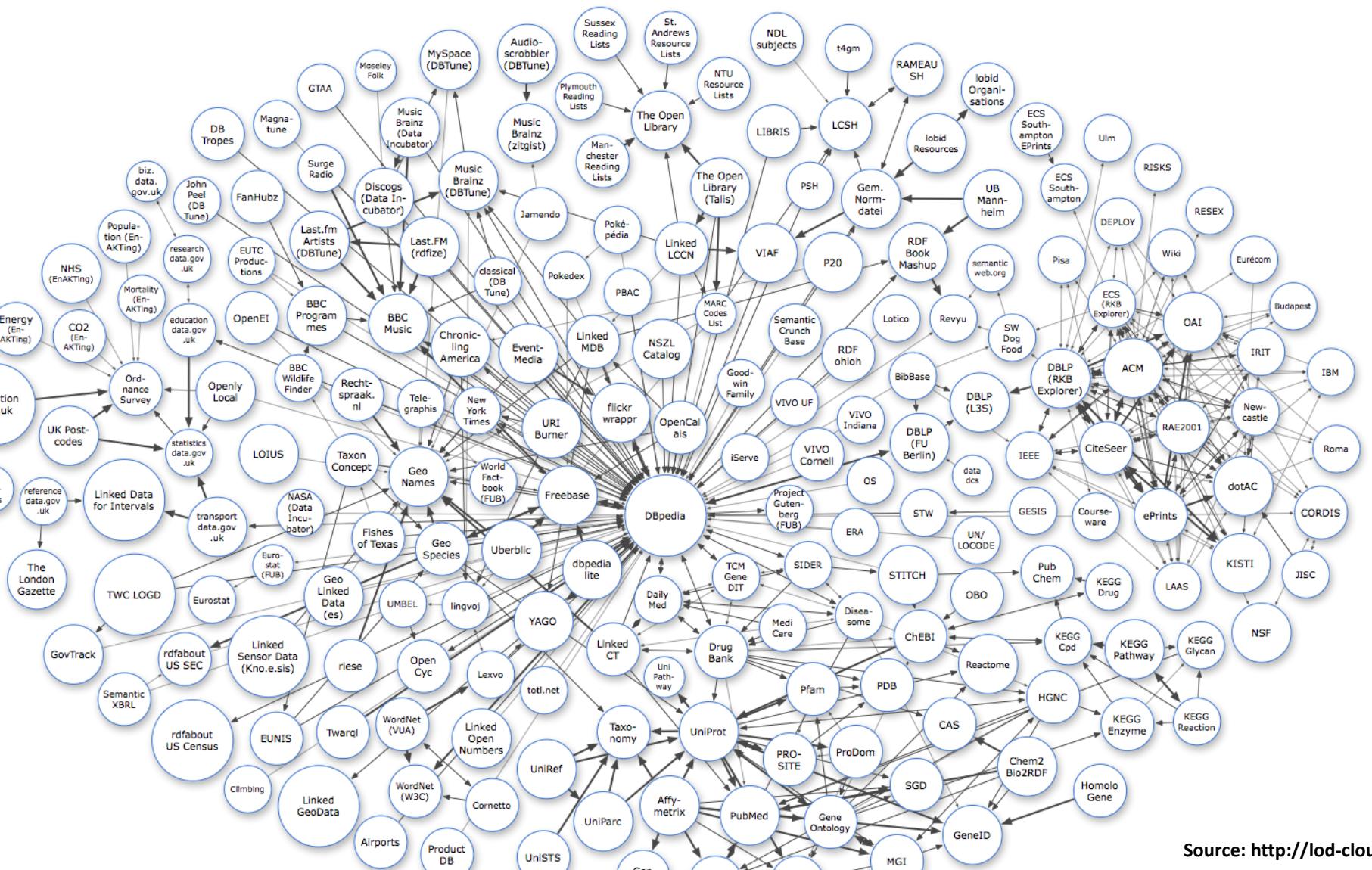
## Datasets published following Linked Data ‘format’: 2008

# Linked Data



Datasets published following Linked Data ‘format’: **2009**

# Linked Data

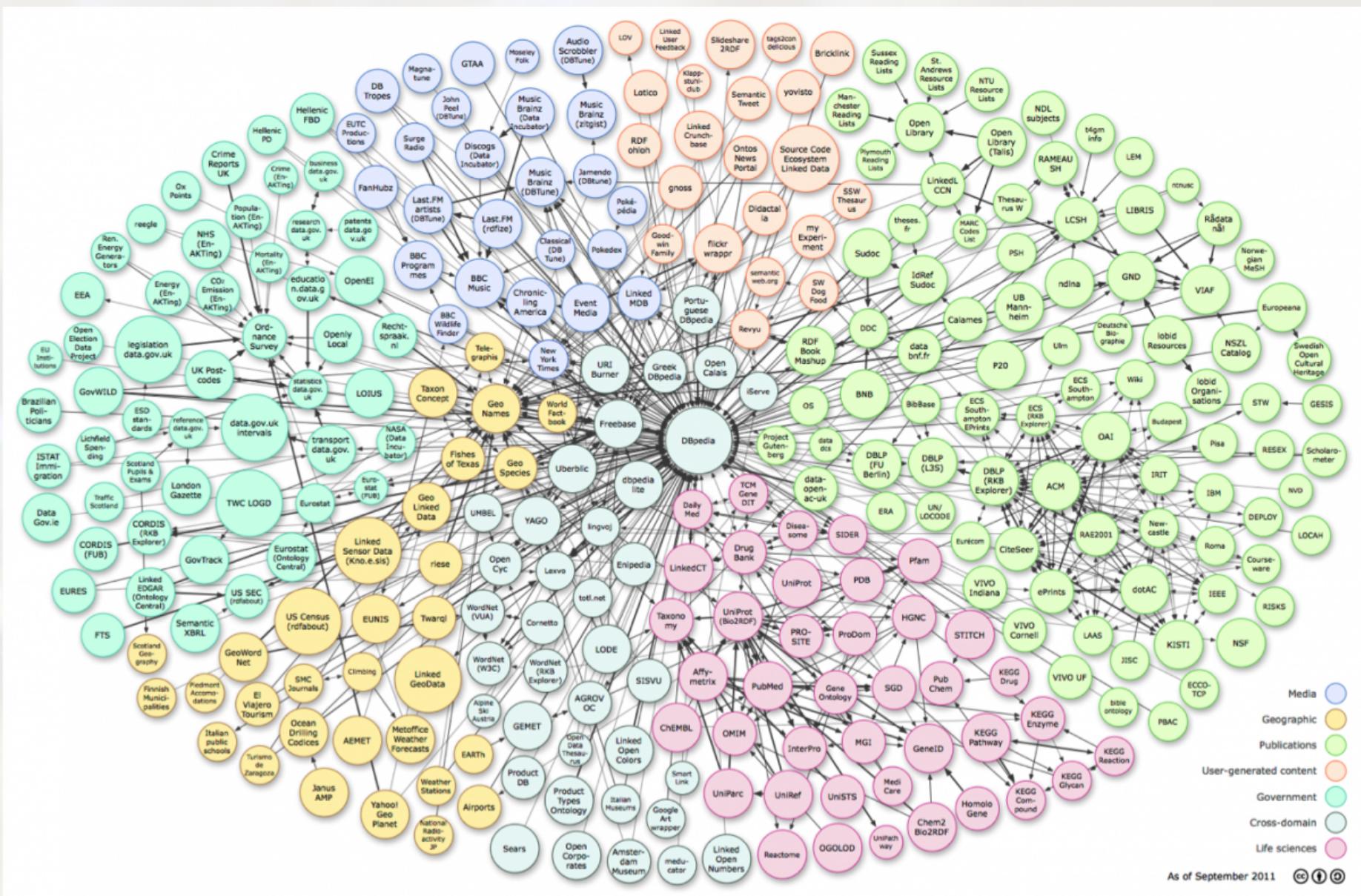


Source: <http://lod-cloud.net/>

As of September 2010   

Datasets published following Linked Data ‘format’: **2010**

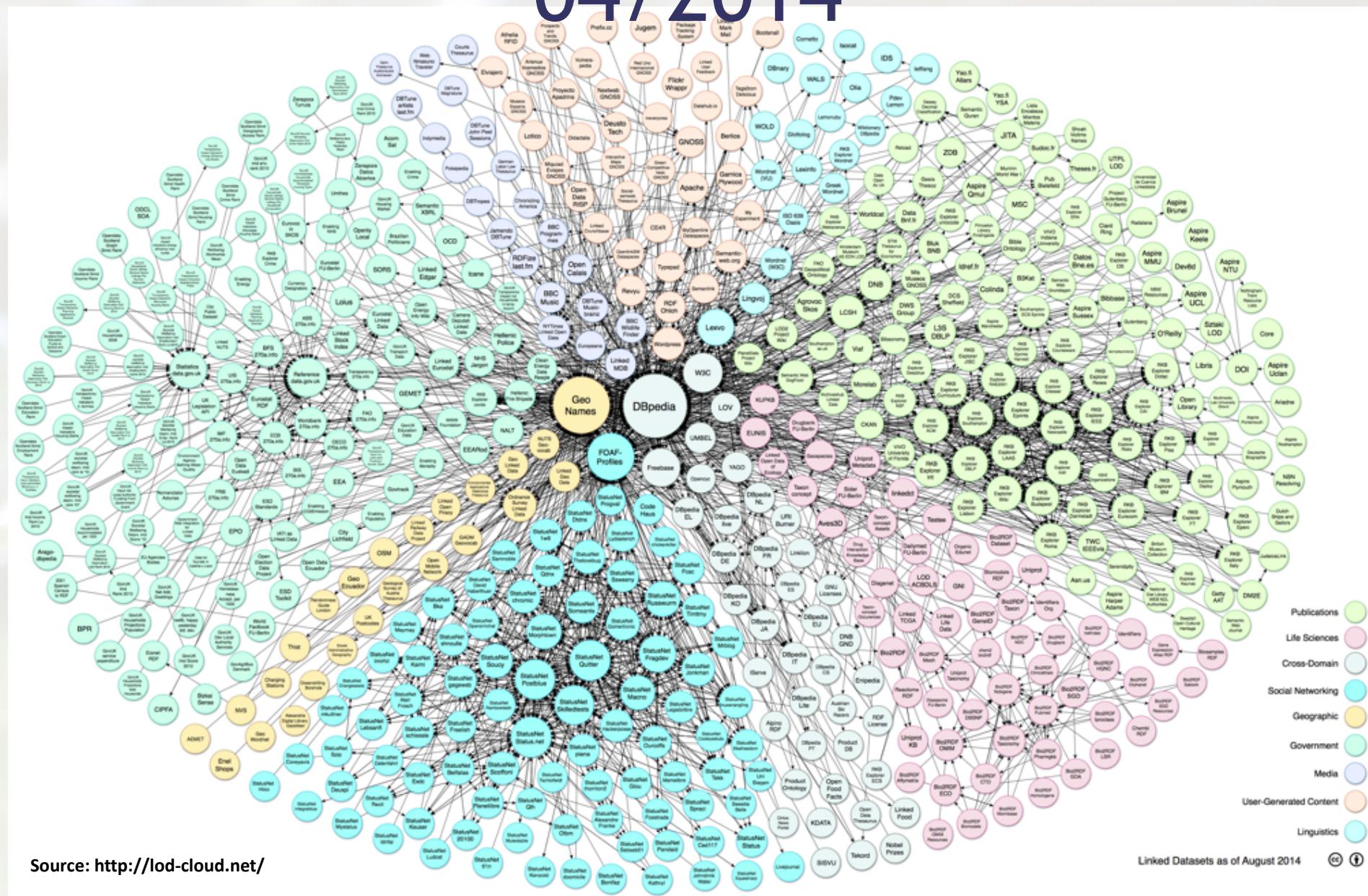
# Linked Data



Datasets published following Linked Data ‘format’: **2011**

# Linked Data

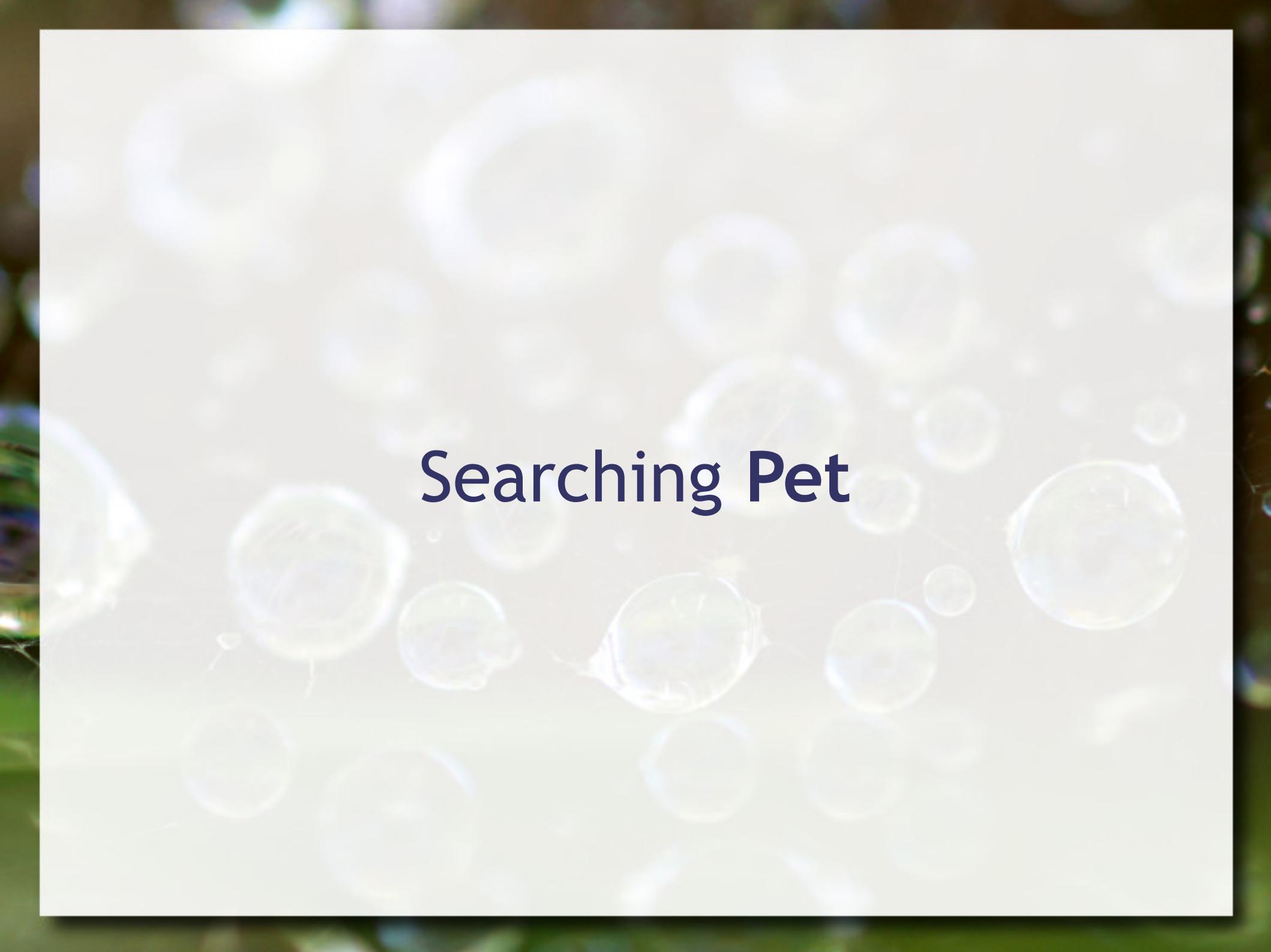
## 04/2014





Why Graphs?

# Social Networks and Social Content



# Searching Pet

# Pet Shop Boys

Anirudh Koul



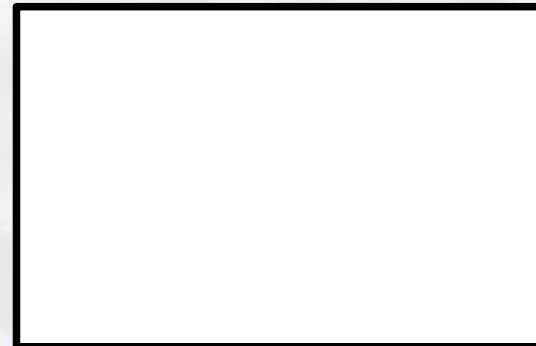
pet shop boys, pet, concert, tour



cat, kittens, eyes,  
ears, pet, animal



dog, pet, animal,  
funny, glasses



dog, pet, alaskan malamute



cat, kitty, eyes,  
pretty



cat, kitten, garden,  
pet



recycle, pet, plastic bottle,  
polyethylene terephthalate



recycle, pet, plastic bottle,  
polyethylene terephthalate



wine, pet, bottle

s froehlich1121



cat, kittens, eyes,  
ears, pet, animal

shorty\_nz\_2000



dog, pet, animal,  
funny, glasses

Edward Corpuz



dog, pet, alaskan malamute

s froehlich1121



cat, kitty, eyes,  
pretty

Jay Woodworth



cat, kitten, garden,  
pet

Nemo's great uncle



recycle, pet, plastic bottle,  
polyethylene terephthalate

Karl Baron



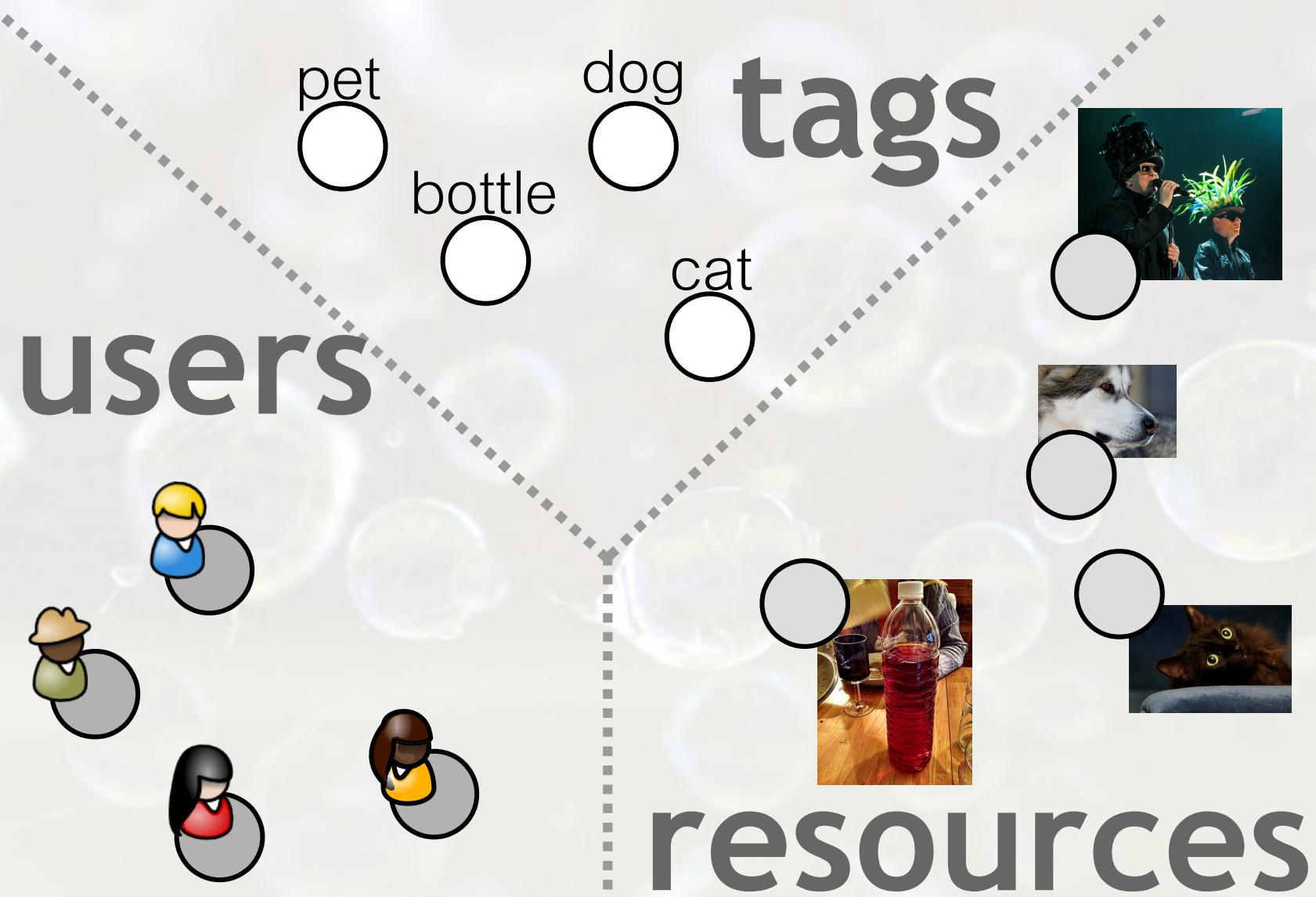
wine, pet, bottle

FaceMePLS

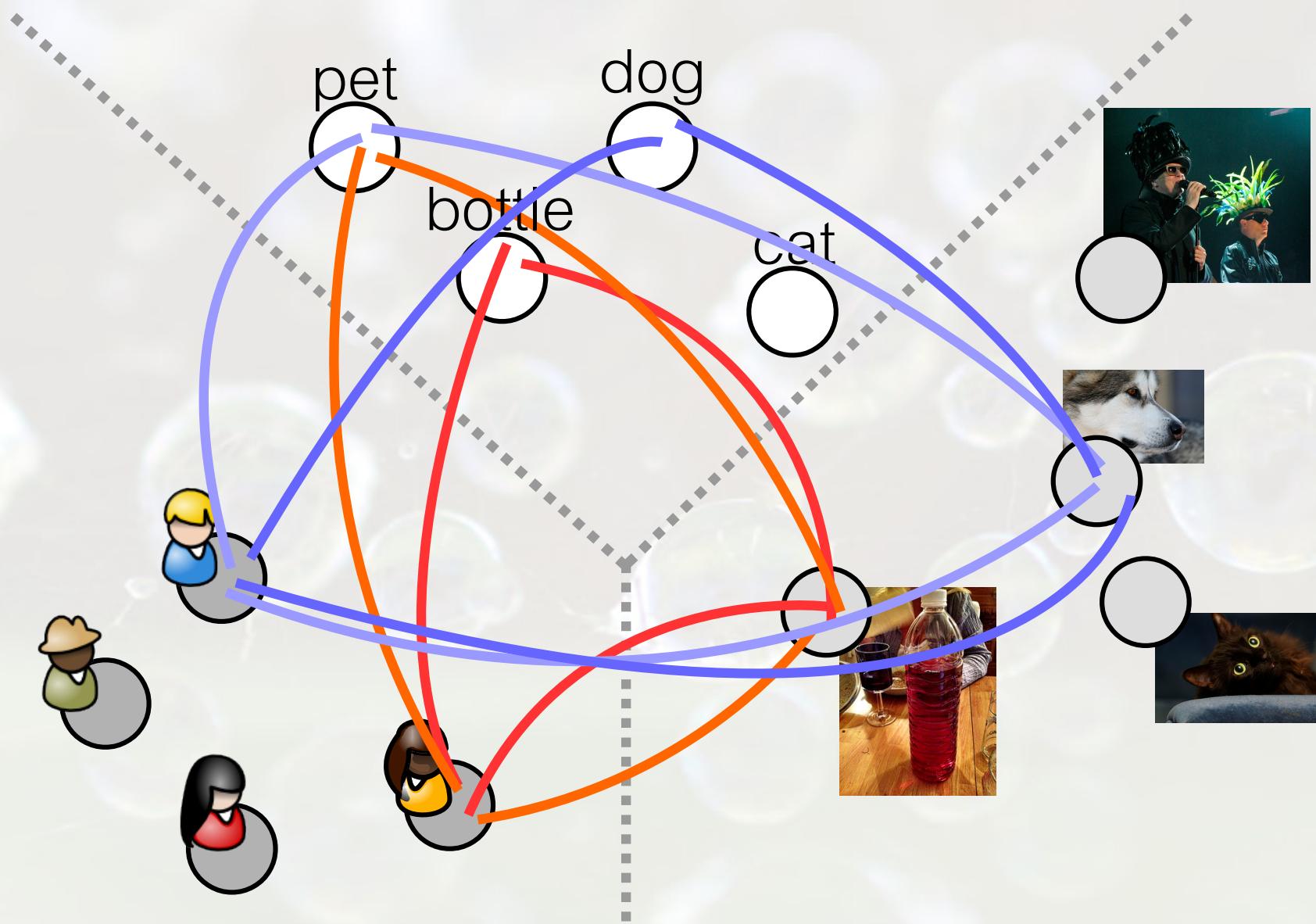


recycle, pet, plastic bottle,  
polyethylene terephthalate

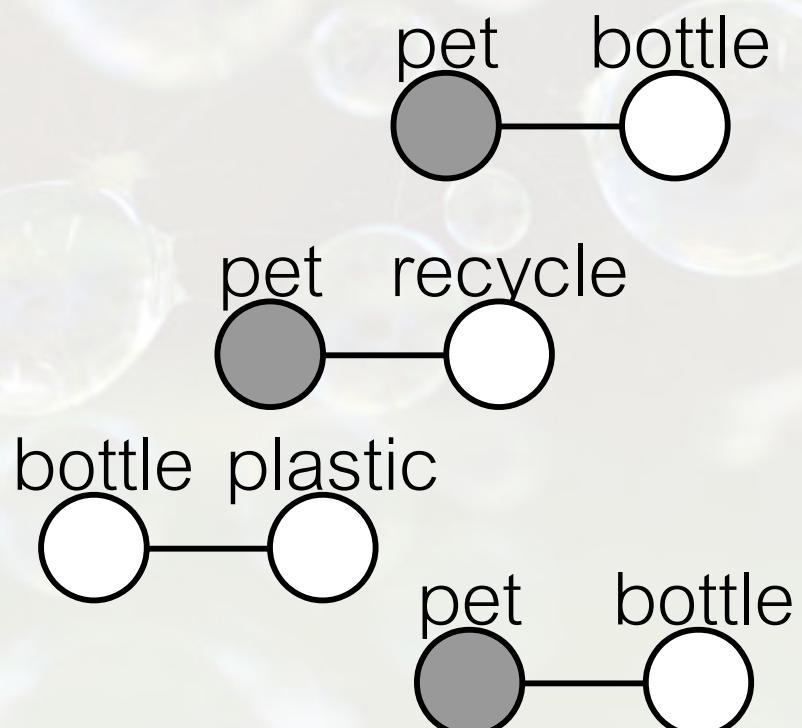
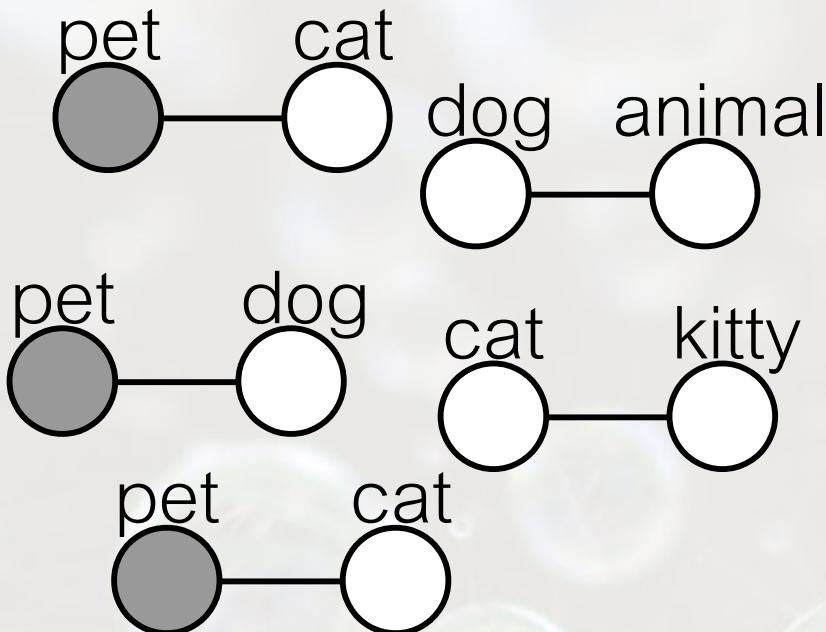
# Graph users/tags/resources



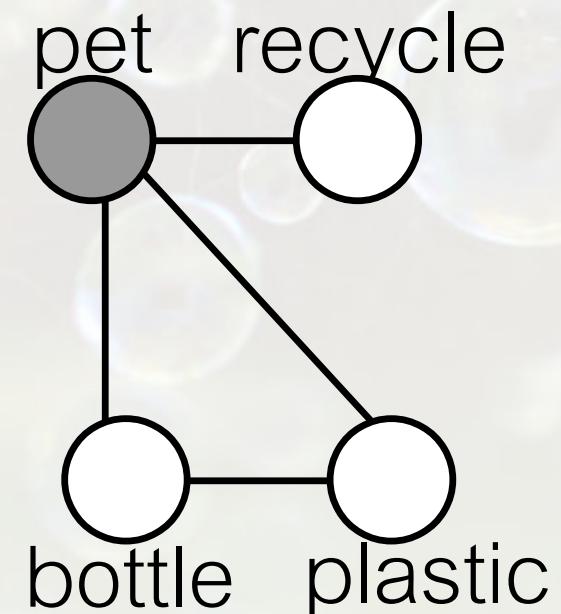
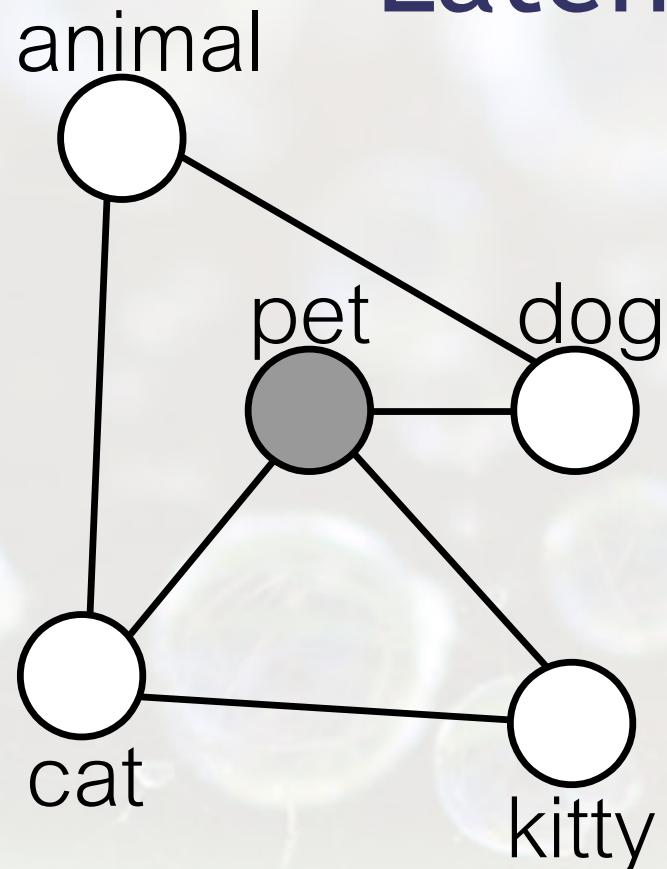
# Graph users/tags/resources



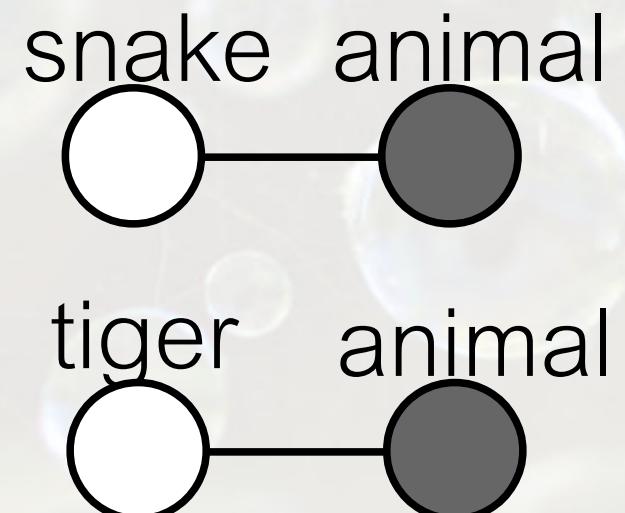
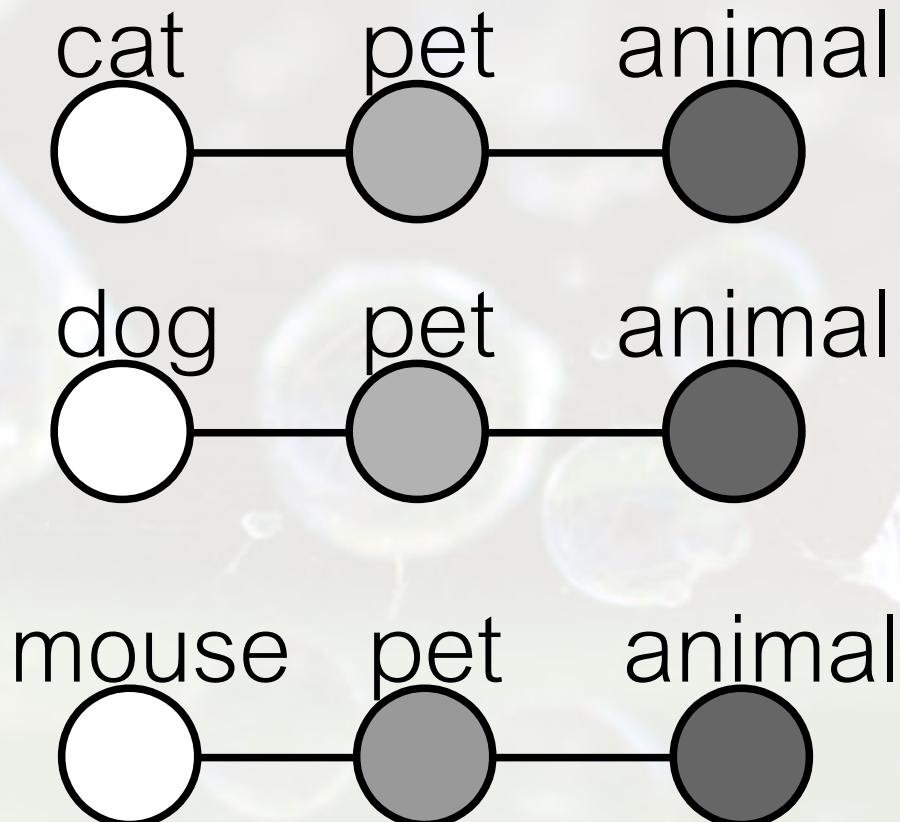
# Co-occurrences and Latent Semantics



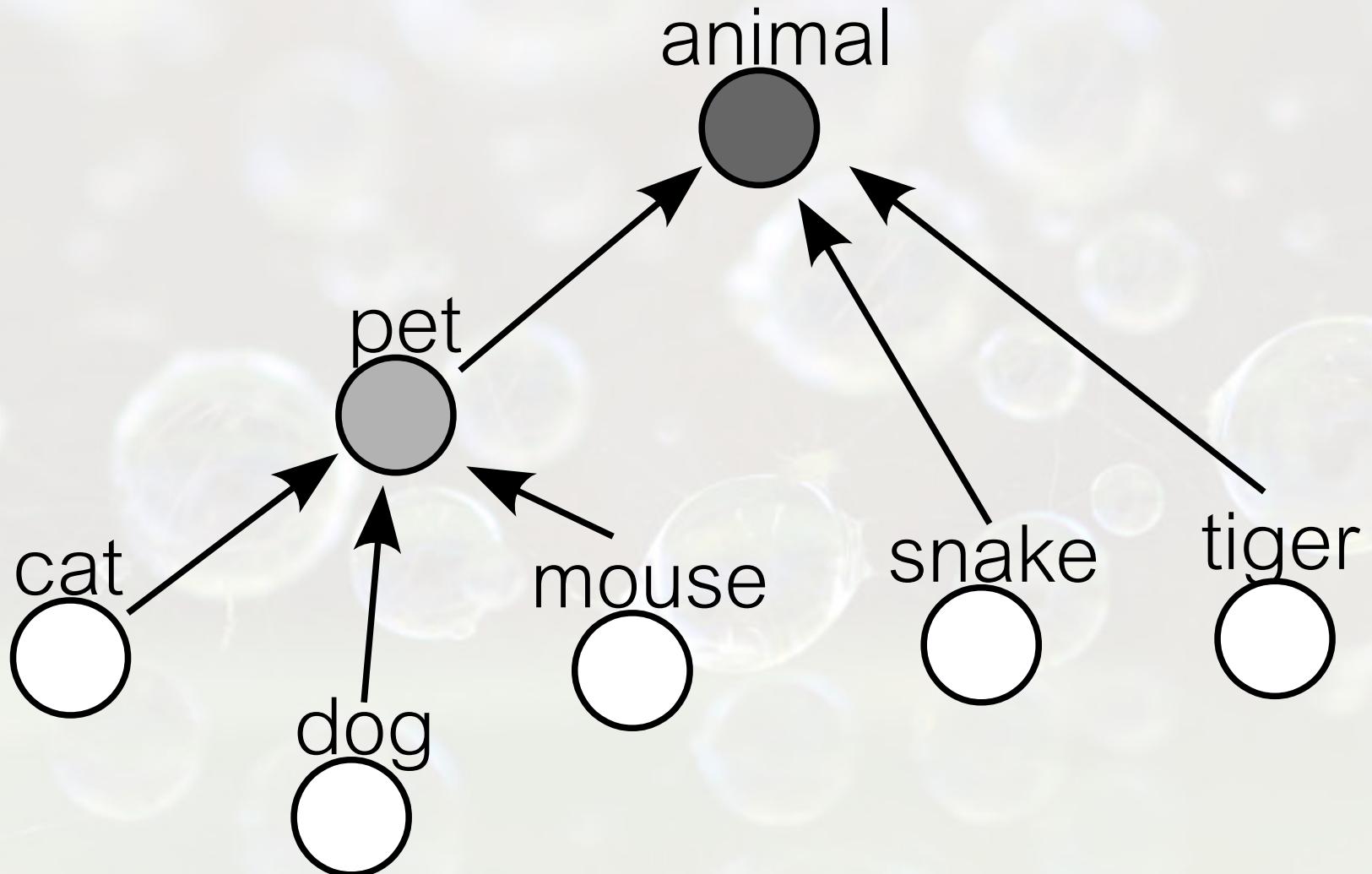
# Co-occurrences and Latent Semantics



# Co-occurrences and Latent Semantics



# Co-occurrences and Latent Semantics



# Social Effect Suggested/Reinforced Tags

sfroehlich1121



black, cat, kitty, katze, long, hair, blue, eyes, pretty,  
canon, tli, 500d, ef 100mm f/2.8 usm macro

# Social Effect Suggested/Reinforced Tags

sfroehlich1121



black, cat, kitty, katze, long,

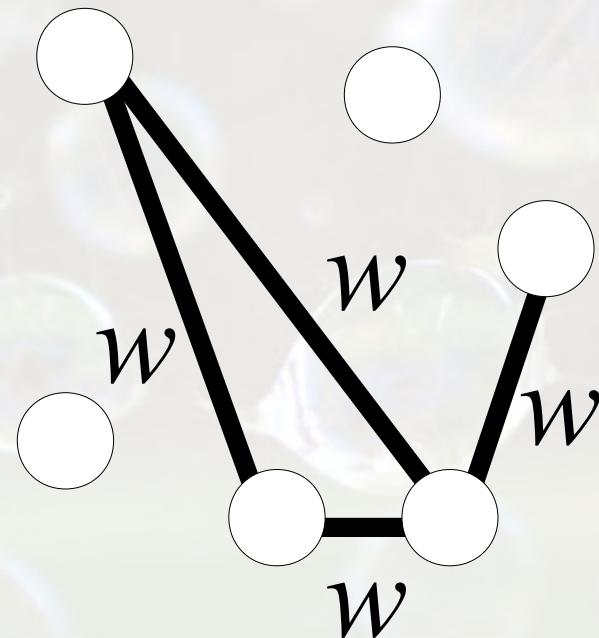
hair, blue, eyes, pretty,

canon, tli, 500d, ef 100mm f/2.8 usm macro

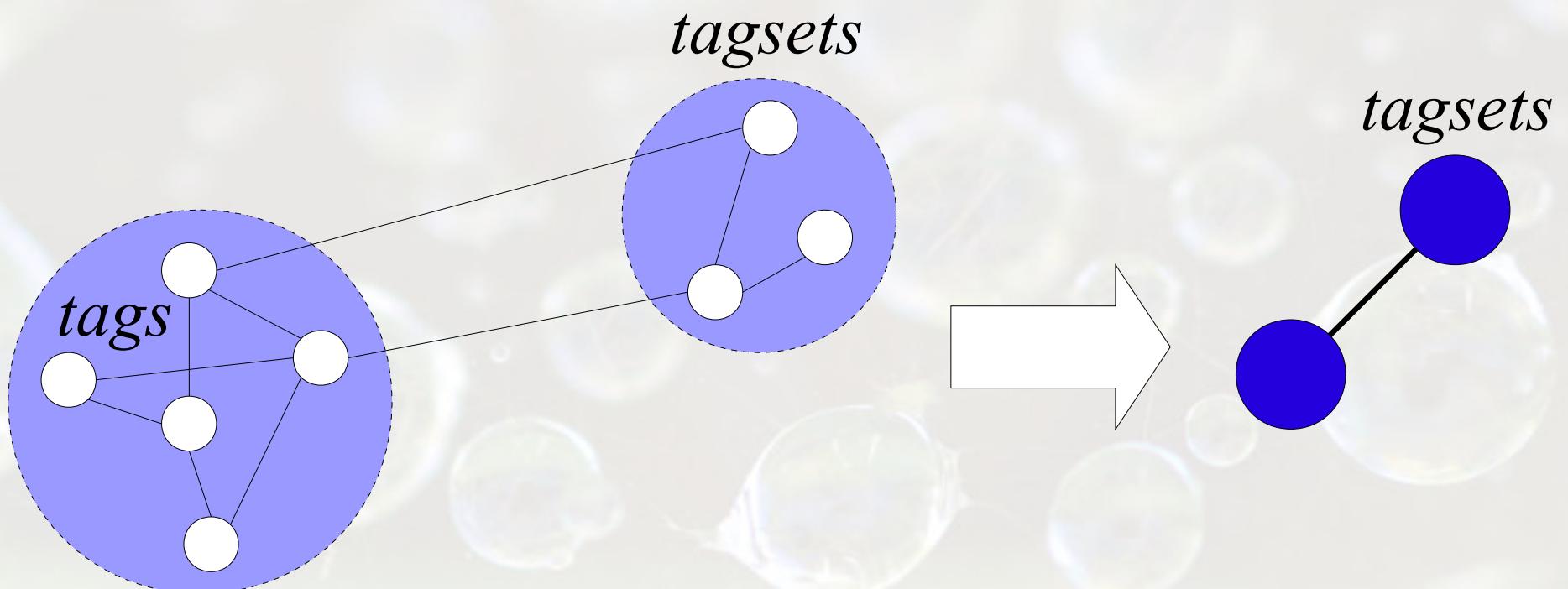
M1

# co-occurrences in a folksonomy

*tags / co-occurrences*



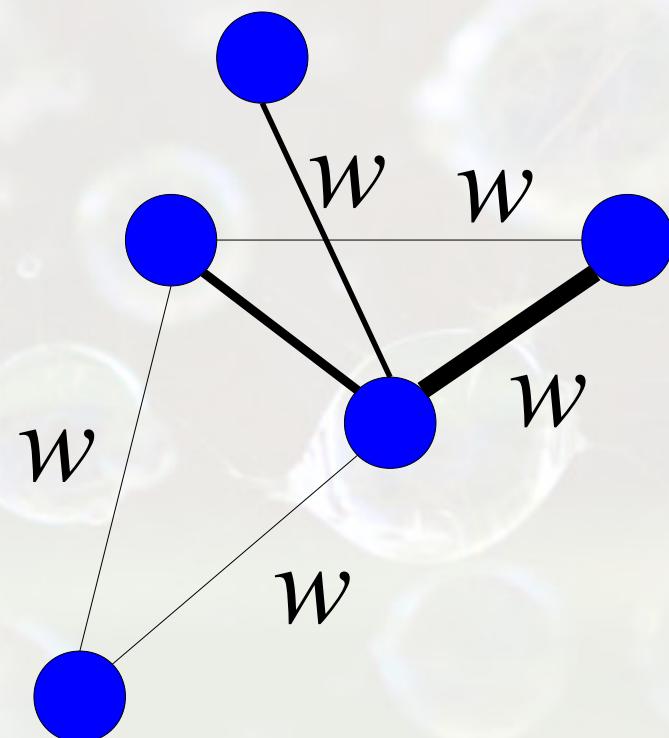
# Tagsets



# M2

## tagsets and their co-occurrences

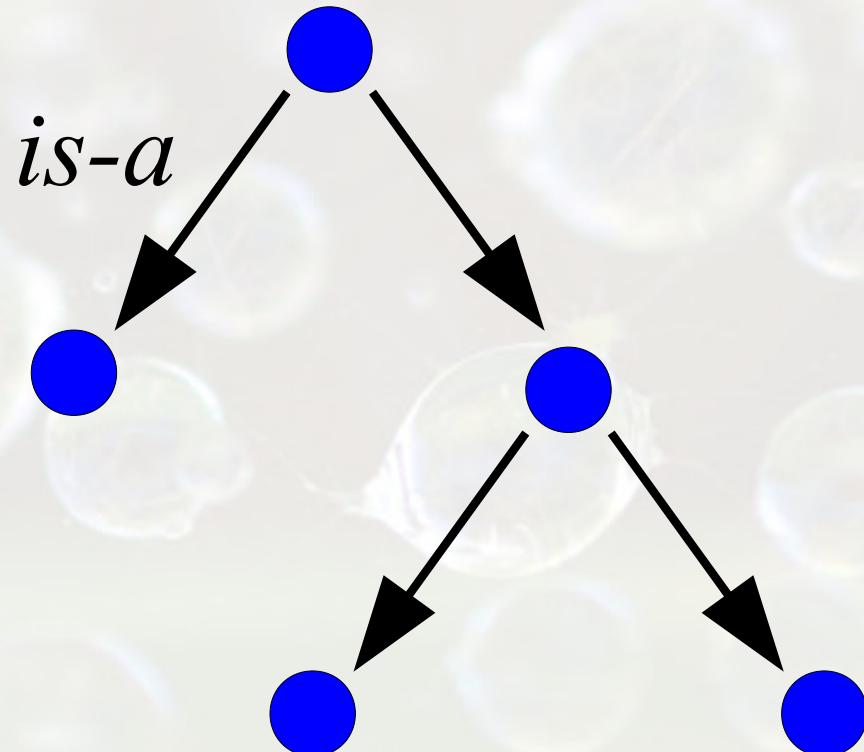
*tagsets/co-occurrences*



# M4

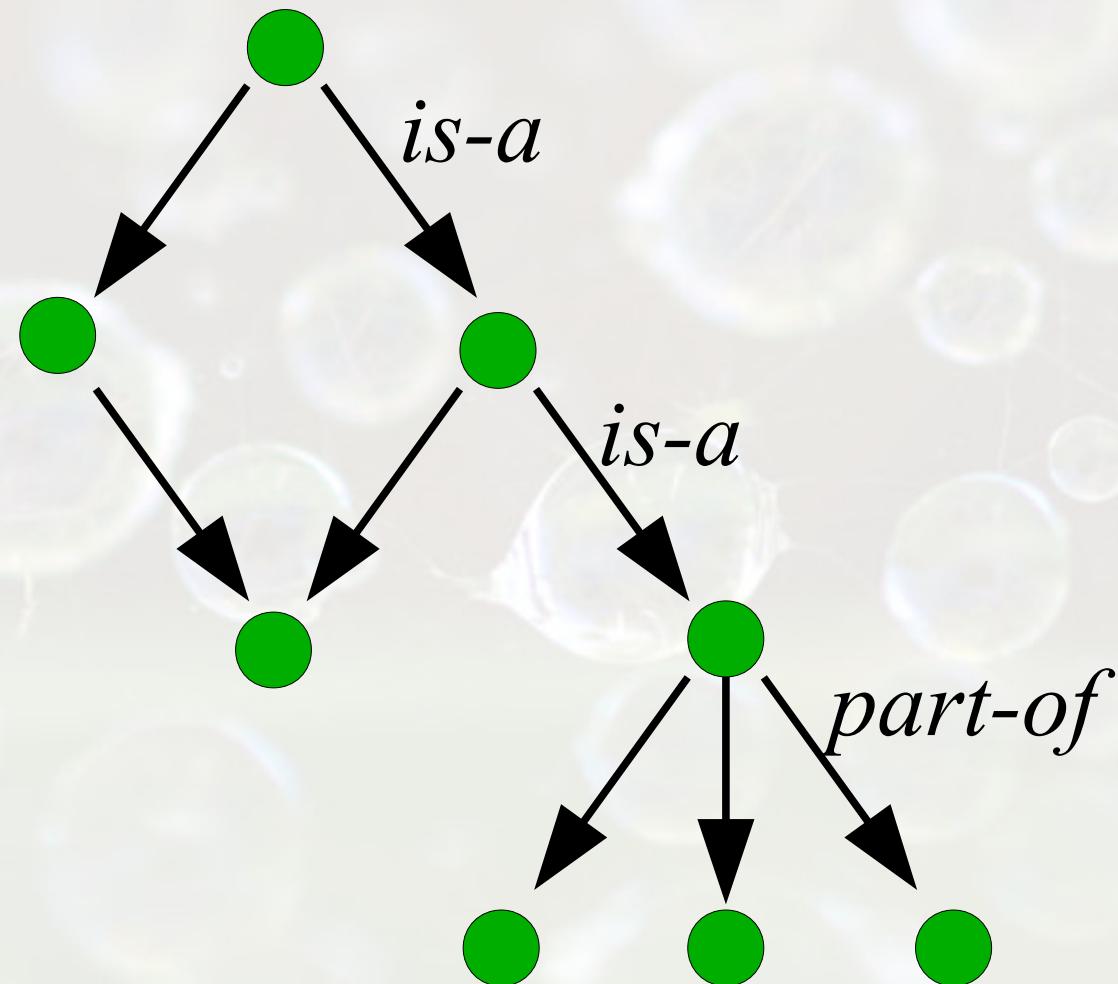
## Social Ontology

*tagsets / typed relations*



# M3 simplified model of an ontology

*concepts/relations*



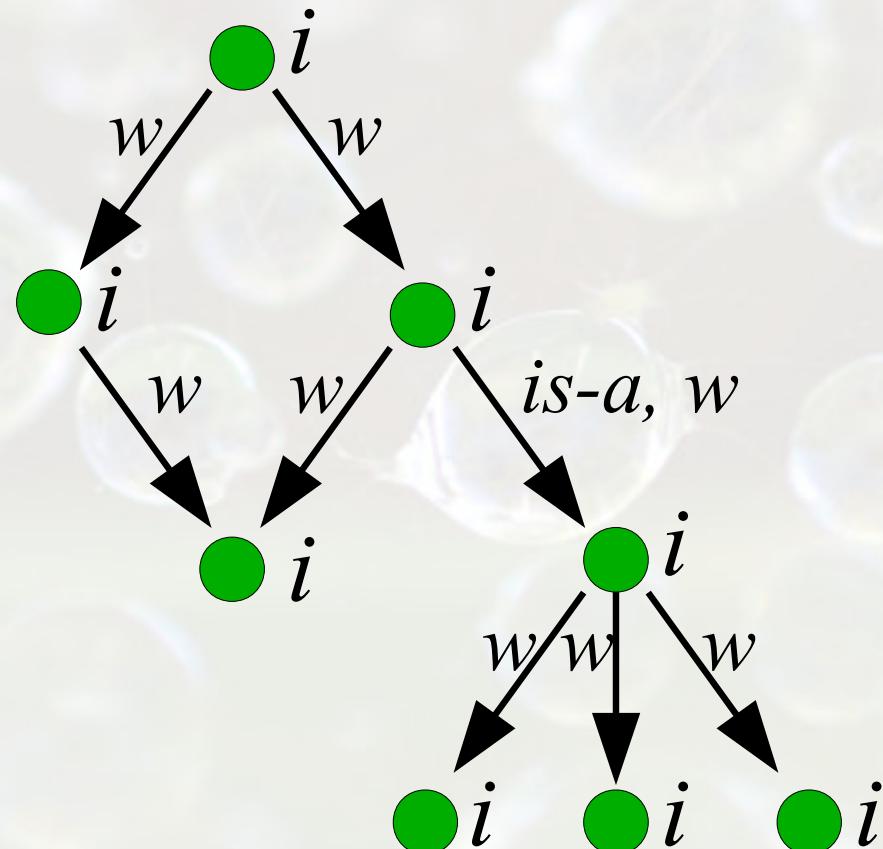
# M5

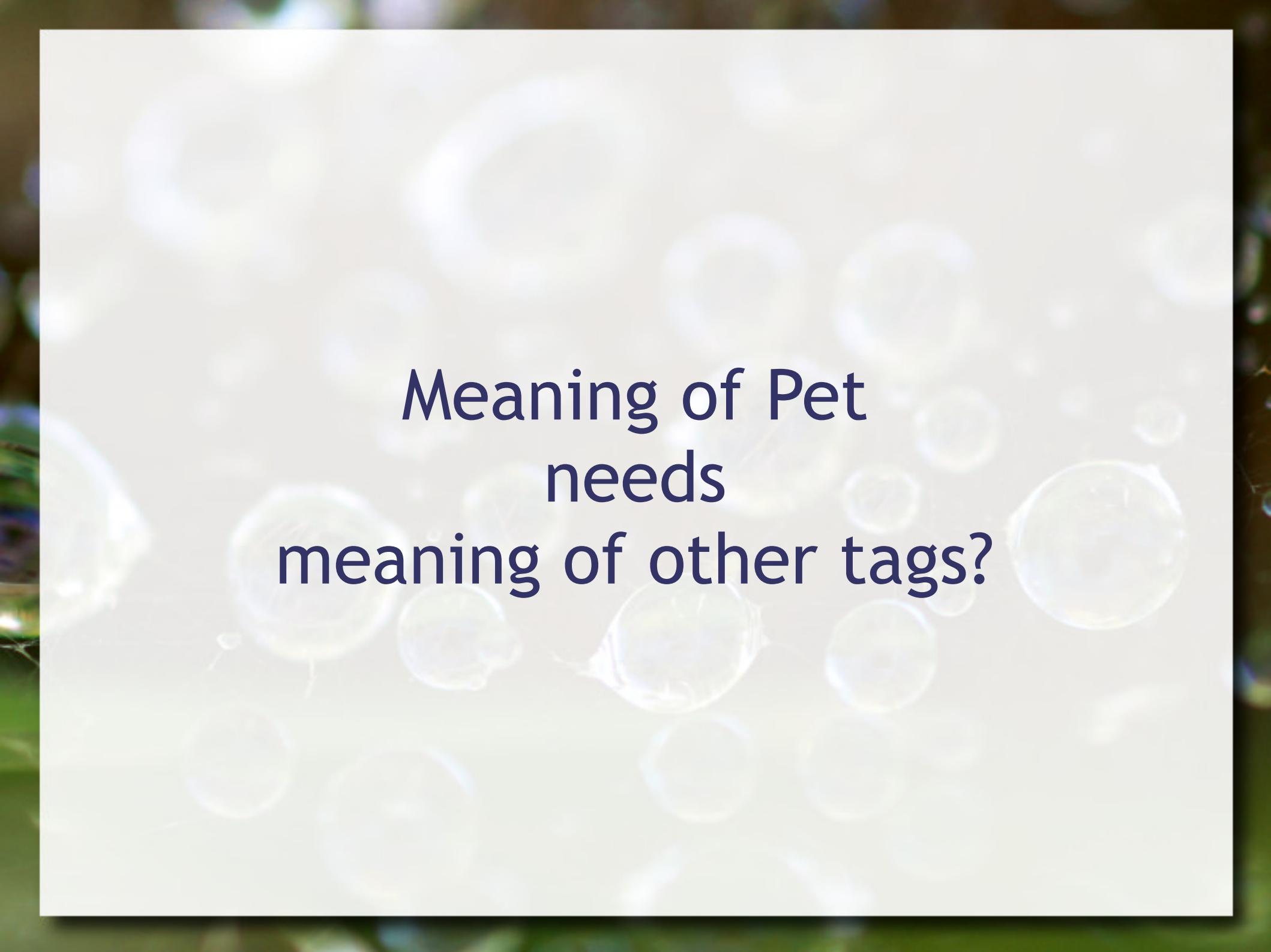
## Folksonomized Ontology (FO)

*concepts / typed relations (ontology)*

+

*information content / co-occurrences (folksonomy)*





Meaning of Pet  
needs  
meaning of other tags?

# Matrix

What is the meaning of the word  
Love?

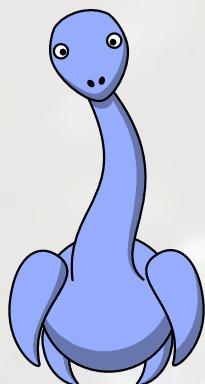


Why Graphs?

# Sharing and Interconnecting

# Models to Describe

# Describing Prehistoric Animals



**MNHN A. C. 8592**

<b>Is a</b>	Plesiosaurus dolichodeirus
<b>Origin</b>	Lyme Regis
<b>Recognized</b>	1824
<b>Size</b>	5

# Describing Prehistoric Animals



SIPB R 90

Is a	Plesiosaurus dolichodeirus	
Origin	Lyme Regis	England
Recognized	1830	
Size	5	



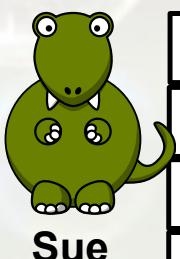
STC223

Is a	Plesiosaurus gurgitis	
Origin	St. Croix	Switzerland
Recognized	1964	
Size	3.5	



MNHN 1912.20

Is a	Triceratops horridus	
Origin	Lance Creek	EUA
Recognized	1889	
Size	9	



Sue

FMNH PR2081

Is a	Tyrannosaurus rex	
Origin	Hell Creek	EUA
Recognized	1990	
Size	12.3	

# Table

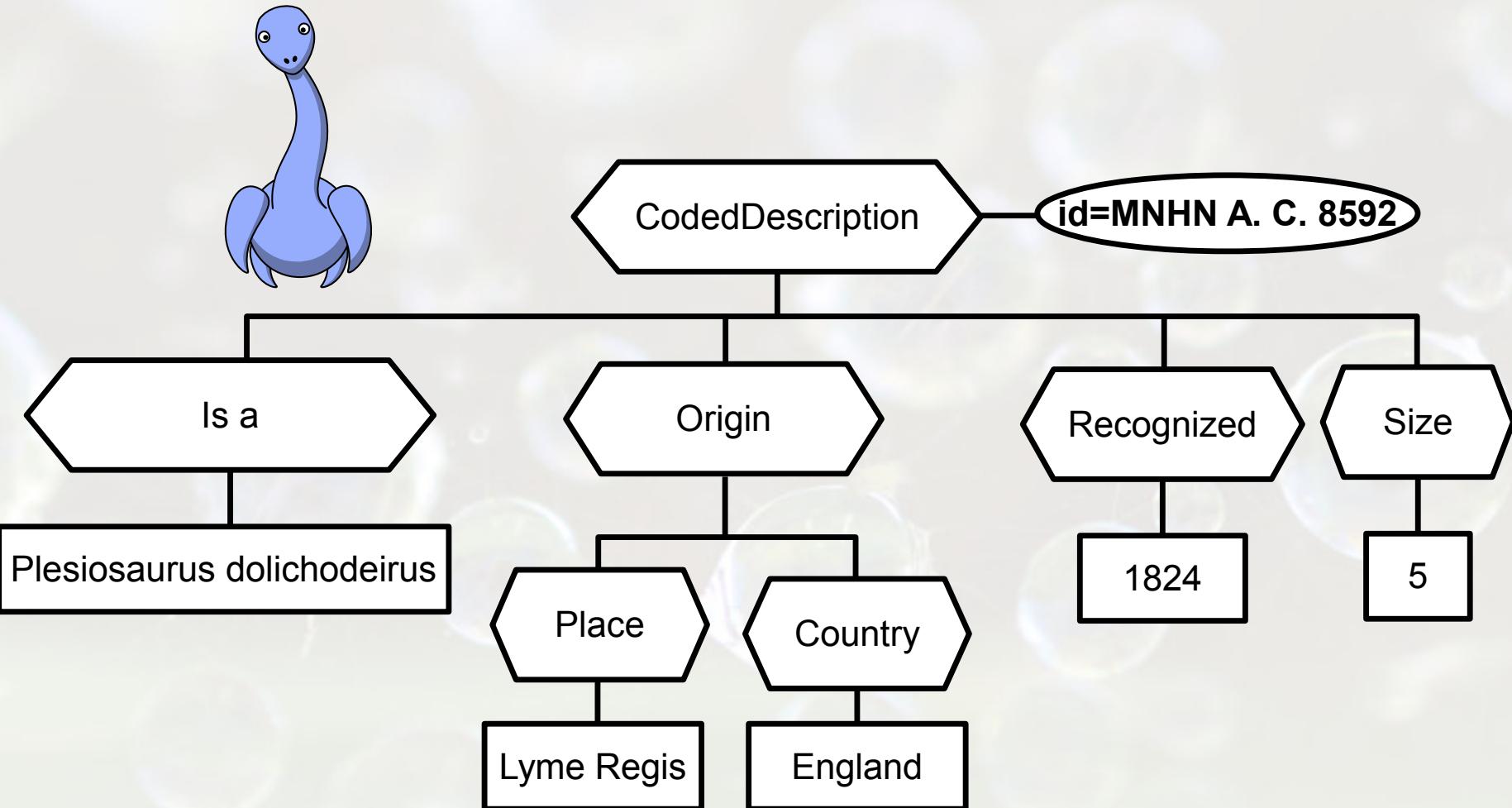
Id	Is a	Origin Place	Origin Country	Recognized	Size
 MHN A. C. 8592	Plesiosaurus dolichodeirus	Lyme Regis	England	1824	5
 SIPB R 90	Plesiosaurus dolichodeirus	Lyme Regis	England	1830	5
 STC223	Plesiosaurus gurgitis	St. Croix	Switzerland	1964	3.5
 MHN 1912.20	Triceratops horridus	Lance Creek	EUA	1889	9
 FMNH PR2081	Tyrannosaurus rex	Hell Creek	EUA	1990	12.3

# Table

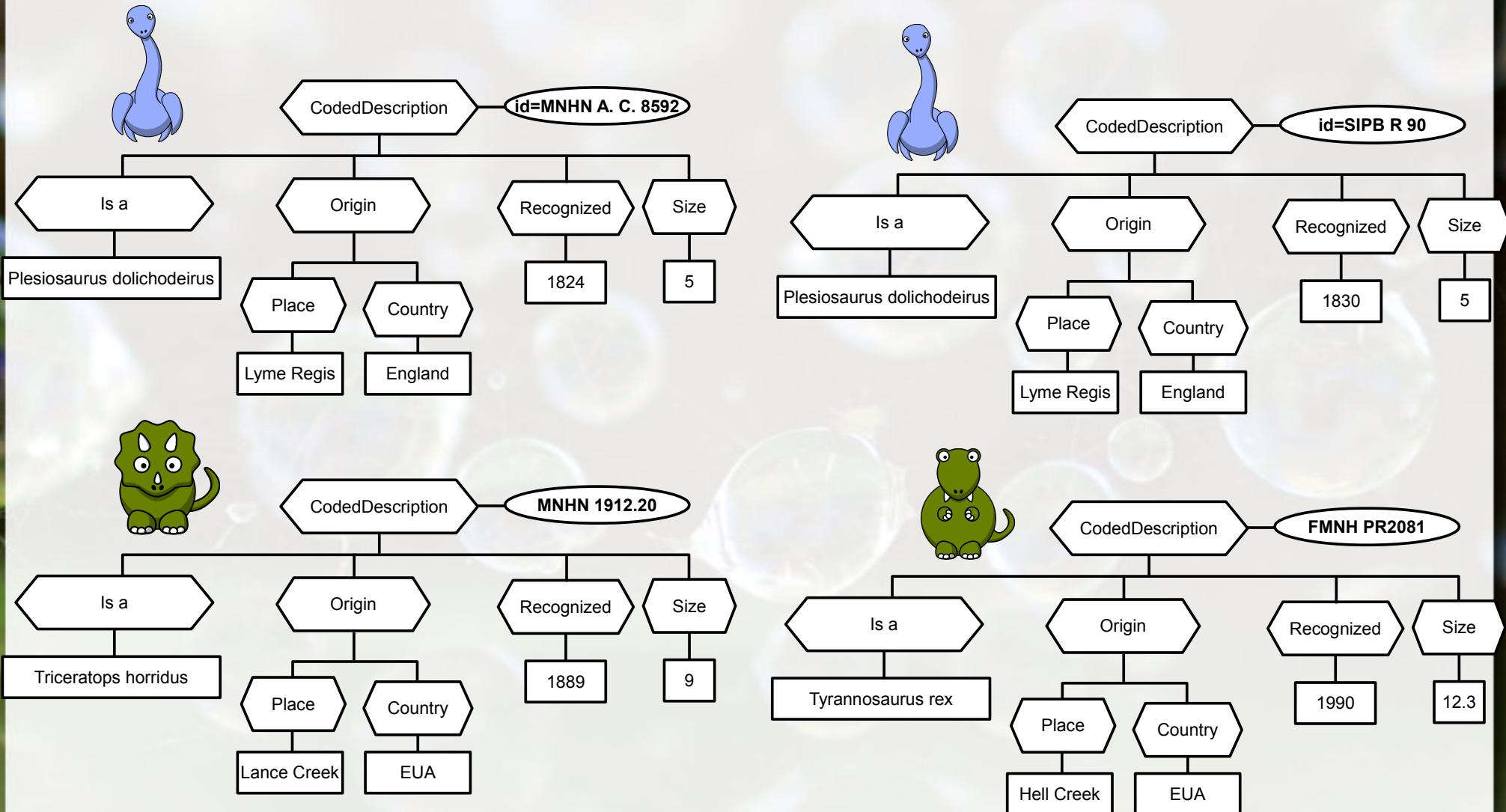
Excellent to Manage Data with  
Predictable Static Schema

Sharing?

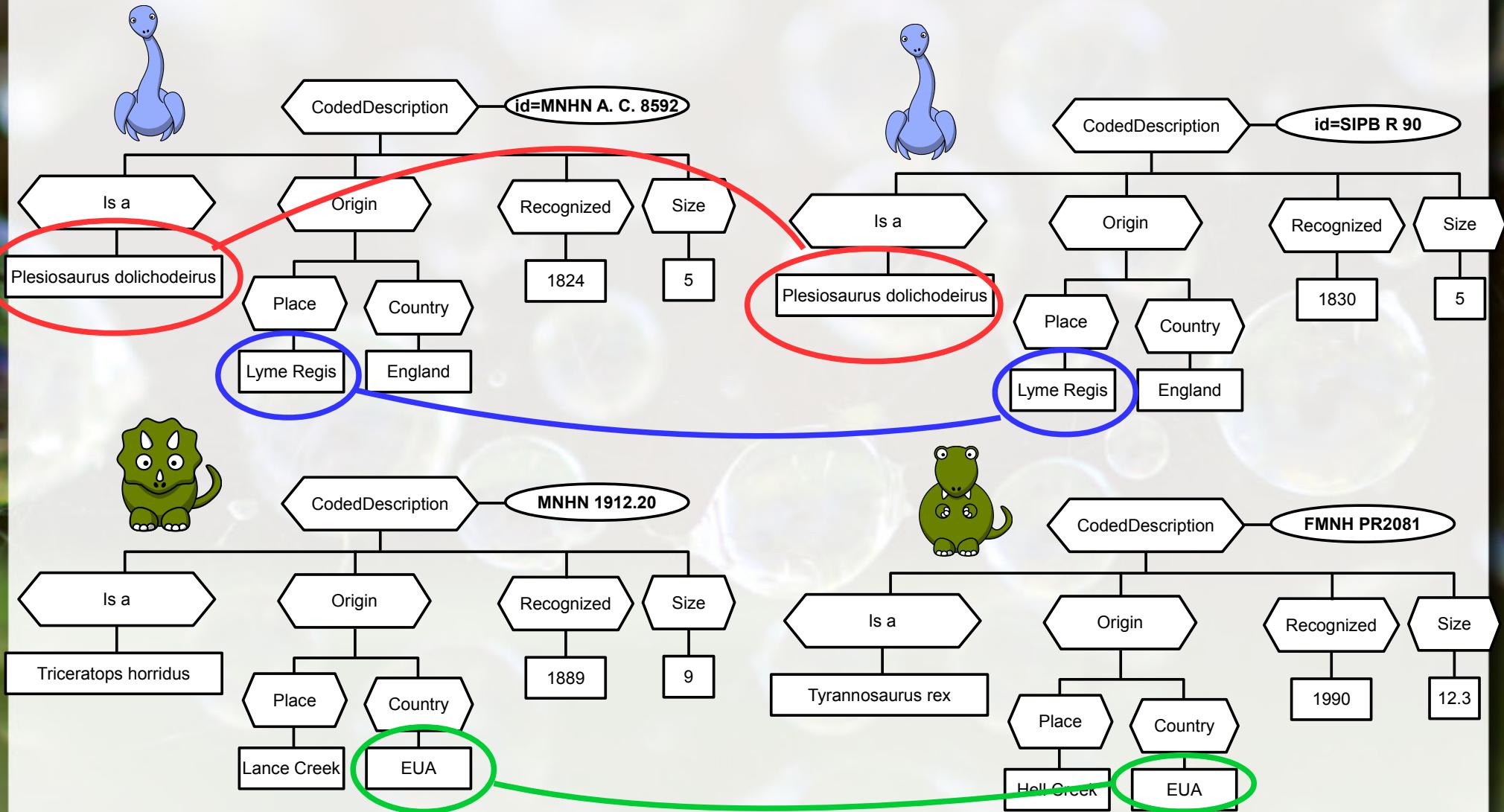
# Documents and XML

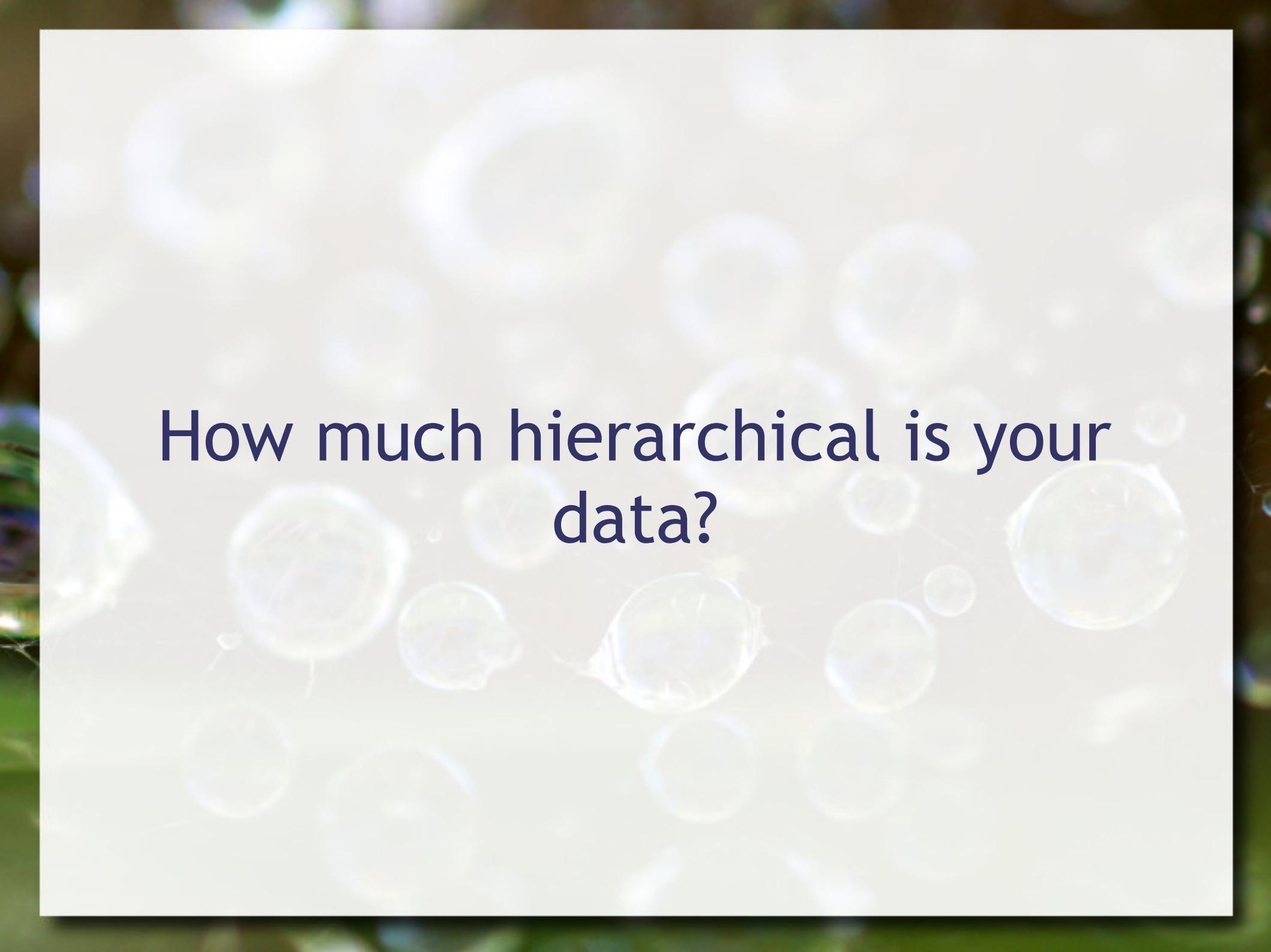


# Documents and XML



# Documents and XML





How much hierarchical is your  
data?

# Back to the Table

Le Muséum national d'Histoire naturelle

includes

Id	Is a	Origin Place	Origin Country	Recognized	Size
 MHN A. C. 8592	Plesiosaurus dolichodeirus	Lyme Regis	England	1824	5
 SIPB R 90	Plesiosaurus dolichodeirus	Lyme Regis	England	1830	5
 STC223	Plesiosaurus gurgitis	St. Croix	Switzerland	1964	3.5
 MHN 1912.20	Triceratops calicornis	Lance Creek	EUA	1888	9
 MHN 1912.20b	Triceratops horridus	Lance Creek	EUA	1889	9
 FMNH PR2081	Tyrannosaurus rex	Hell Creek	EUA	1990	12.3

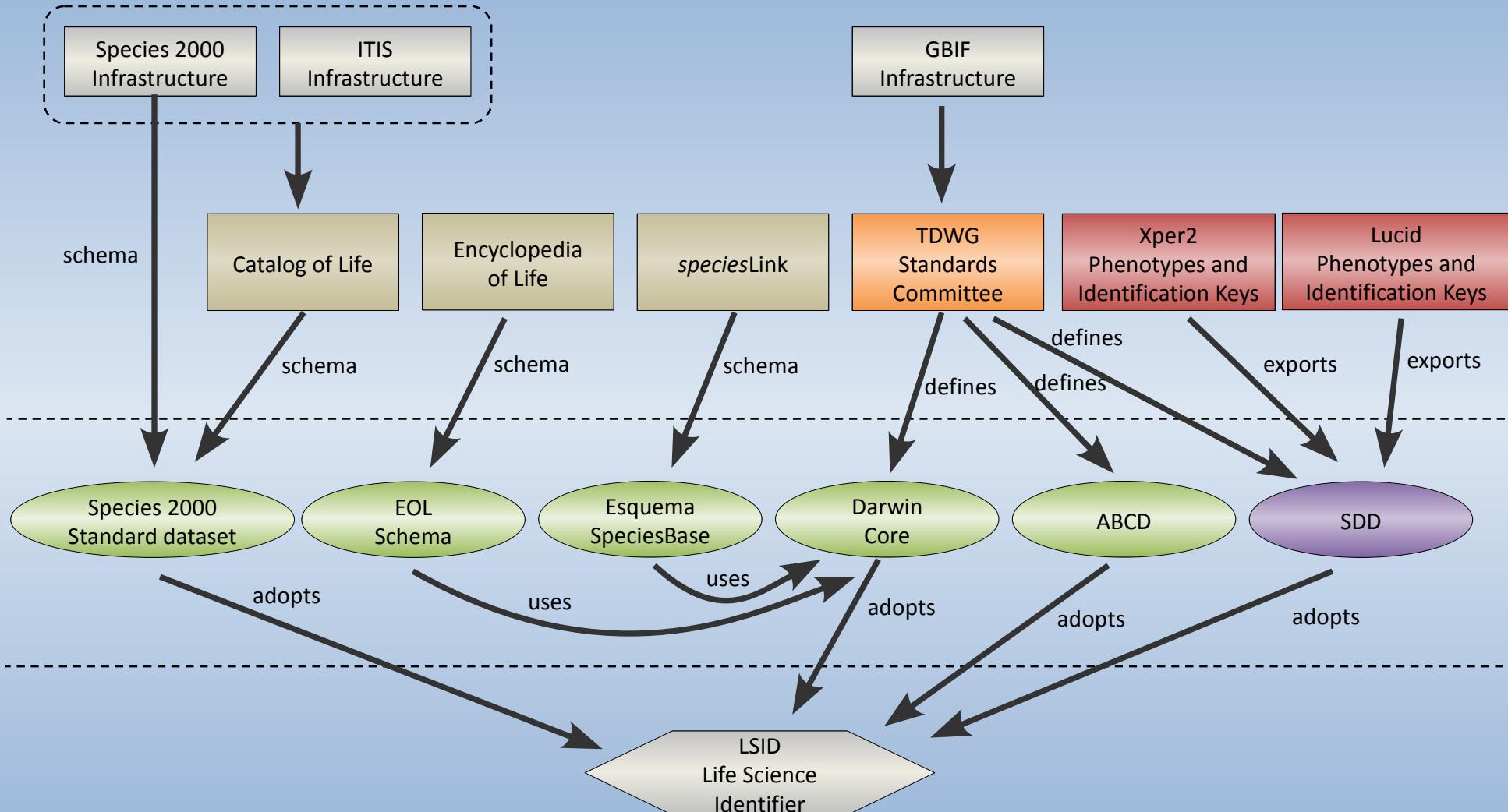
renamed

near

# The tradeoff of Static Schemas

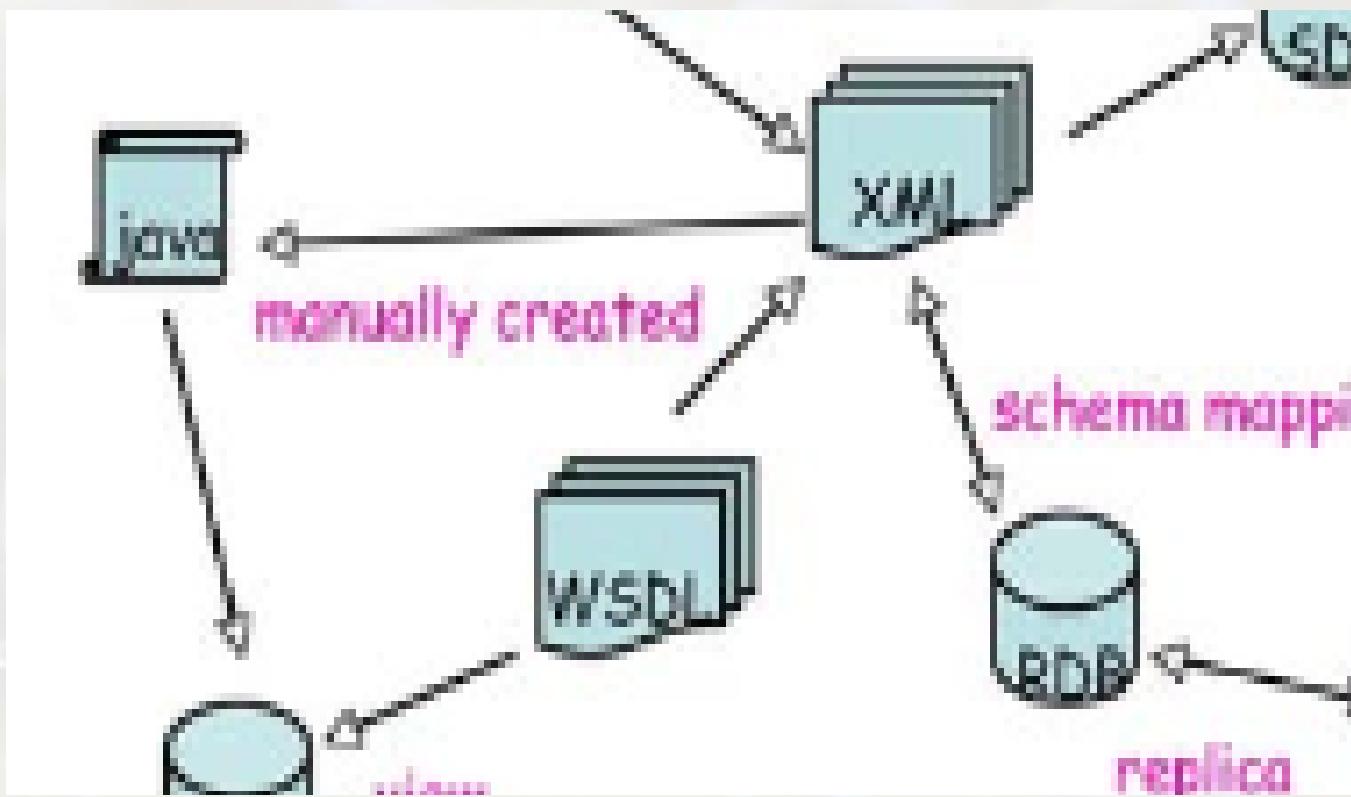
# XML Schema

## Not designed to mix



# DataSpaces

## Pay-as-you-go Integration



Franklin, M., Halevy, A., & Maier, D. (2005). From databases to dataspaces: a new abstraction for information management. SIGMOD Rec., 34(4), 27–33.

# DataSpaces and Linked Data



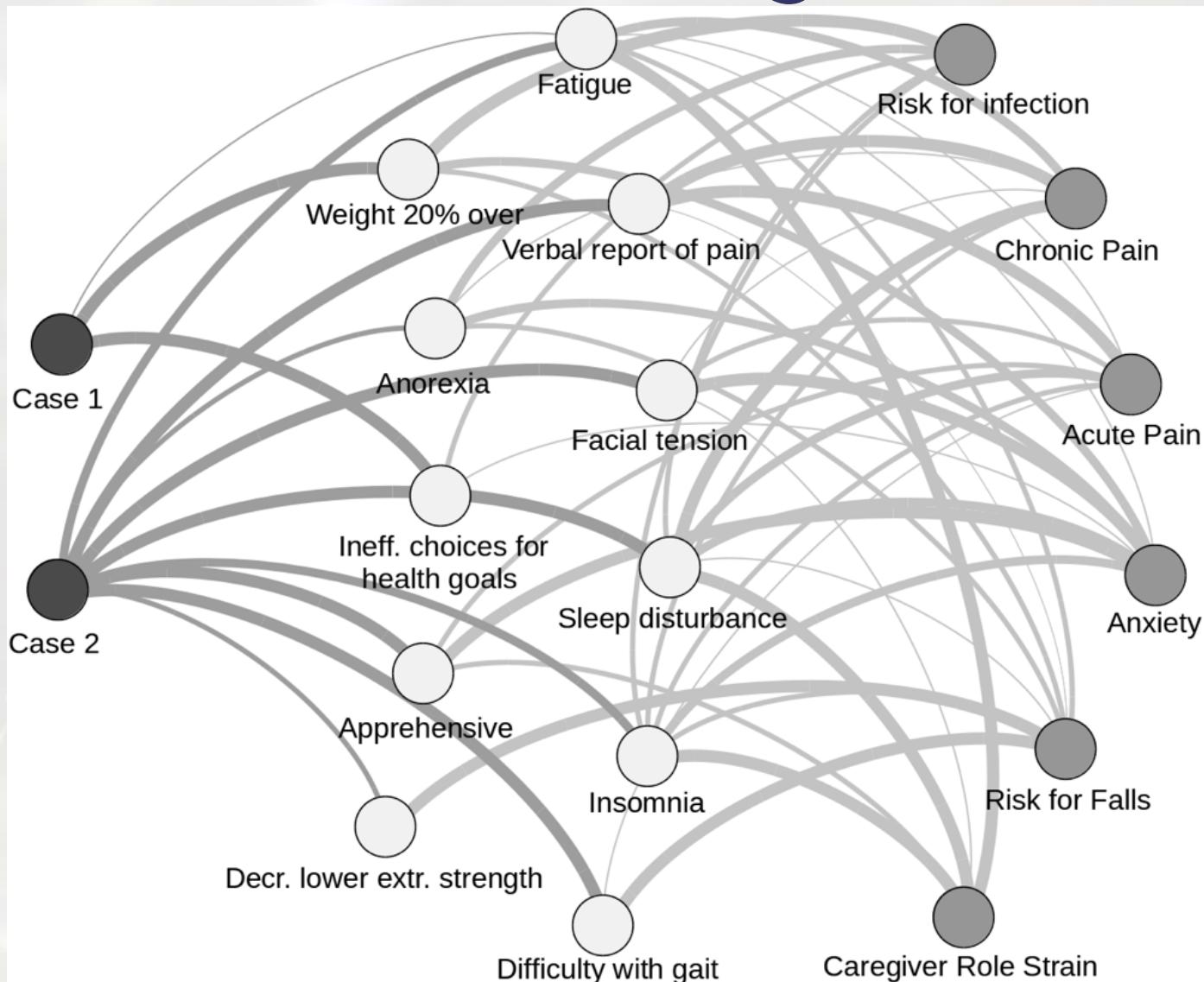
# Why Graphs? Complex Networks

# Complex Networks

- Developed steadily since 1999
- Discrete systems are represented in terms of entities and relationships

(Luciano da F. Costa, 2013)

# Patient x Diagnosis



(Gomes-Jr, 2013)

# Complex Networks

- Physical relationships
  - neurons - nodes; connections - edges
- Force relationships
  - grains - nodes; force vectors - edges
- Social relationships
- Conceptual relationships

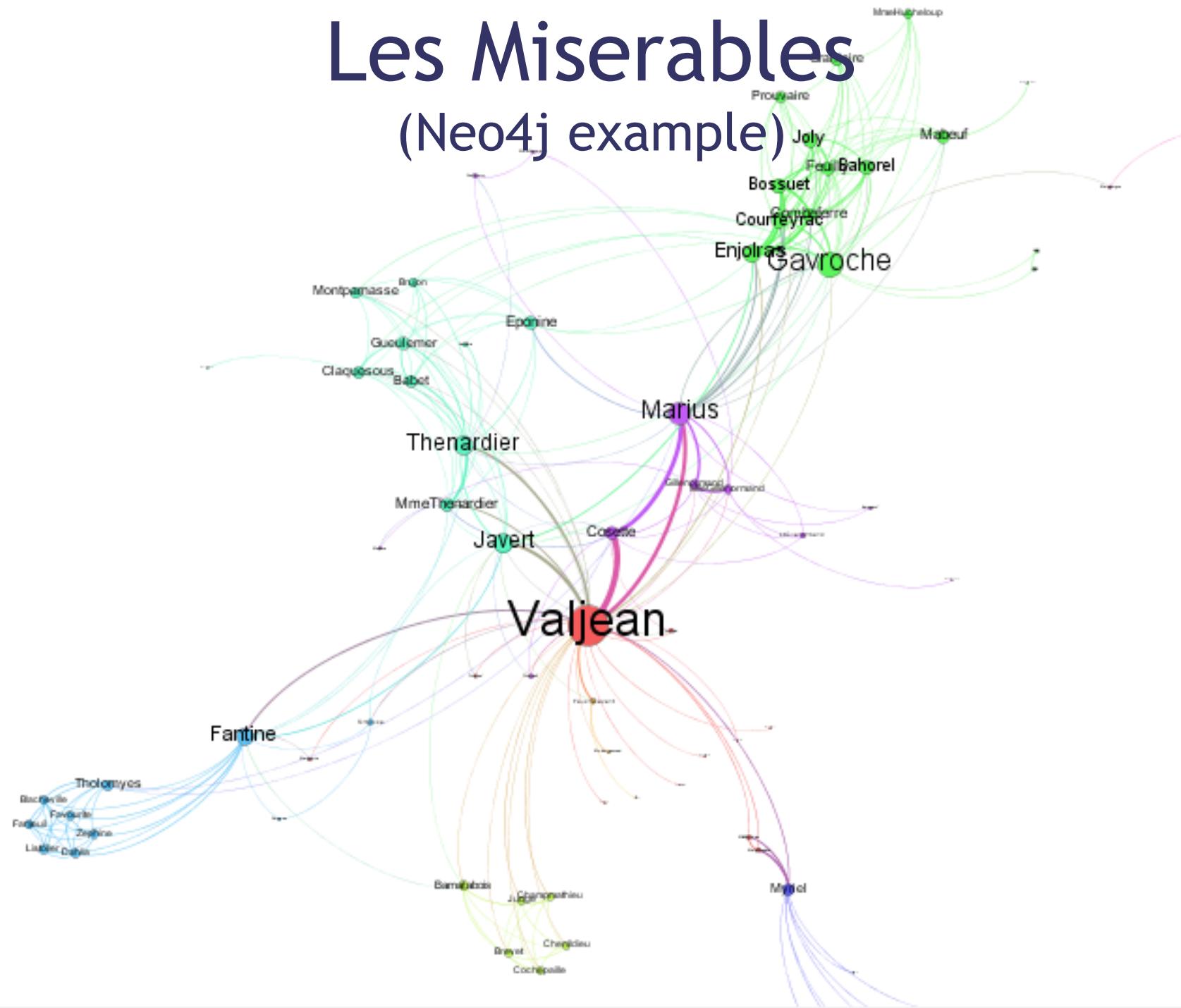
# Complex Networks Examples

<http://www-personal.umich.edu/~mejn/networks/>

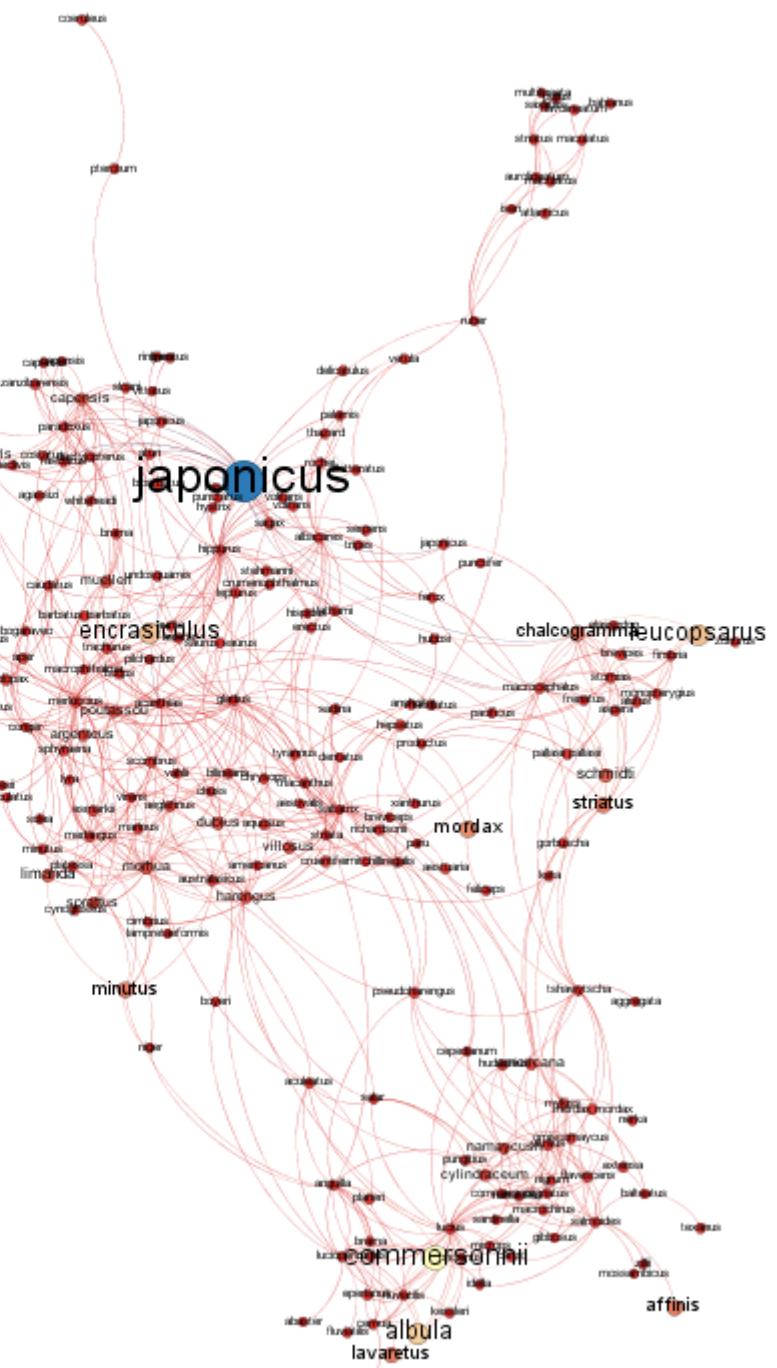
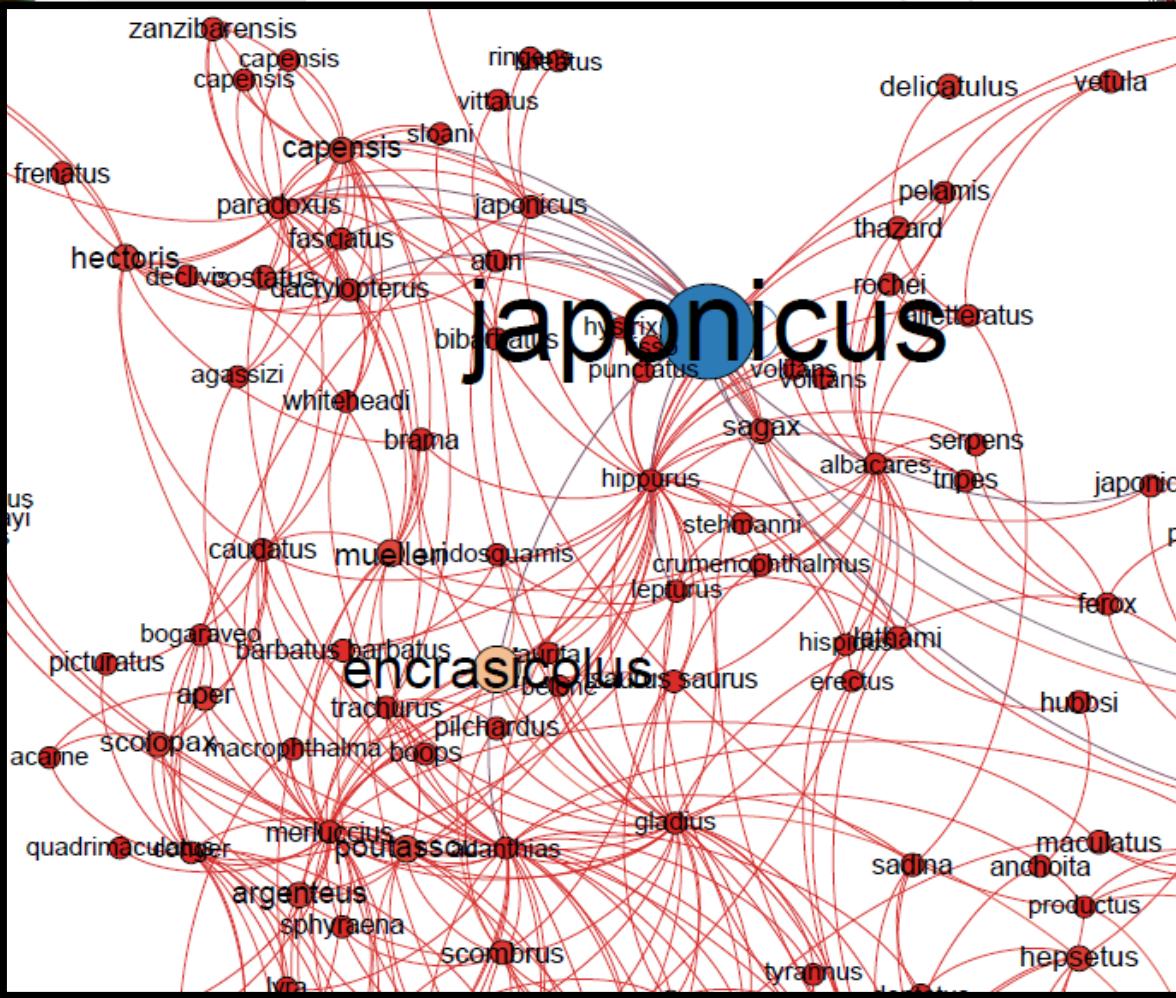
# Les Miserables

Proudhon Joly

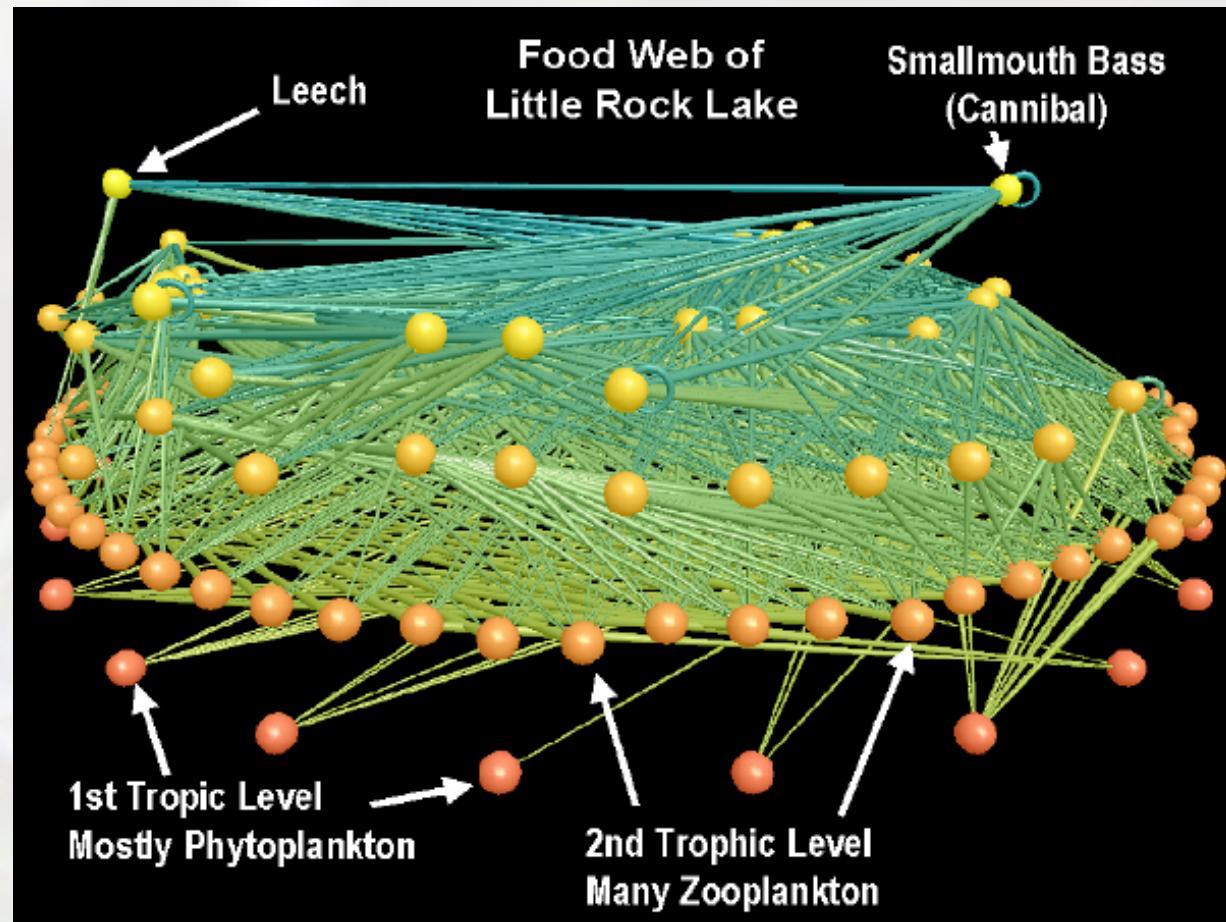
## (Neo4j example)



# Cadeia Alimentar FishBase

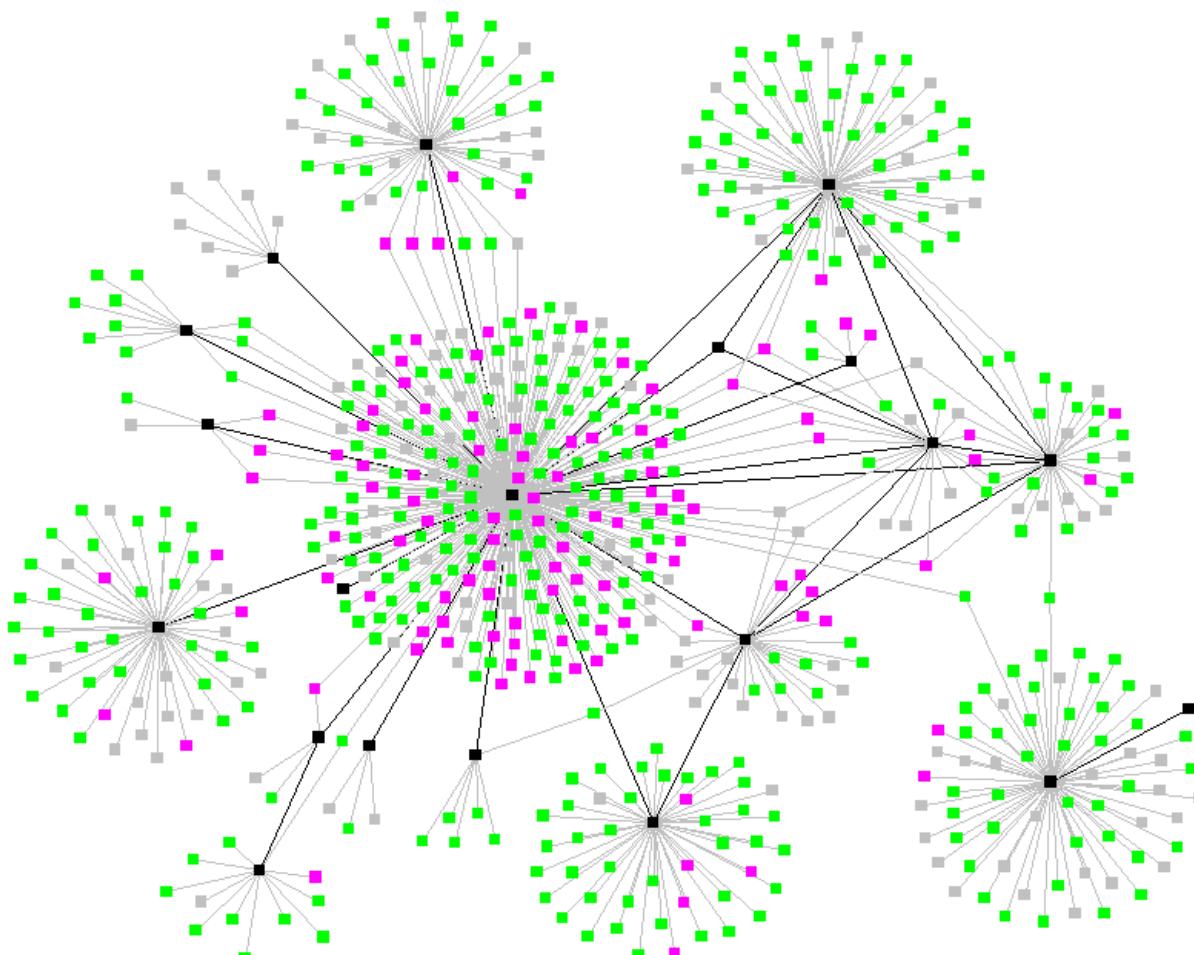


# Freshwater food web



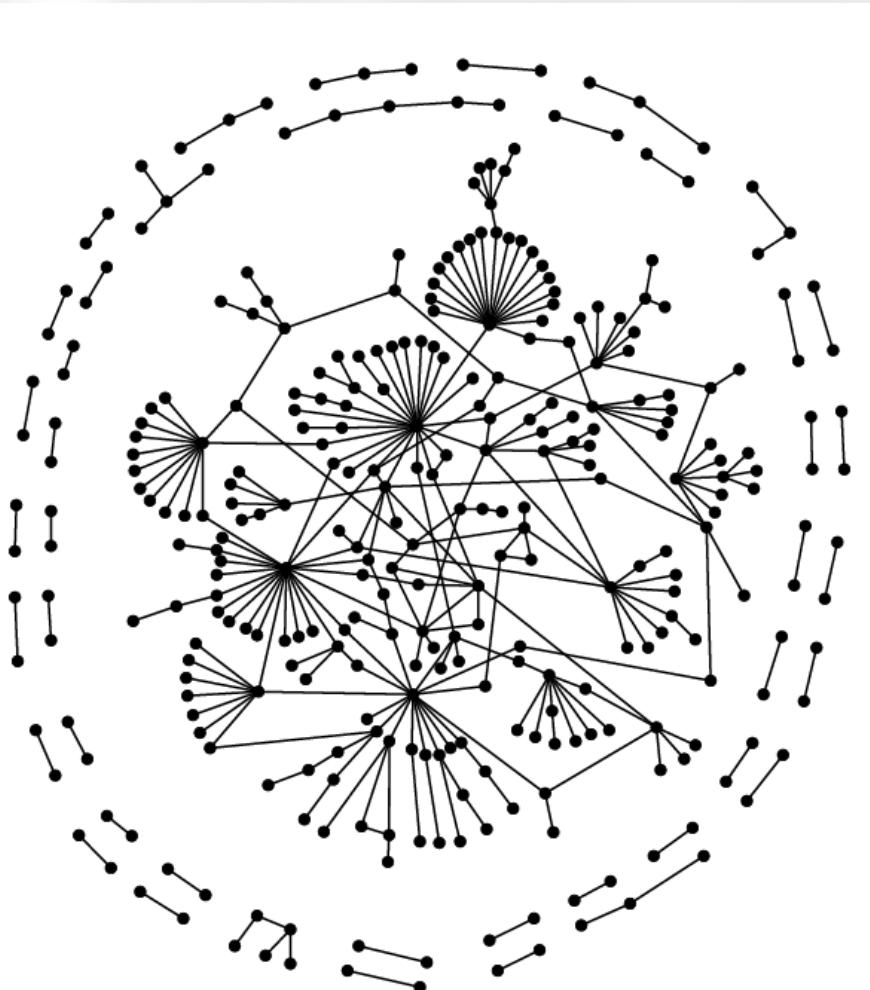
Freshwater food web: Neo Martinez and Richard Williams.

# Contagion of TB

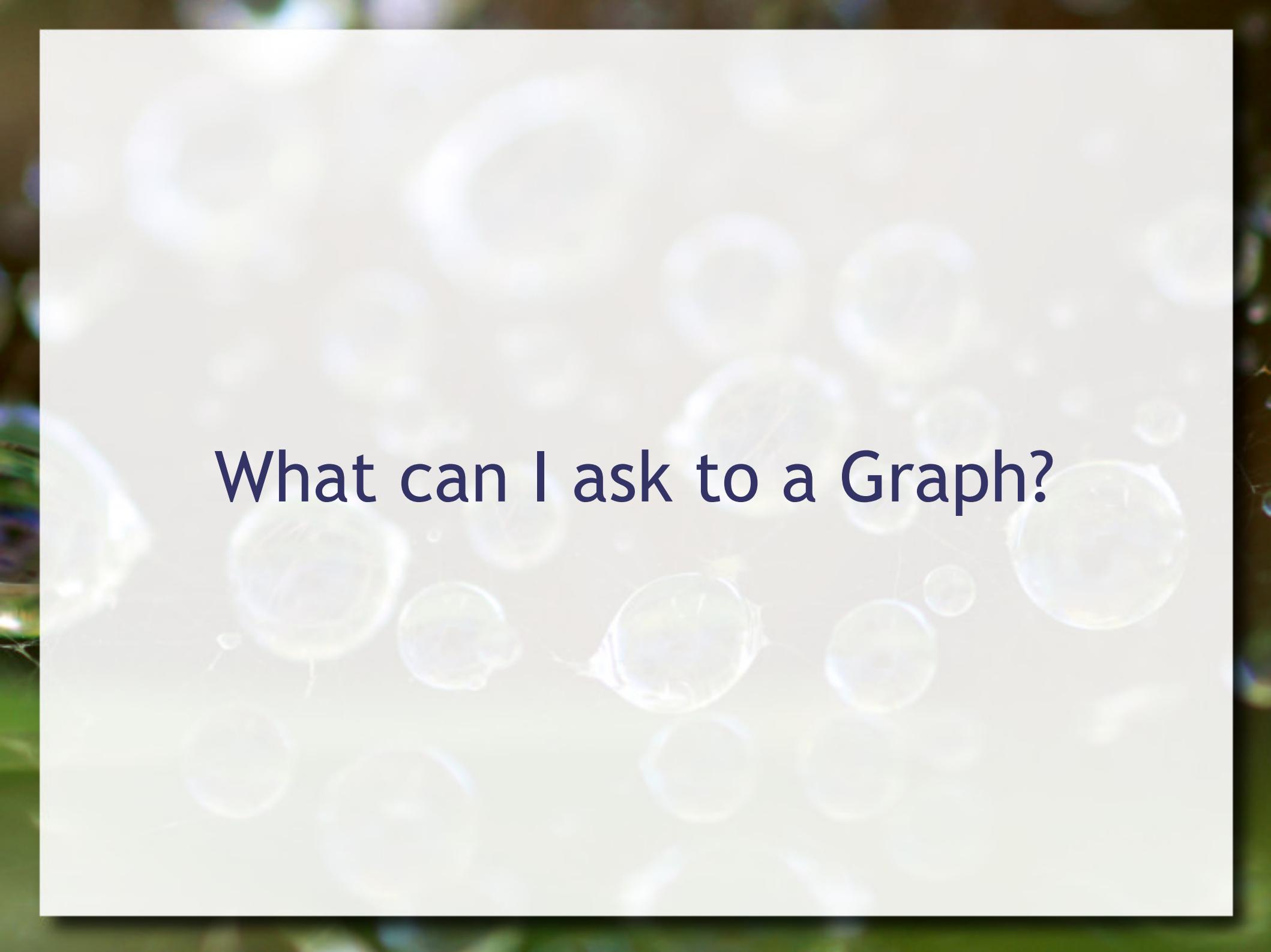


Contagion of TB, books on politics: Valdis Krebs, [www.orgnet.com](http://www.orgnet.com).

# Yeast proteins

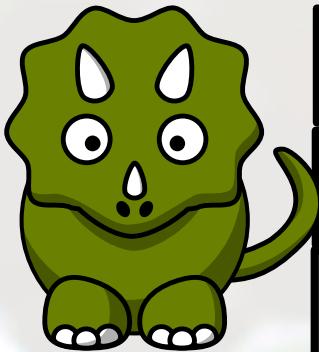


Yeast proteins: Sergei Maslov and Kim Sneppen,  
**Specificity and stability in topology of protein networks,**  
Science 296, 910-913 (2002).



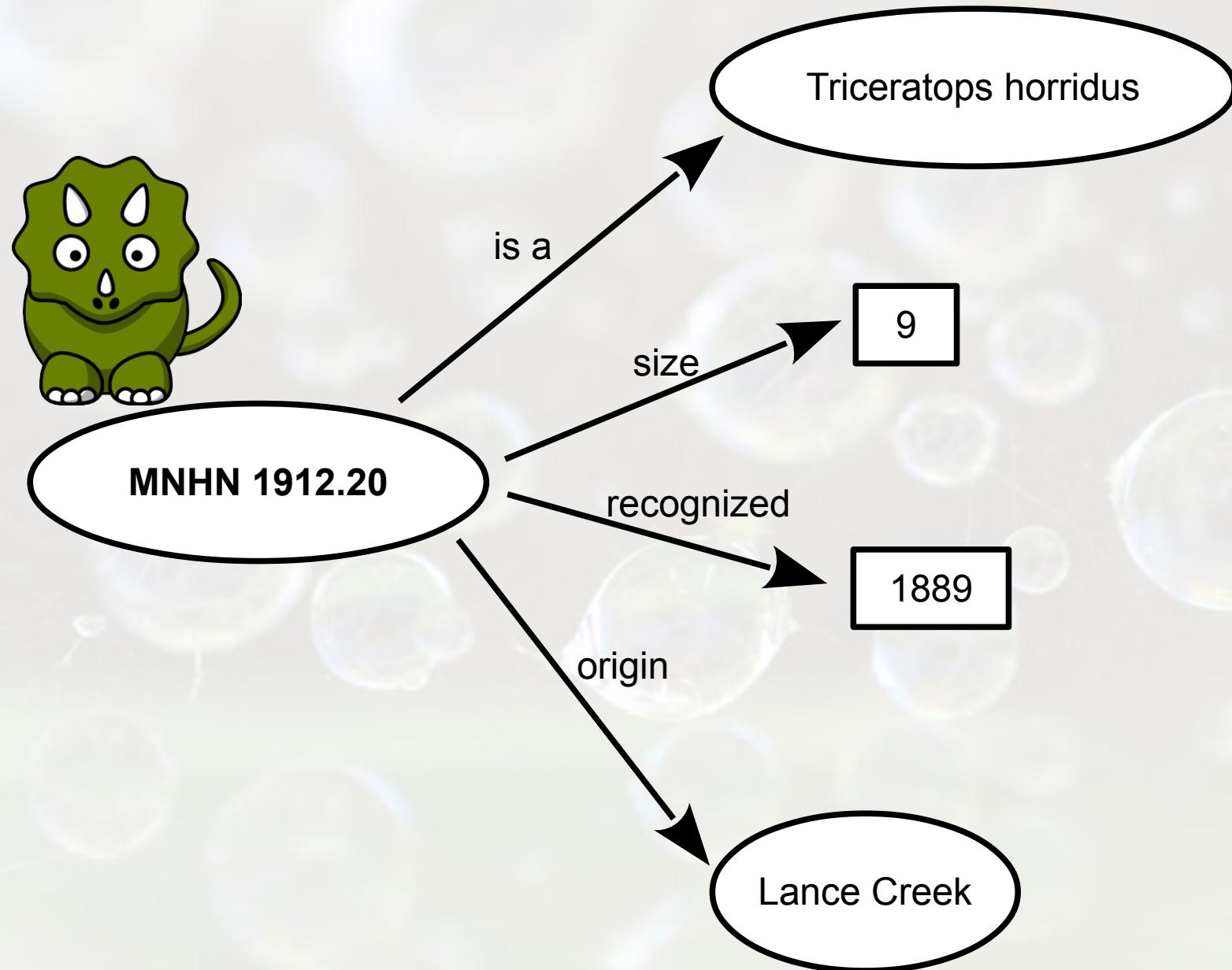
# What can I ask to a Graph?

# Triceratops in a Graph

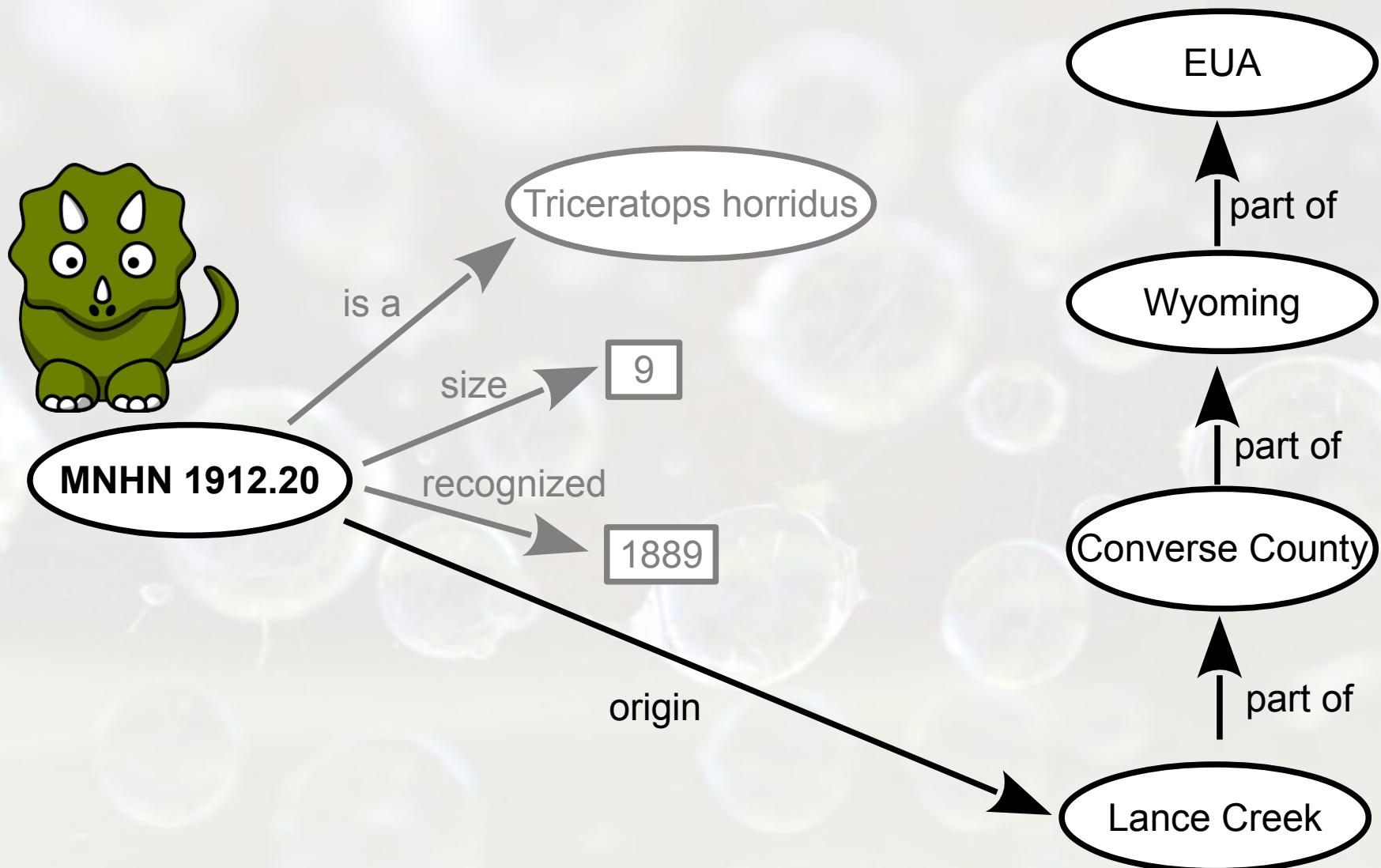


<b>MNHN 1912.20</b>		
Is a	Triceratops horridus	
Origin	Lance Creek	EUA
Recognized	1889	
Size	9	

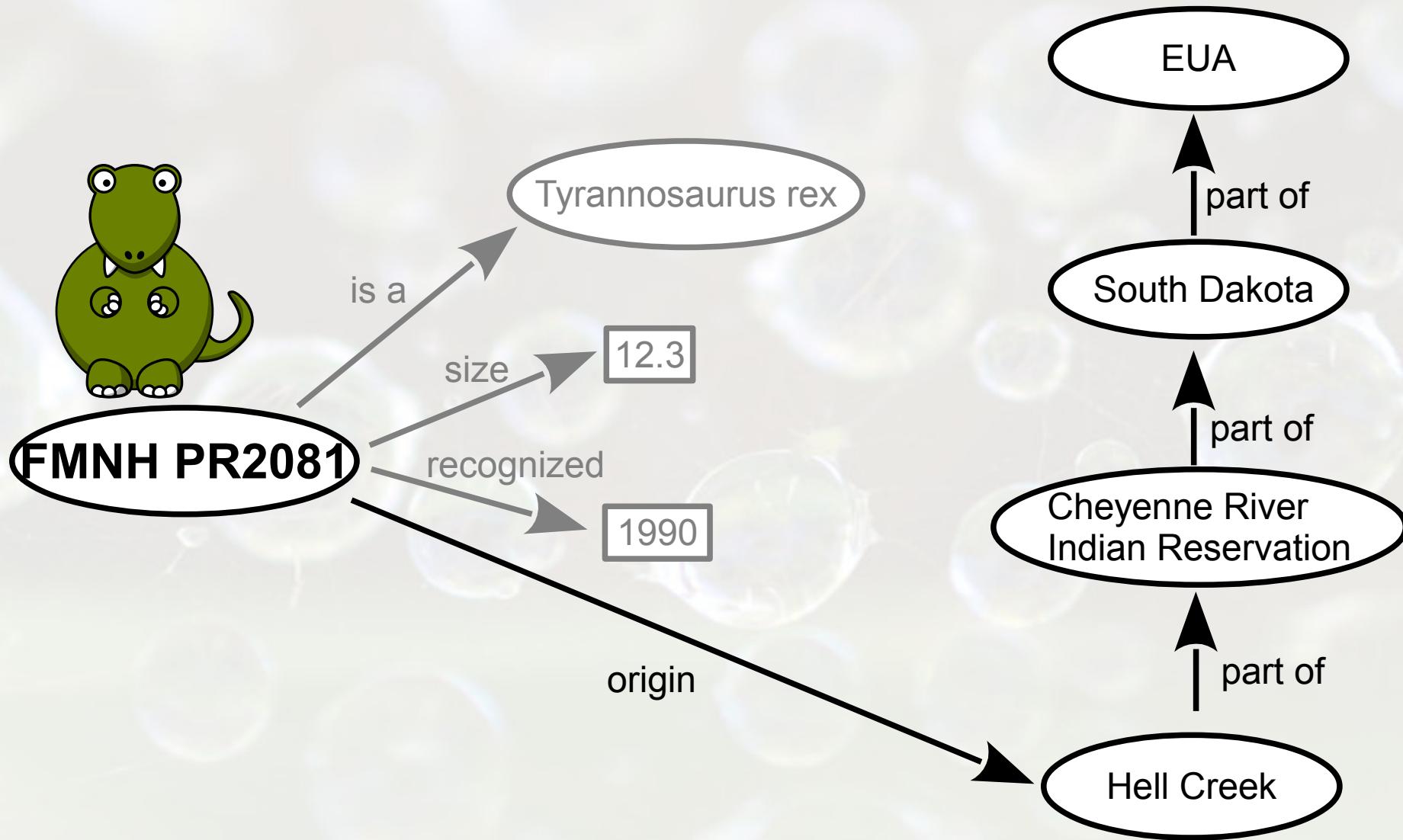
# Triceratops in a Graph



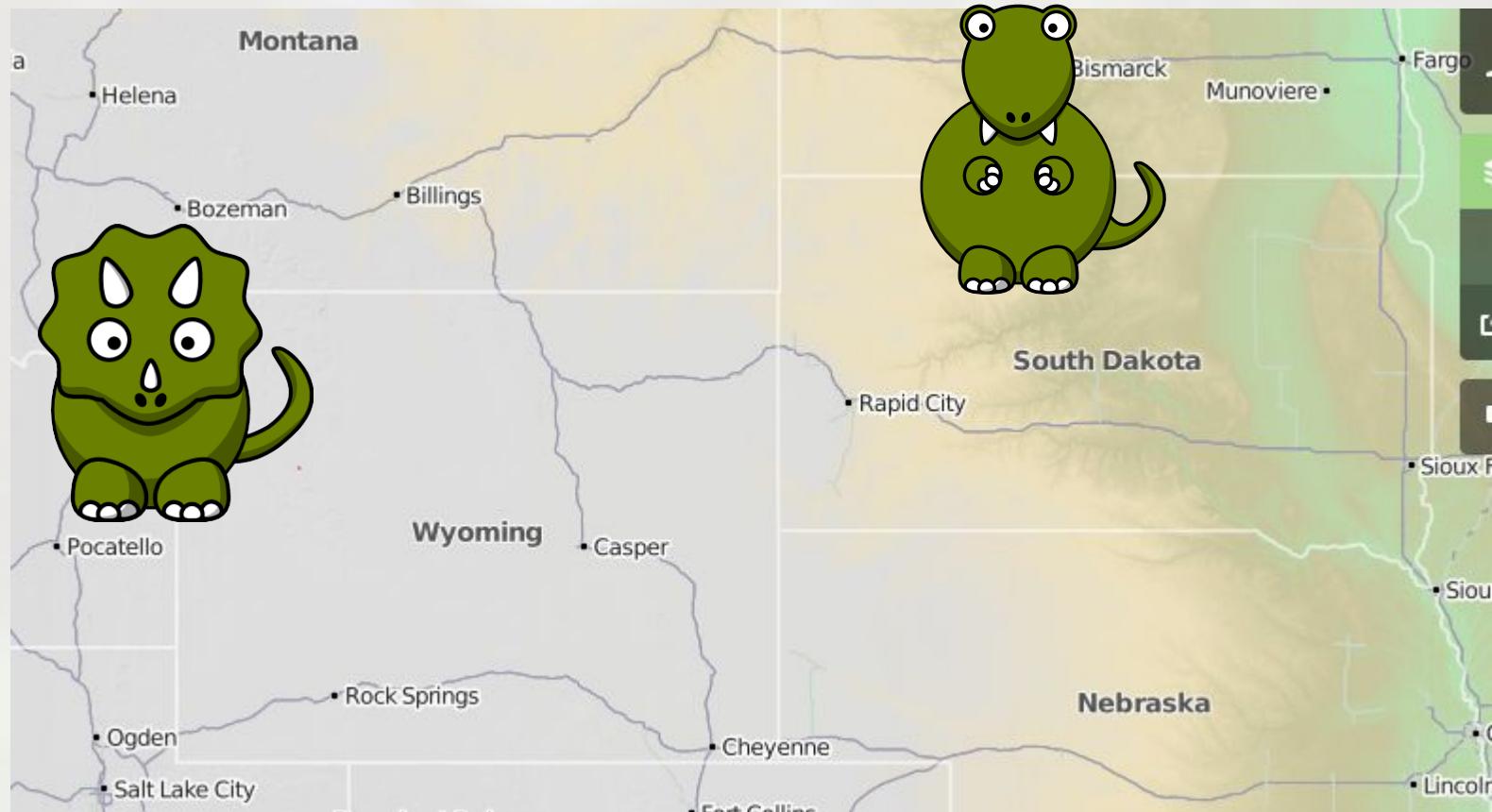
# Triceratops in a Graph



# Tyrannosaurus in a Graph

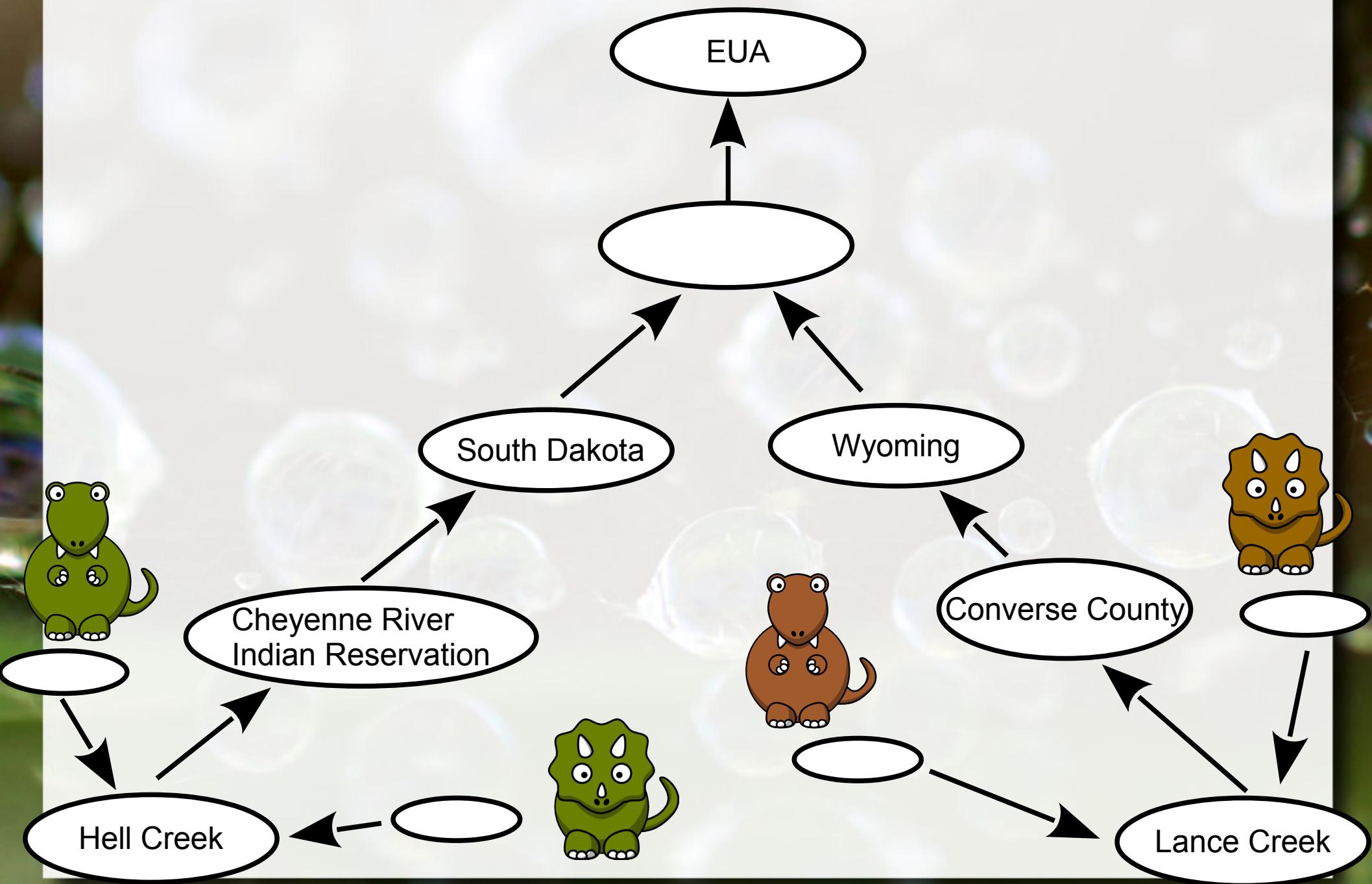


# Analyzing in the Space

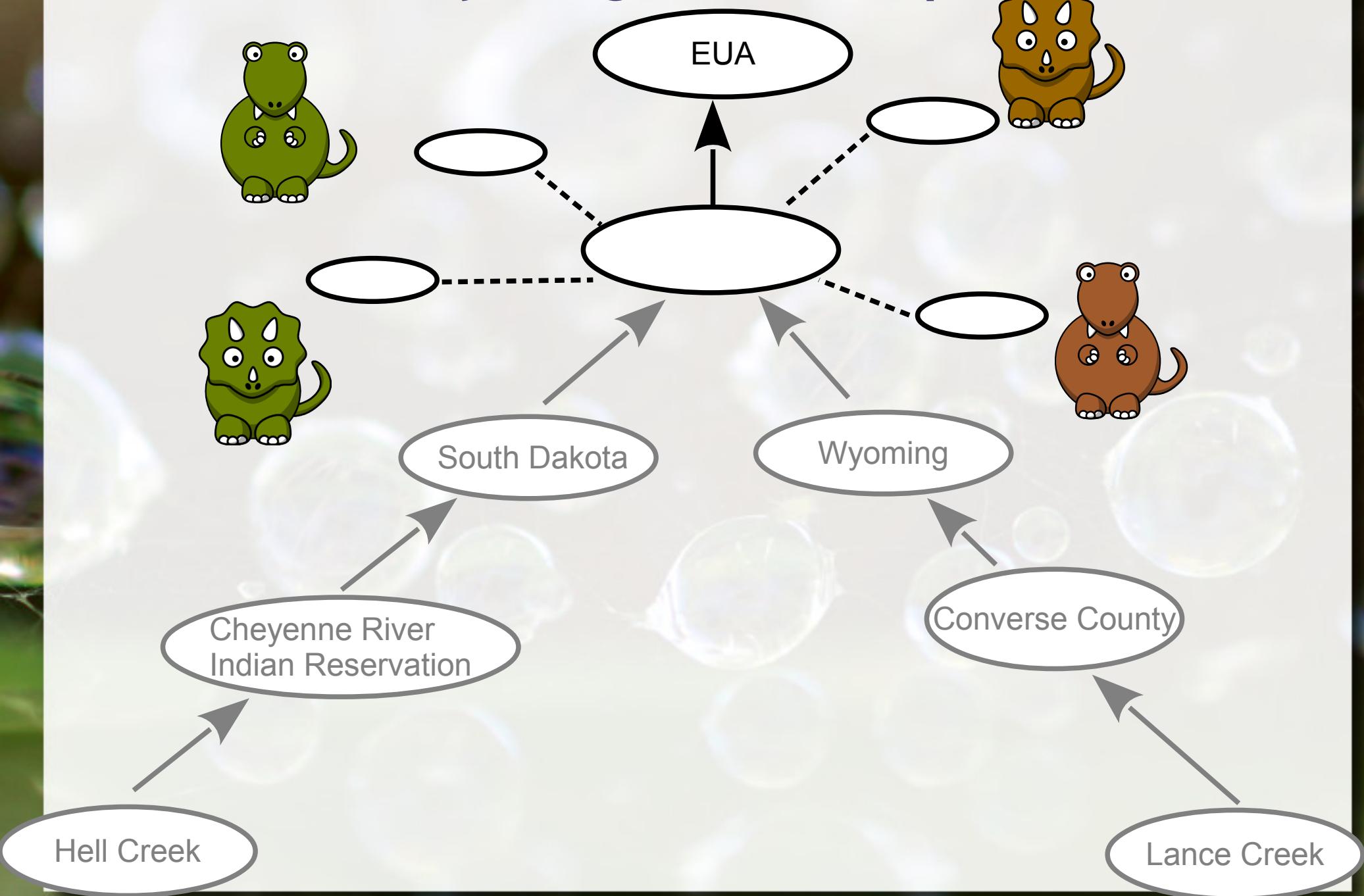


OpenStreetMap

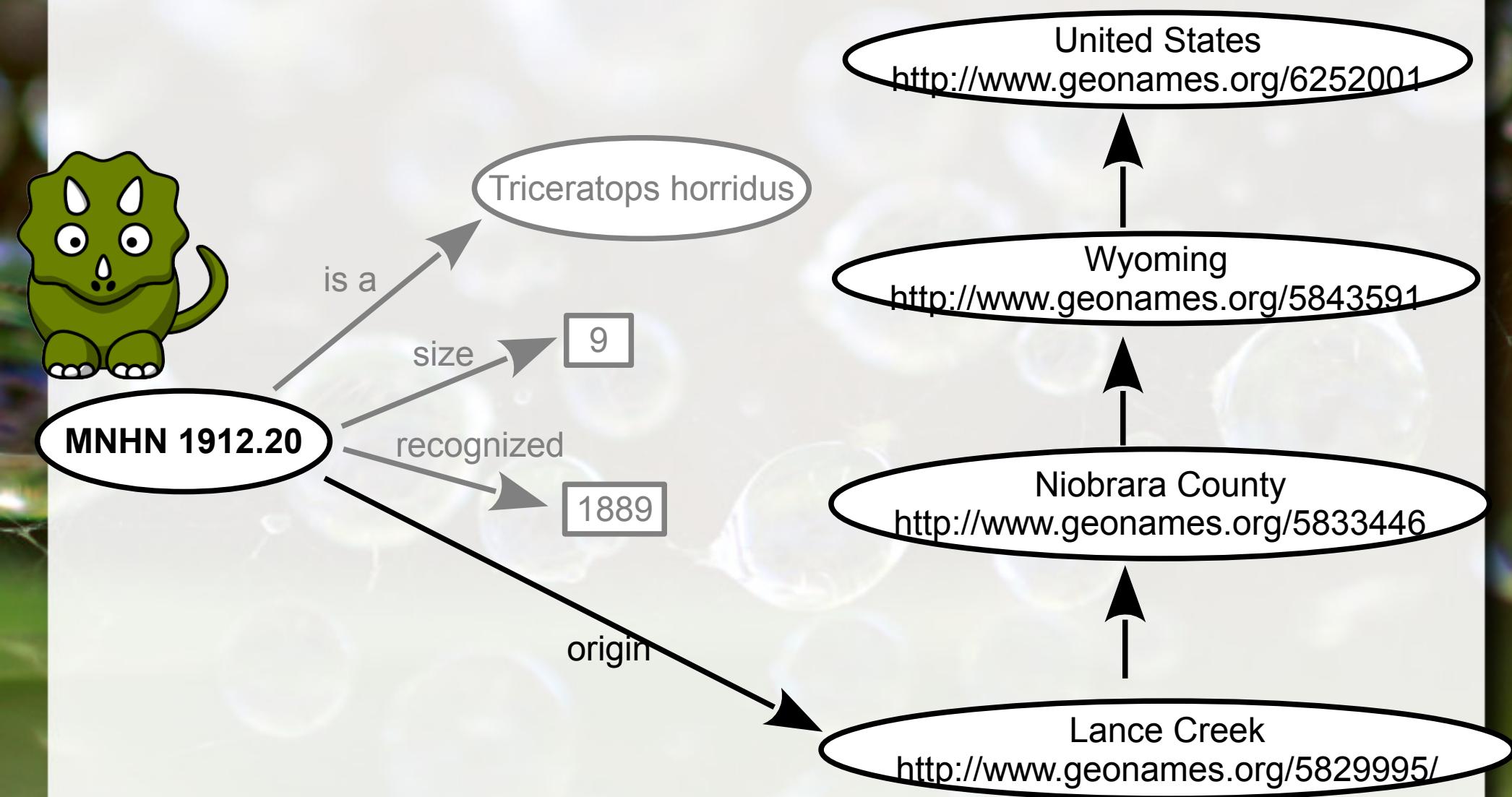
# Analyzing in the Space



# Analyzing in the Space



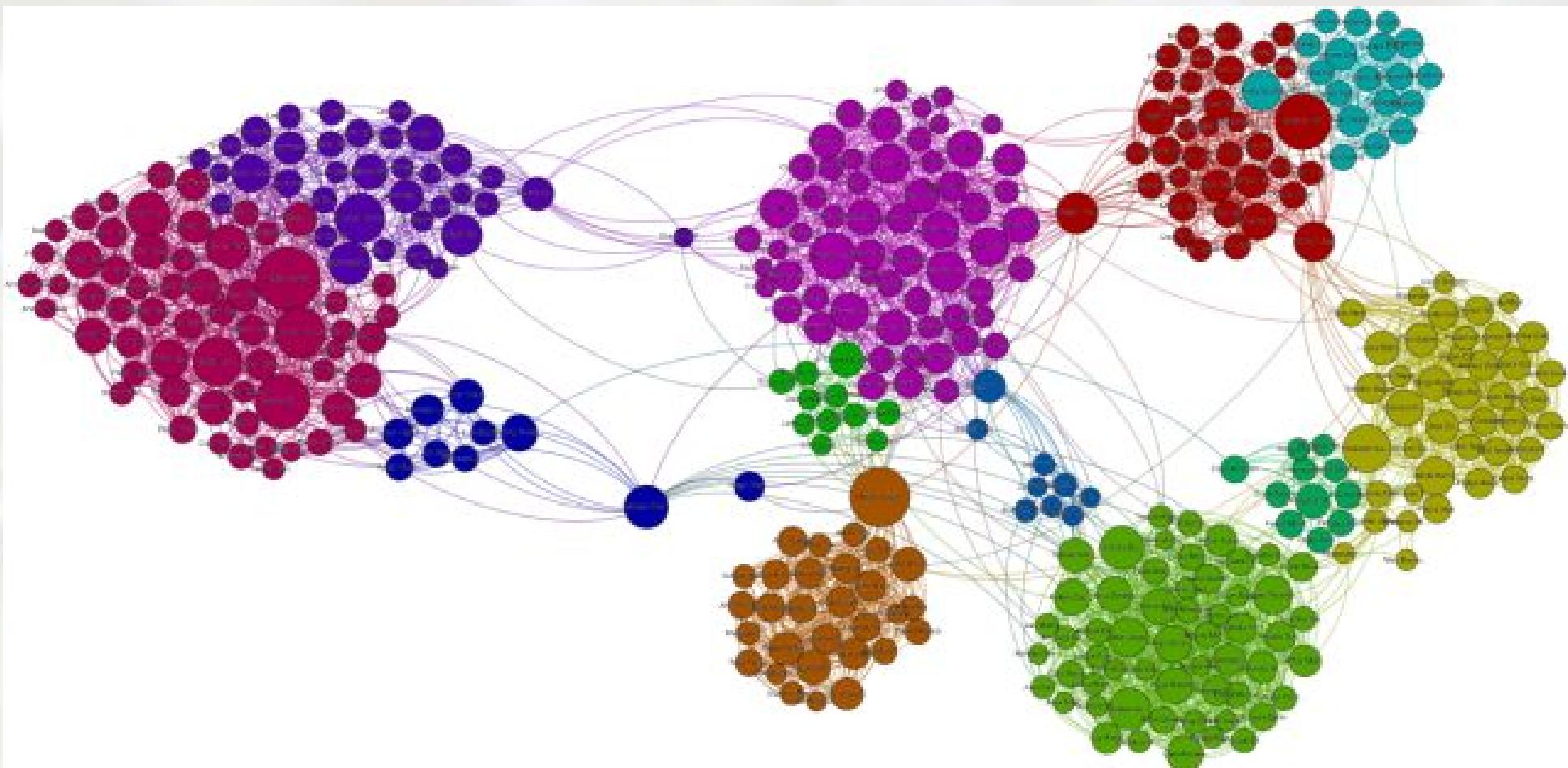
# GeoNames



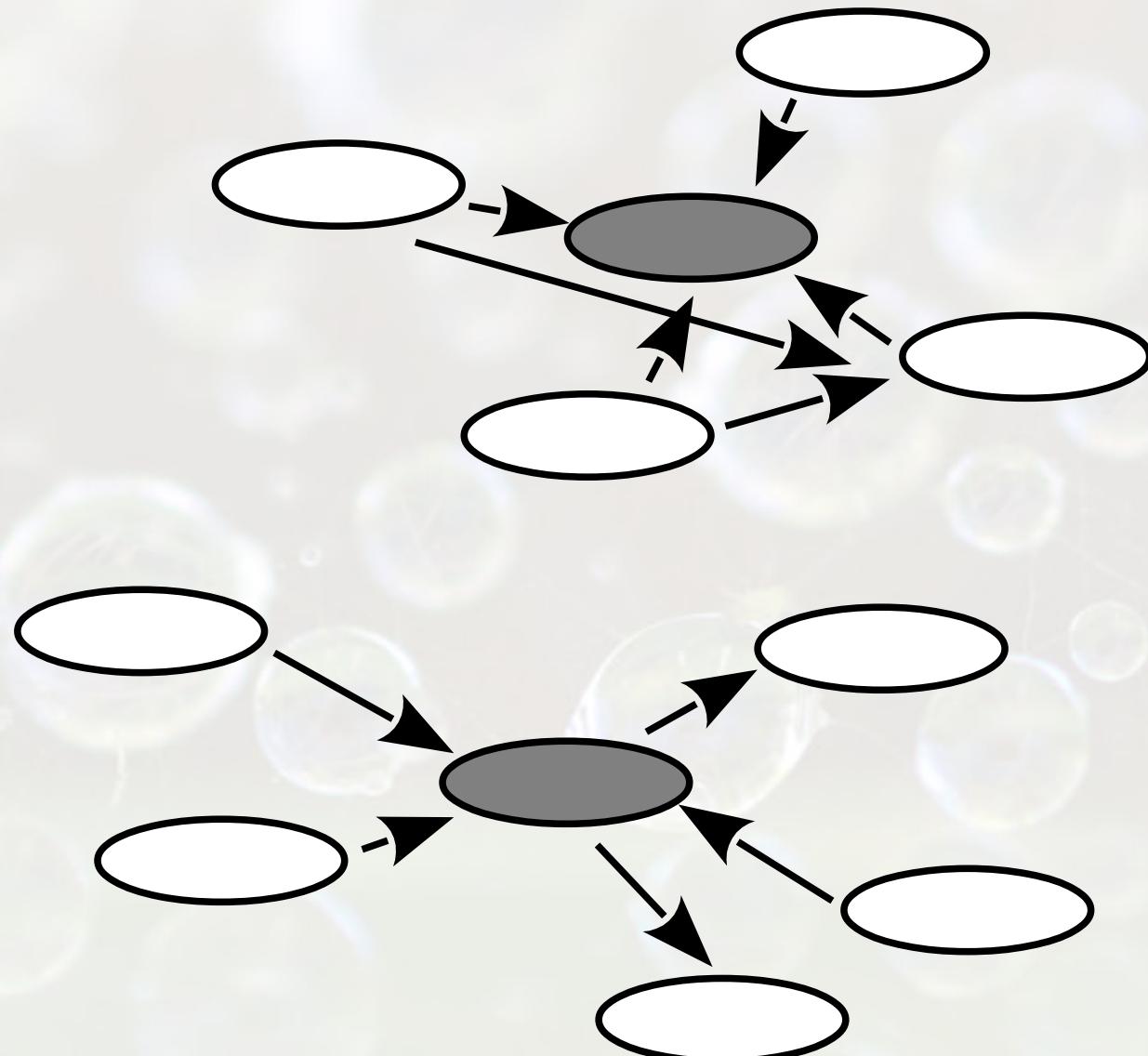
# Processing the Query

- Process by pattern
  - Find [species] whose [origin] → (0..\*) part of → EUA
- Process by inference
  - Find [species] whose [origin] → EUA
  - Rules:
    - If (A) origin (B) and (B) part of (C) => (A) origin (C)
    - If (A) part of (B) and (B) part of (C) => (A) part of (C)

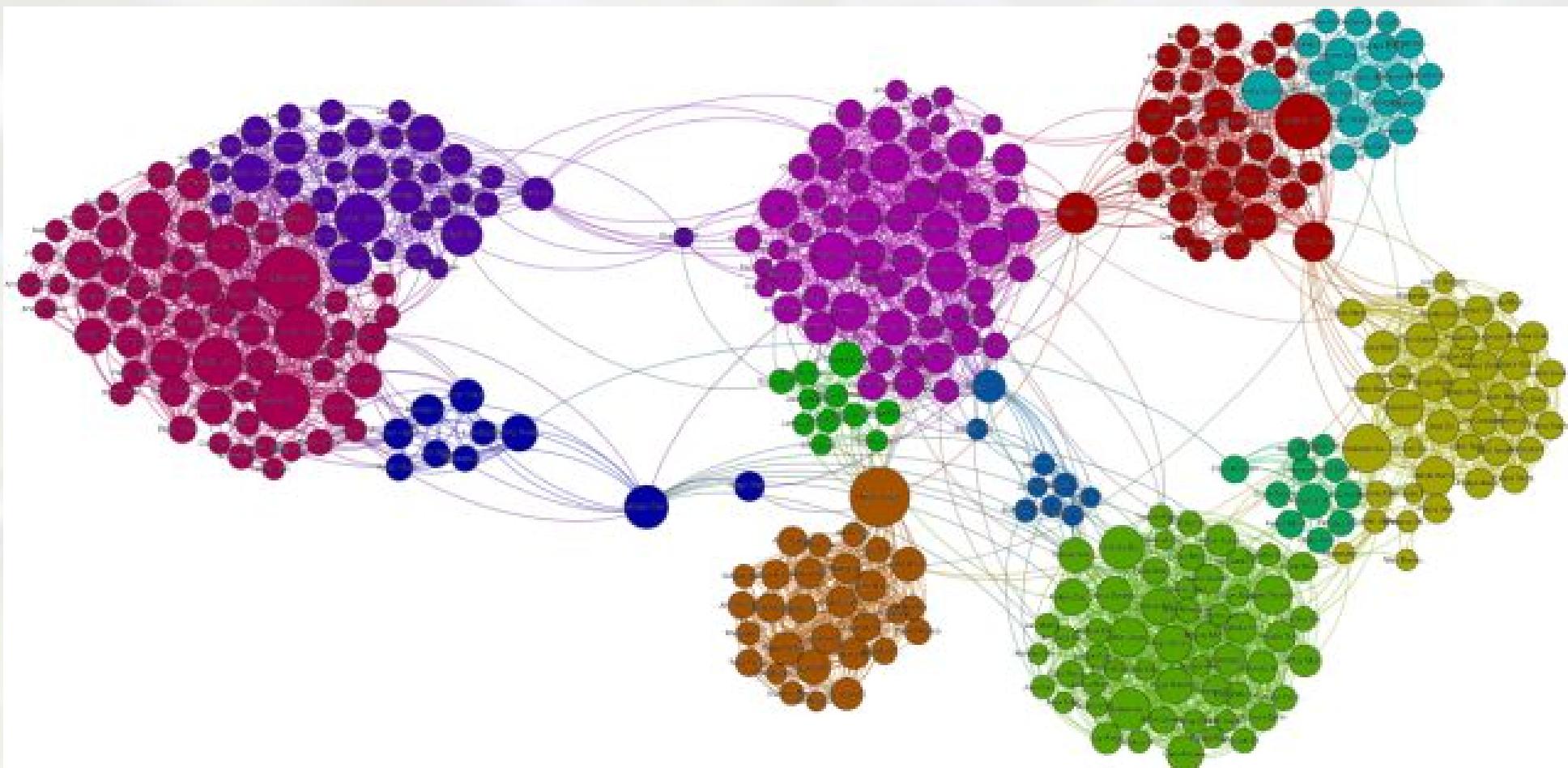
# Looking the Topology



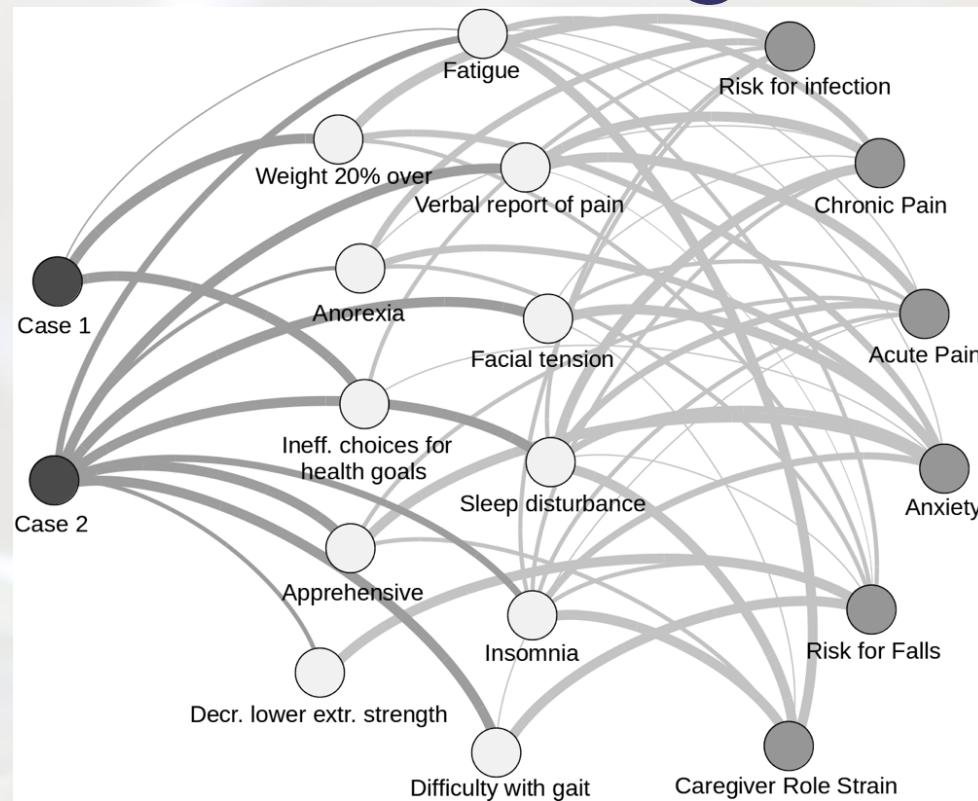
# Graph Topology



# Graph Topology



# Patient x Diagnosis

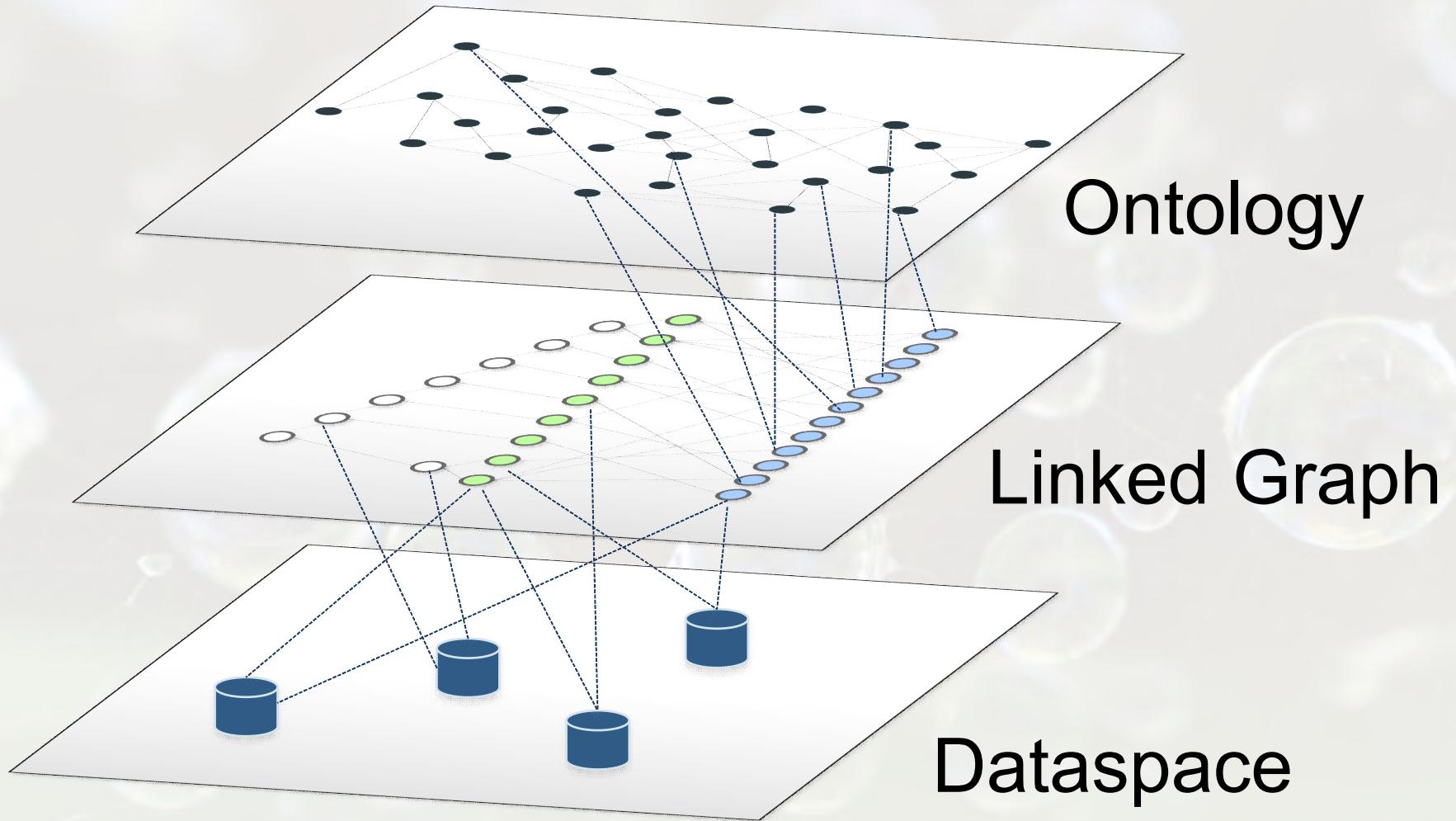


- Retrieve candidate **diagnosis** relevant to a given **patient** based on her **symptoms** and correlations with other similar patients.

(Gomes-Jr, 2013)

# From Dataspaces to Ontologies

# From Dataspaces to Ontologies



# Building Patterns

ID 

Schema 

Instances 

	A	B	C	D	E	F	G
1	Espécie	Área	Data	Hora	Isca	Amostra	coleta
23	<i>Eulaema cingulata</i>	Bom Jardim	8/7/2008	11:01	S.M.	1	2
24	<i>Eulaema meriana</i>	Bom Jardim	8/7/2008	13:01	S.M.	1	2
25	<i>Eulaema meriana</i>	Bom Jardim	8/7/2008	12:55	A.B.	1	2
26	<i>Eulaema meriana</i>	Bom Jardim	8/7/2008	13:11	S.M.	1	2
27	<i>Eulaema meriana</i>	Bom Jardim	8/7/2008	16:10	A.B.	1	2
28	<i>Eulaema meriana</i>	Bom Jardim	8/7/2008	12:09	S.M.	1	2
29	<i>Eulaema meriana</i>	Bom Jardim	8/7/2008	14:44	A.B.	1	2
30	<i>Eulaema meriana</i>	Bom Jardim	8/7/2008	11:51	A.B.	1	2
31	<i>Eulaema meriana</i>	Bom Jardim	8/7/2008	14:44	A.B.	1	2
32	<i>Eulaema meriana</i>	Bom Jardim	8/7/2008	12:04	A.B.	1	2
33	<i>Exaerete frontalis</i>	Bom Jardim	8/7/2008	11:27	S.M.	1	2
34	<i>Eufriesea flaviventris</i>	Bom Jardim	8/7/2008	12:40	S.M.	1	2
35	<i>Eufriesea flaviventris</i>	Bom Jardim	8/7/2008	13:25	S.M.	1	2
36	<i>Euglossa cognata</i>	Bom Jardim	8/7/2008	9:20	S.M.	1	2
37	<i>Euglossa ignita</i>	Bom Jardim	8/7/2008	11:08	A.B.	1	2
38	<i>Euglossa ignita</i>	Bom Jardim	8/7/2008	10:42	A.B.	1	2
39	<i>Euglossa imperialis</i>	Bom Jardim	8/7/2008	14:32	S.M.	1	2

when 

what 

where 

what 

what 

	Data	N. da Estação	Latitude	Longitude	Prof.	Classificação
3						
4	14/12/1997	6644	25° 45.80'	45° 11.77'	485	Litoclástico
5	14/12/1997	6645	25° 44.09'	45° 13.93'	256	Litobioclástico
6	14/12/1997	6646	25° 43.78'	45° 16.06'	198	Biolitoclástico

**1 Planilha Geral dos Dados da Biota**

**2 Arquivo de origem: Abelhas 2o. Periodo**

**3 ID \*** Número da unidade de coleta \*\*

**4** **5** **6** **7** **8** **9** **10** **11** **12** **13** **14** **15** **16** **17** **18** **19** **20** **21** **22** **23** **24** **25** **26** **27** **28** **29** **30** **31** **32** **33** **34** **35** **36** **37** **38** **39** **40** **41** **42** **43** **44** **45** **46** **47** **48** **49** **50** **51** **52** **53** **54** **55** **56** **57** **58** **59** **60** **61** **62** **63** **64** **65** **66** **67** **68** **69** **70** **71** **72** **73** **74** **75** **76** **77** **78** **79** **80** **81** **82** **83** **84** **85** **86** **87** **88** **89** **90** **91** **92** **93** **94** **95** **96** **97** **98** **99** **100** **101** **102** **103** **104** **105** **106** **107** **108** **109** **110** **111** **112** **113** **114** **115** **116** **117** **118** **119** **120** **121** **122** **123** **124** **125** **126** **127** **128** **129** **130** **131** **132** **133** **134** **135** **136** **137** **138** **139** **140** **141** **142** **143** **144** **145** **146** **147** **148** **149** **150** **151** **152** **153** **154** **155** **156** **157** **158** **159** **160** **161** **162** **163** **164** **165** **166** **167** **168** **169** **170** **171** **172** **173** **174** **175** **176** **177** **178** **179** **180** **181** **182** **183** **184** **185** **186** **187** **188** **189** **190** **191** **192** **193** **194** **195** **196** **197** **198** **199** **200** **201** **202** **203** **204** **205** **206** **207** **208** **209** **210** **211** **212** **213** **214** **215** **216** **217** **218** **219** **220** **221** **222** **223** **224** **225** **226** **227** **228** **229** **230** **231** **232** **233** **234** **235** **236** **237** **238** **239** **240** **241** **242** **243** **244** **245** **246** **247** **248** **249** **250** **251** **252** **253** **254** **255** **256** **257** **258** **259** **260** **261** **262** **263** **264** **265** **266** **267** **268** **269** **270** **271** **272** **273** **274** **275** **276** **277** **278** **279** **280** **281** **282** **283** **284** **285** **286** **287** **288** **289** **290** **291** **292** **293** **294** **295** **296** **297** **298** **299** **300** **301** **302** **303** **304** **305** **306** **307** **308** **309** **310** **311** **312** **313** **314** **315** **316** **317** **318** **319** **320** **321** **322** **323** **324** **325** **326** **327** **328** **329** **330** **331** **332** **333** **334** **335** **336** **337** **338** **339** **340** **341** **342** **343** **344** **345** **346** **347** **348** **349** **350** **351** **352** **353** **354** **355** **356** **357** **358** **359** **360** **361** **362** **363** **364** **365** **366** **367** **368** **369** **370** **371** **372** **373** **374** **375** **376** **377** **378** **379** **380** **381** **382** **383** **384** **385** **386** **387** **388** **389** **390** **391** **392** **393** **394** **395** **396** **397** **398** **399** **400** **401** **402** **403** **404** **405** **406** **407** **408** **409** **410** **411** **412** **413** **414** **415** **416** **417** **418** **419** **420** **421** **422** **423** **424** **425** **426** **427** **428** **429** **430** **431** **432** **433** **434** **435** **436** **437** **438** **439** **440** **441** **442** **443** **444** **445** **446** **447** **448** **449** **450** **451** **452** **453** **454** **455** **456** **457** **458** **459** **460** **461** **462** **463** **464** **465** **466** **467** **468** **469** **470** **471** **472** **473** **474** **475** **476** **477** **478** **479** **480** **481** **482** **483** **484** **485** **486** **487** **488** **489** **490** **491** **492** **493** **494** **495** **496** **497** **498** **499** **500** **501** **502** **503** **504** **505** **506** **507** **508** **509** **510** **511** **512** **513** **514** **515** **516** **517** **518** **519** **520** **521** **522** **523** **524** **525** **526** **527** **528** **529** **530** **531** **532** **533** **534** **535** **536** **537** **538** **539** **540** **541** **542** **543** **544** **545** **546** **547** **548** **549** **550** **551** **552** **553** **554** **555** **556** **557** **558** **559** **560** **561** **562** **563** **564** **565** **566** **567** **568** **569** **570** **571** **572** **573** **574** **575** **576** **577** **578** **579** **580** **581** **582** **583** **584** **585** **586** **587** **588** **589** **590** **591** **592** **593** **594** **595** **596** **597** **598** **599** **600** **601** **602** **603** **604** **605** **606** **607** **608** **609** **610** **611** **612** **613** **614** **615** **616** **617** **618** **619** **620** **621** **622** **623** **624** **625** **626** **627** **628** **629** **630** **631** **632** **633** **634** **635** **636** **637** **638** **639** **640** **641** **642** **643** **644** **645** **646** **647** **648** **649** **650** **651** **652** **653** **654** **655** **656** **657** **658** **659** **660** **661** **662** **663** **664** **665** **666** **667** **668** **669** **670** **671** **672** **673** **674** **675** **676** **677** **678** **679** **680** **681** **682** **683** **684** **685** **686** **687** **688** **689** **690** **691** **692** **693** **694** **695** **696** **697** **698** **699** **700** **701** **702** **703** **704** **705** **706** **707** **708** **709** **710** **711** **712** **713** **714** **715** **716** **717** **718** **719** **720** **721** **722** **723** **724** **725** **726** **727** **728** **729** **730** **731** **732** **733** **734** **735** **736** **737** **738** **739** **7310** **7311** **7312** **7313** **7314** **7315** **7316** **7317** **7318** **7319** **7320** **7321** **7322** **7323** **7324** **7325** **7326** **7327** **7328** **7329** **7330** **7331** **7332** **7333** **7334** **7335** **7336** **7337** **7338** **7339** **7340** **7341** **7342** **7343** **7344** **7345** **7346** **7347** **7348** **7349** **7350** **7351** **7352** **7353** **7354** **7355** **7356** **7357** **7358** **7359** **7360** **7361** **7362** **7363** **7364** **7365** **7366** **7367** **7368** **7369** **7370** **7371** **7372** **7373** **7374** **7375** **7376** **7377** **7378** **7379** **7380** **7381** **7382** **7383** **7384** **7385** **7386** **7387** **7388** **7389** **7390** **7391** **7392** **7393** **7394** **7395** **7396** **7397** **7398** **7399** **73100** **73101** **73102** **73103** **73104** **73105** **73106** **73107** **73108** **73109** **73110** **73111** **73112** **73113** **73114** **73115** **73116** **73117** **73118** **73119** **73120** **73121** **73122** **73123** **73124** **73125** **73126** **73127** **73128** **73129** **73130** **73131** **73132** **73133** **73134** **73135** **73136** **73137** **73138** **73139** **73140** **73141** **73142** **73143** **73144** **73145** **73146** **73147** **73148** **73149** **73150** **73151** **73152** **73153** **73154** **73155** **73156** **73157** **73158** **73159** **73160** **73161** **73162** **73163** **73164** **73165** **73166** **73167** **73168** **73169** **73170** **73171** **73172** **73173** **73174** **73175** **73176** **73177** **73178** **73179** **73180** **73181** **73182** **73183** **73184** **73185** **73186** **73187** **73188** **73189** **73190** **73191** **73192** **73193** **73194** **73195** **73196** **73197** **73198** **73199** **73200** **73201** **73202** **73203** **73204** **73205** **73206** **73207** **73208** **73209** **73210** **73211** **73212** **73213** **73214** **73215** **73216** **73217** **73218** **73219** **73220** **73221** **73222** **73223** **73224** **73225** **73226** **73227** **73228** **73229** **73230** **73231** **73232** **73233** **73234** **73235** **73236** **73237** **73238** **73239** **73240** **73241** **73242** **73243** **73244** **73245** **73246** **73247** **73248** **73249** **73250** **73251** **73252** **73253** **73254** **73255** **73256** **73257** **73258** **73259** **73260** **73261** **73262** **73263** **73264** **73265** **73266** **73267** **73268** **73269** **73270** **73271** **73272** **73273** **73274** **73275** **73276** **73277** **73278** **73279** **73280** **73281** **73282** **73283** **73284** **73285** **73286** **73287** **73288** **73289** **73290** **73291** **73292** **73293** **73294** **73295** **73296** **73297** **73298** **73299** **73300** **73301** **73302** **73303** **73304** **73305** **73306** **73307** **73308** **73309** **73310** **73311** **73312** **73313** **73314** **73315** **73316** **73317** **73318** **73319** **73320** **73321** **73322** **73323** **73324** **73325** **73326** **73327** **73328** **73329** **73330** **73331** **73332** **73333** **73334** **73335** **73336** **73337** **73338** **73339** **73340** **73341** **73342** **73343** **73344** **73345** **73346** **73347** **73348** **73349** **73350** **73351** **73352** **73353** **73354** **73355** **73356** **73357** **73358** **73359** **73360** **73361** **73362** **73363** **73364** **73365** **73366** **73367** **73368** **73369** **73370** **73371** **73372** **73373** **73374** **73375** **73376** **73377** **73378** **73379** **73380** **73381** **73382** **73383** **73384** **73385** **73386** **73387** **73388** **73389** **73390** **73391** **73392** **73393** **73394** **73395** **73396** **73397** **73398** **73399** **73400** **73401** **73402** **73403** **73404** **73405** **73406** **73407** **73408** **73409** **73410** **73411** **73412** **73413** **73414** **73415** **73416** **73417** **73418** **73419** **73420** **73421** **73422** **73423** **73424** **73425** **73426** **73427** **73428** **73429** **73430** **73431** **73432** **73433** **73434** **73435** **73436** **73437** **73438** **73439** **73440** **73441** **73442** **73443** **73444** **73445** **73446** **73447** **73448** **73449** **73450** **73451** **73452** **73453** **73454** **73455** **73456** **73457** **73458** **73459** **73460** **73461** **73462** **73463** **73464** **73465** **73466** **73467** **73468** **73469** **73470** **73471** **73472** **73473** **73474** **73475** **73476** **73477** **73478** **73479** **73480** **73481** **73482** **73483** **73484** **73485** **73486** **73487** **73488** **73489** **73490** **73491** **73492** **73493** **73494** **73495** **73496** **73497** **73498** **73499** **73500** **73501** **73502** **73503** **73504** **73505** **73506** **73507** **73508** **73509** **73510** **73511** **73512** **73513** **73514** **73515** **73516** **73517** **73518** **73519** **73520** **73521** **73522** **73523** **73524** **73525** **73526** **73527** **73528** **73529** **73530** **73531** **73532** **73533** **73534** **73535** **73536** **73537** **73538** **73539** **73540** **73541** **73542**

# Data Exploration Prototype

BioSpread Project - LIS IC UNICAMP - Mozilla Firefox 27 °C Sun Jan 13 09:32:22 Matheus Mota

localhost/biospreadPortal/v2/?task=pages/txnavigator

Google Translate BioSpread Project icterus galbula | ... icterus galbula | ... icterus galbula | ... Virtuoso SPARQL... Error HTTP/1.1 4... http://s...debug=on

Comm Academic To Study To Read Utilities Misc Viagem AU Feeling lucky: How i... Of Luck and Success... Linked Open Aalto ... How to access cross...

## BioSpread Project

Integrating data from spreadsheet abroad

LIS UNICAMP

[Home](#) [About](#) [Contact](#) [Examples](#) [Data](#) [Playground](#) [SPARQL Endpoint](#) [License](#)

### A Taxonomy Navigator

A JavaScript navigator for the extracted data. Navigate through the select boxes and retrieve more information about species

animalia chordata aves passeriformes icteridae icterus icterus galbula

Taxonomy Path: animalia> chordata> aves> passeriformes> icteridae> icterus> icterus galbula

About: "icterus galbula"

Associated Events: Records(200)

Map showing collection locations in South America, with a callout for a record from Mato Grosso, Brazil.

Item uri: <http://purl.org/biospread/resource/collectionitem/30707> [About the collected item](#) [About the species](#)

Location Description: "mato grosso general carneiro"

Country: Brazil

Permalink: -40.15722, -27.39128

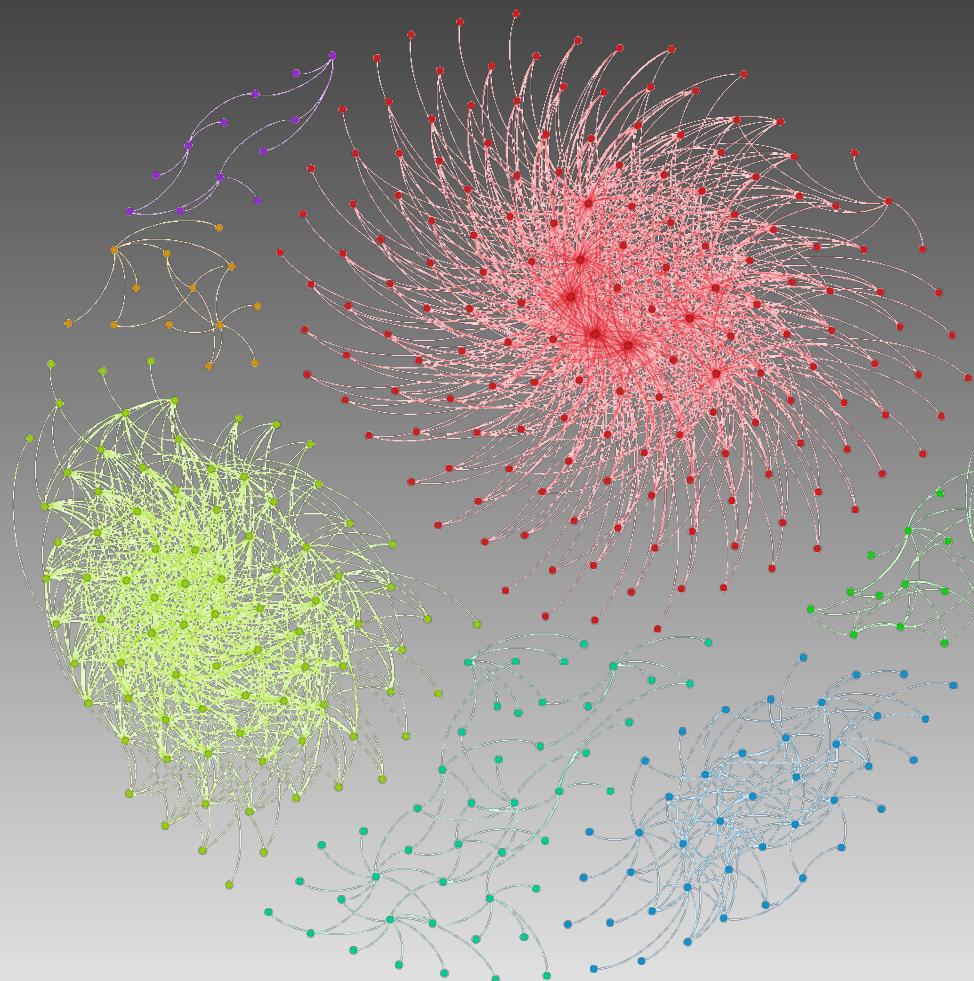
Powered by Google Map data ©2013 - Terms of Use

Find: paola

# Integrating phenotypes

7 distinct morphological descriptions:

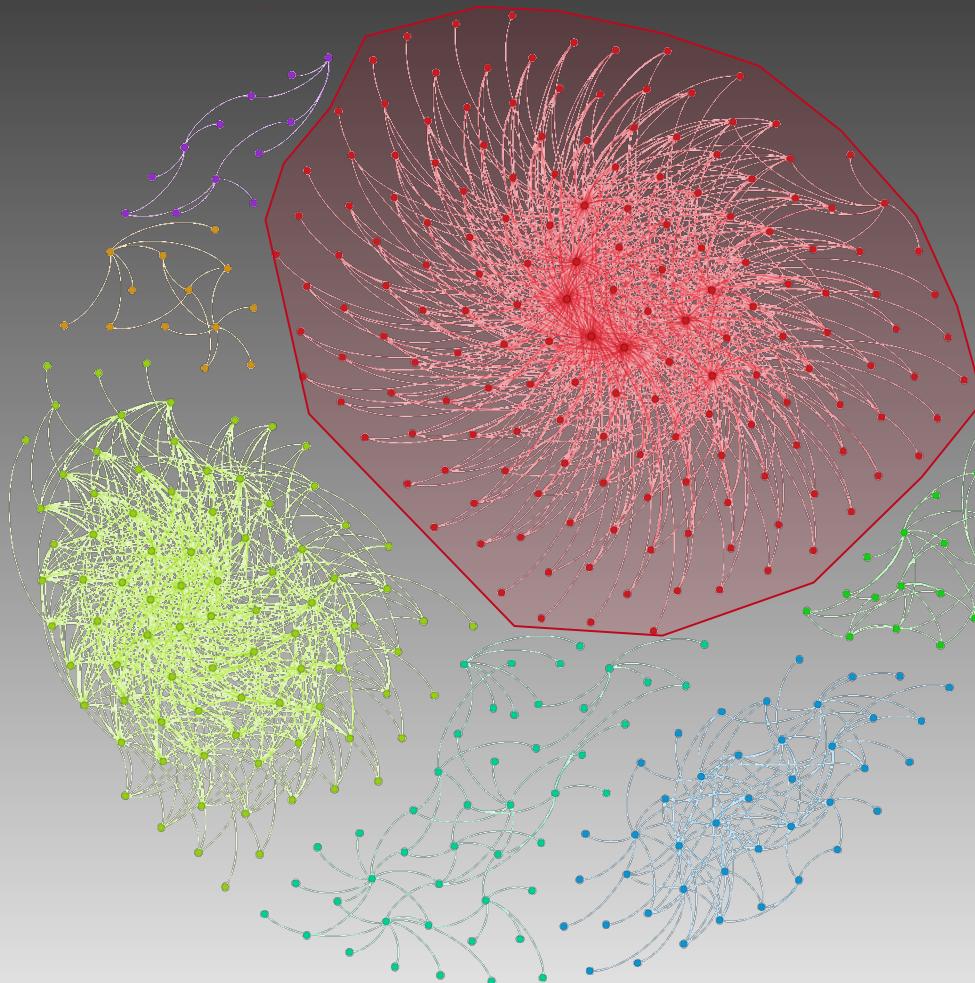
- genus *Varanus*
- *Varanus gouldii*
- *Varanus indicus*
- *Varanus prasinus*
- *Varanus salvator*
- *Varanus spiny*
- *Varanus timorensis*



# Our approach

7 distinct morphological descriptions:

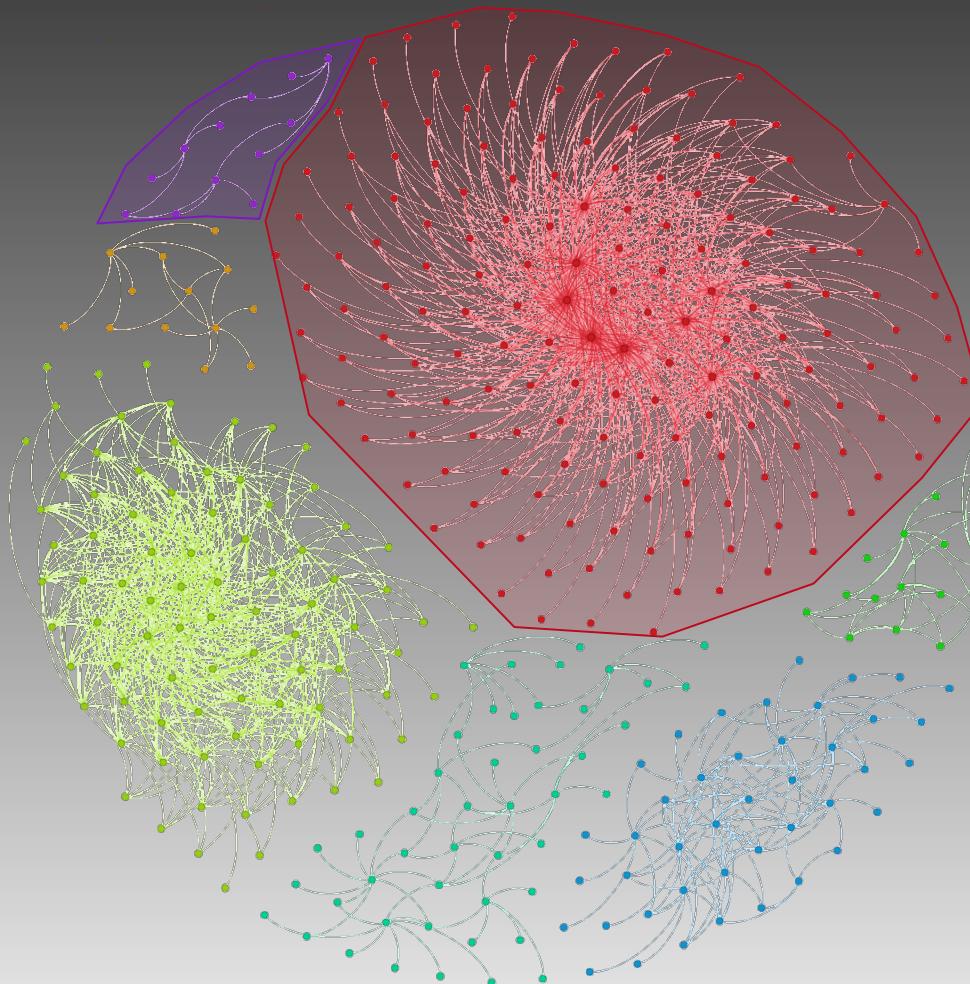
- genus *Varanus*
- *Varanus gouldii*
- *Varanus indicus*
- *Varanus prasinus*
- *Varanus salvator*
- *Varanus spiny*
- *Varanus timorensis*



# Our approach

7 distinct morphological descriptions:

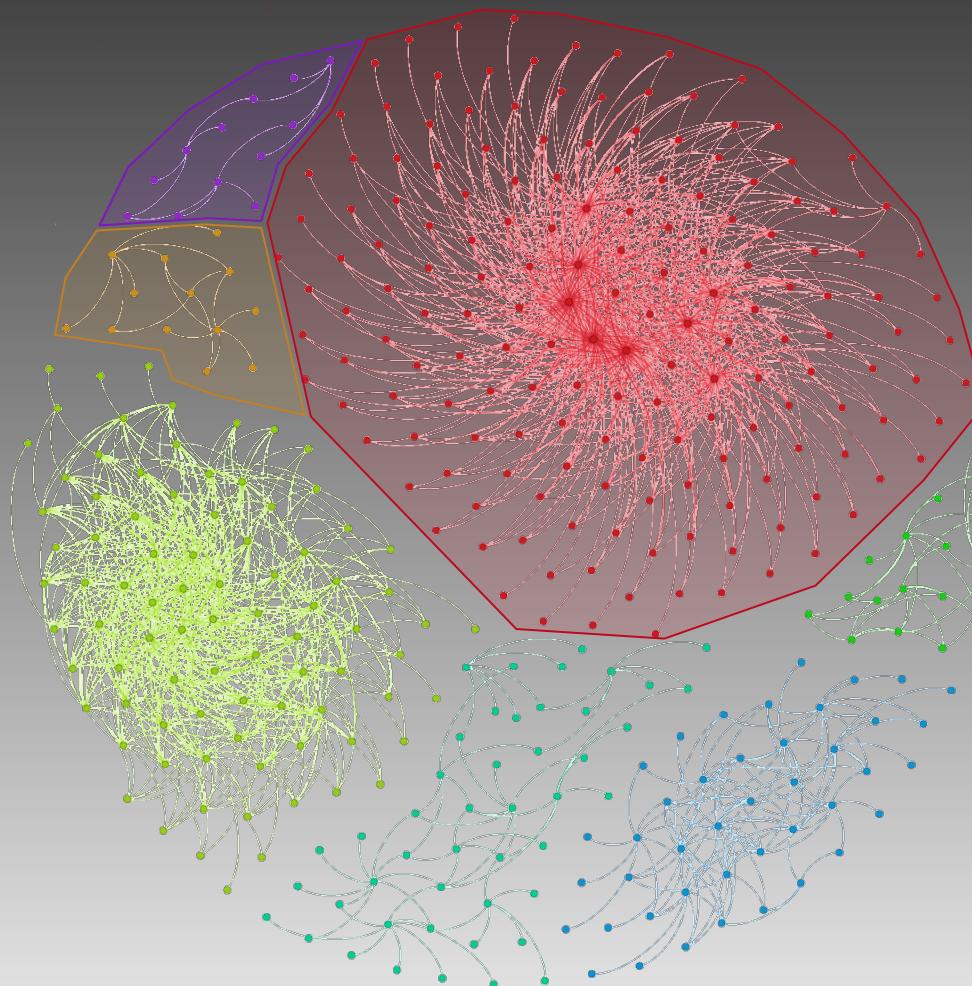
- genus *Varanus*
- *Varanus gouldii*
- *Varanus indicus*
- *Varanus prasinus*
- *Varanus salvator*
- *Varanus spiny*
- *Varanus timorensis*



# Our approach

7 distinct morphological descriptions:

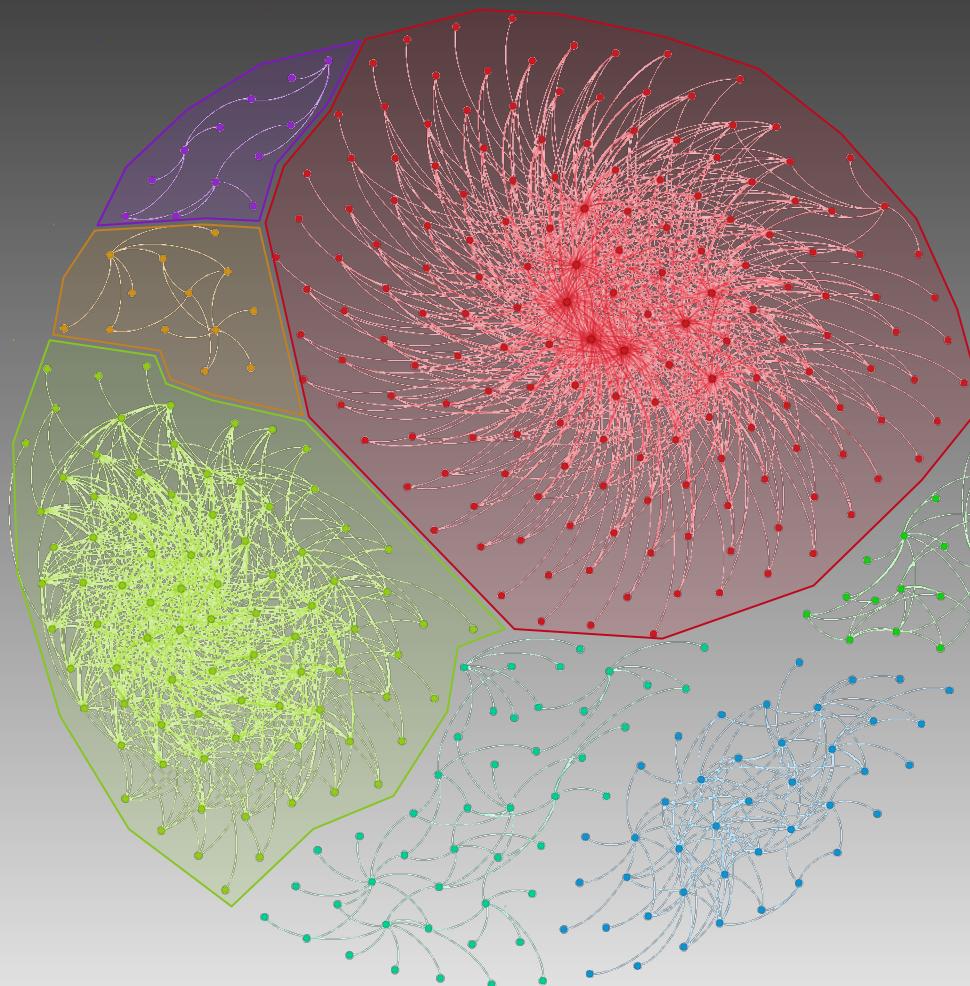
- genus *Varanus*
- *Varanus gouldii*
- *Varanus indicus*
- *Varanus prasinus*
- *Varanus salvator*
- *Varanus spiny*
- *Varanus timorensis*



# Our approach

7 distinct morphological descriptions:

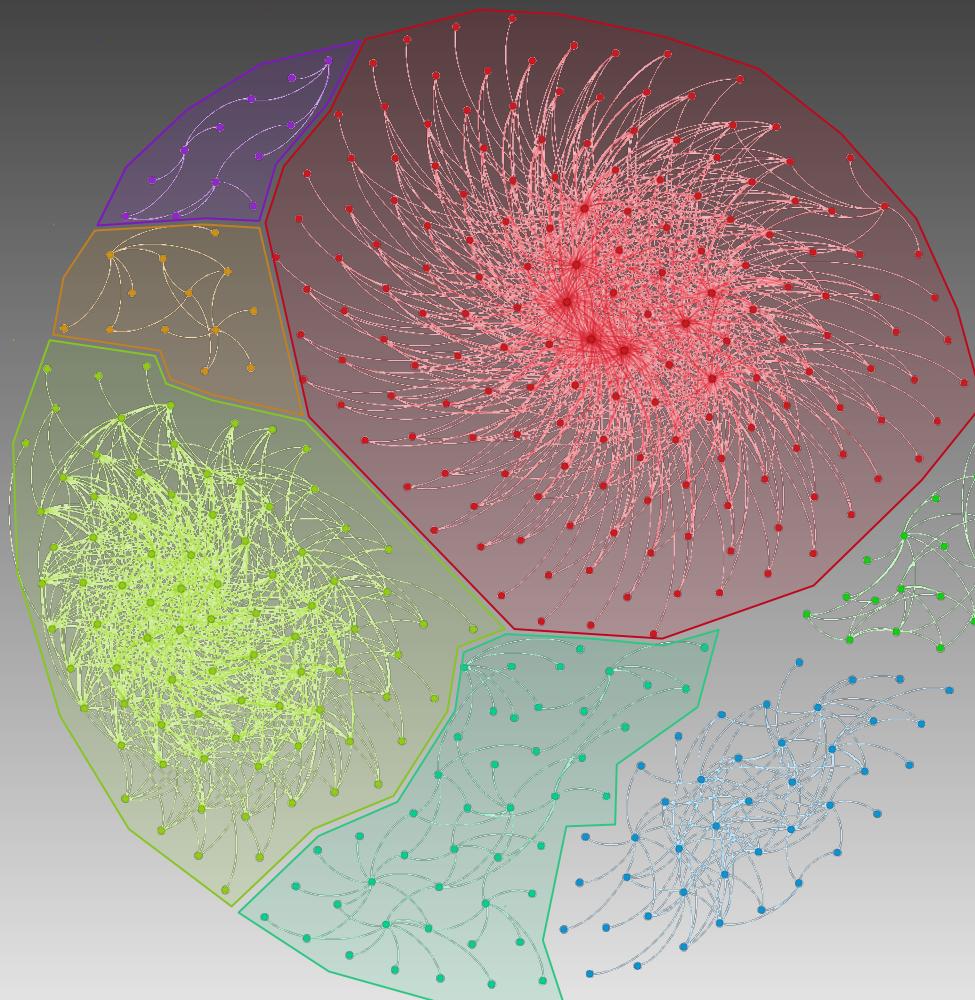
- genus *Varanus*
- *Varanus gouldii*
- *Varanus indicus*
- *Varanus prasinus*
- *Varanus salvator*
- *Varanus spiny*
- *Varanus timorensis*



# Our approach

7 distinct morphological descriptions:

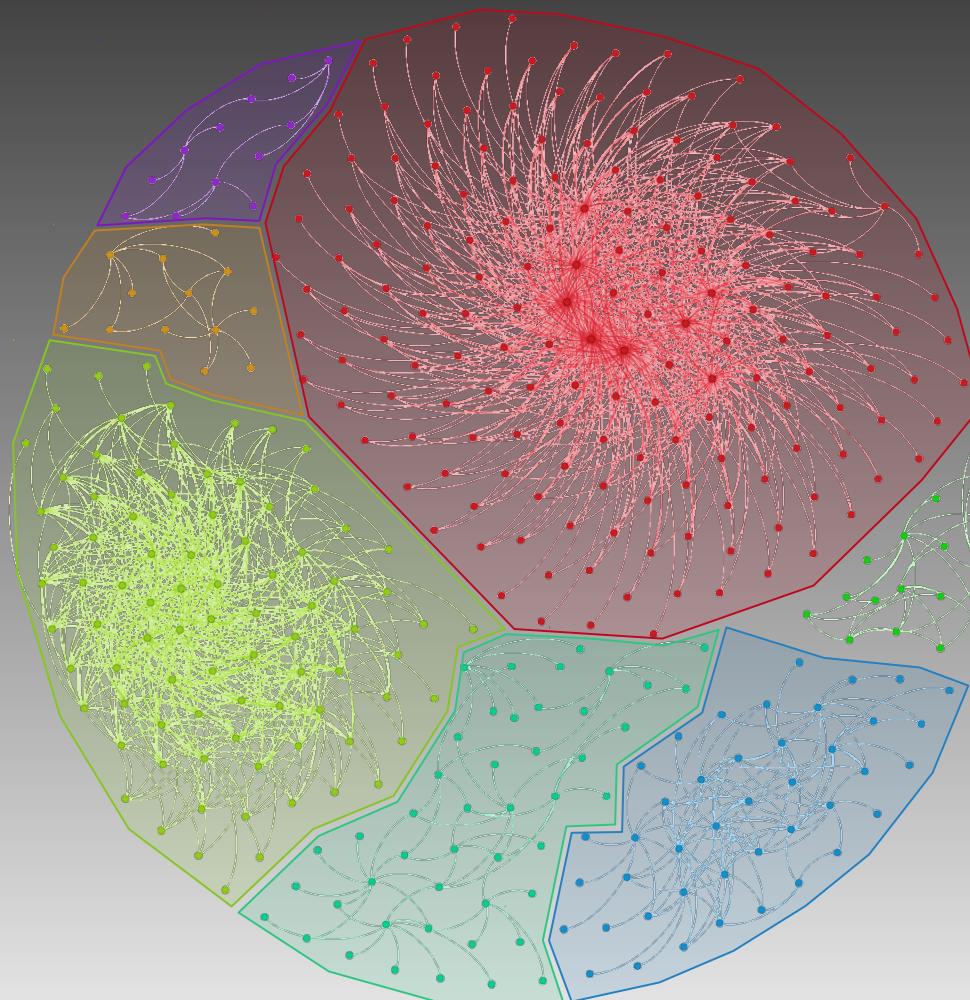
- genus *Varanus*
- *Varanus gouldii*
- *Varanus indicus*
- *Varanus prasinus*
- *Varanus salvator*
- *Varanus spiny*
- *Varanus timorensis*



# Our approach

7 distinct morphological descriptions:

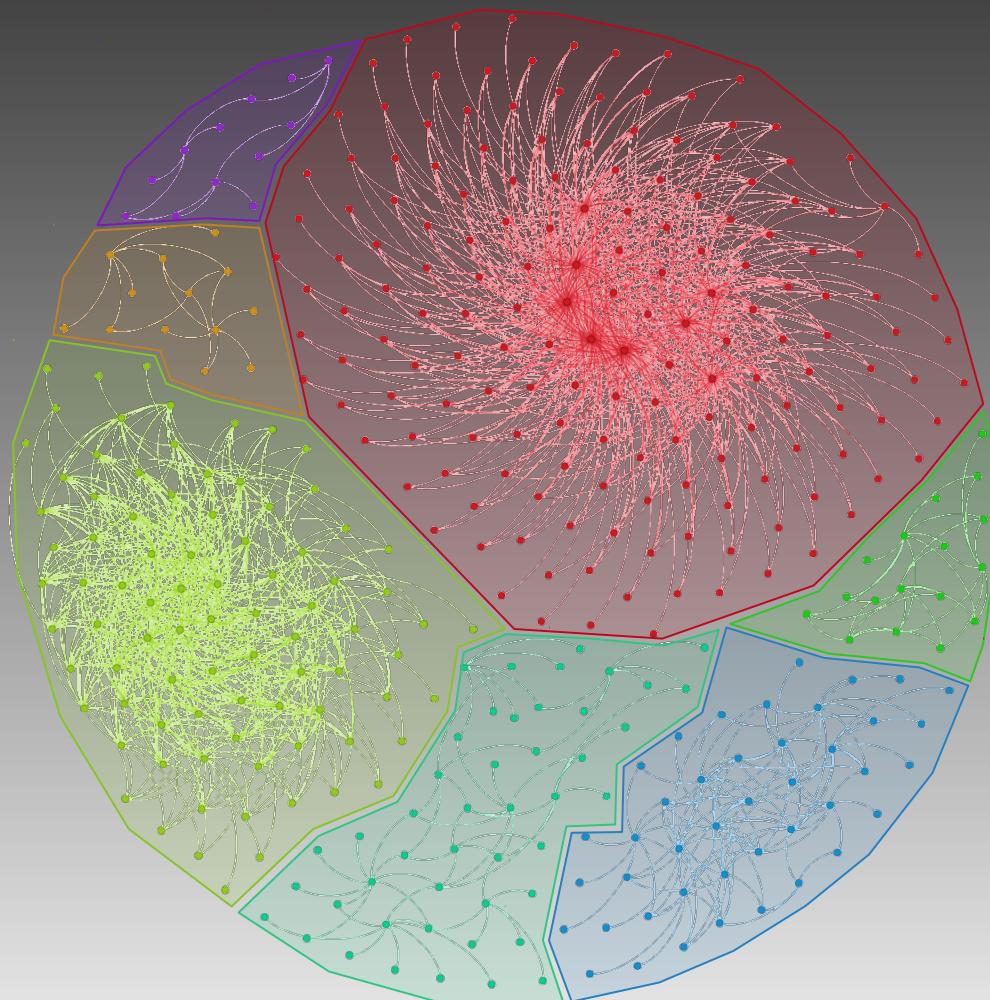
- genus *Varanus*
- *Varanus gouldii*
- *Varanus indicus*
- *Varanus prasinus*
- *Varanus salvator*
- *Varanus spiny*
- *Varanus timorensis*



# Our approach

7 distinct morphological descriptions:

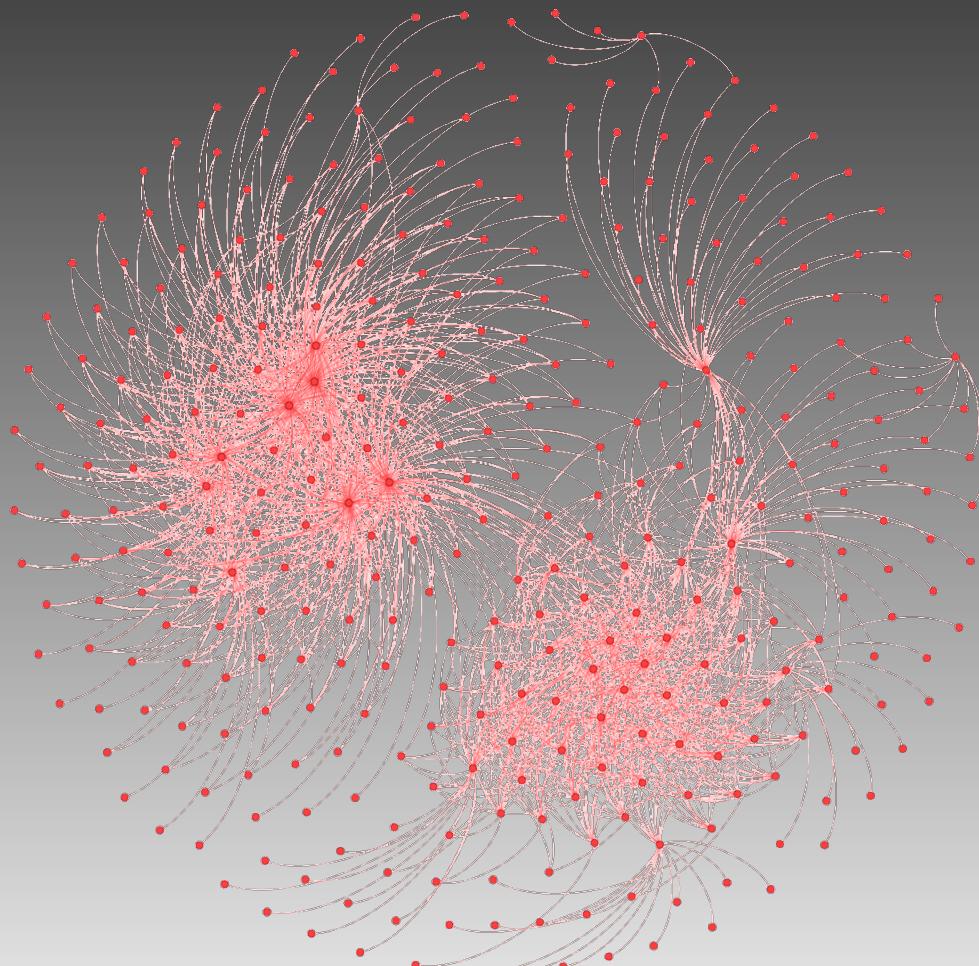
- genus *Varanus*
- *Varanus gouldii*
- *Varanus indicus*
- *Varanus prasinus*
- *Varanus salvator*
- *Varanus spiny*
- *Varanus timorensis*



# Our approach

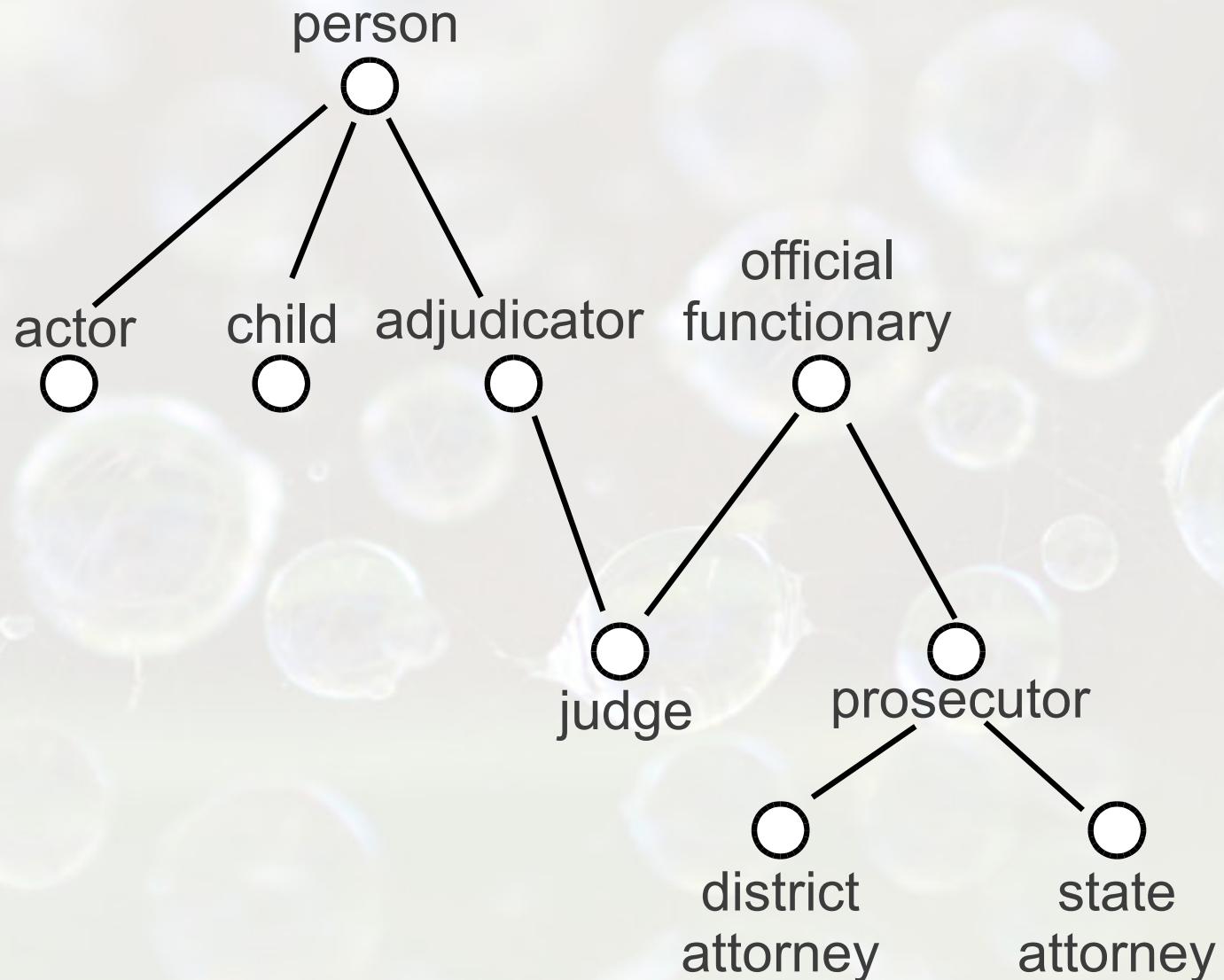
7 distinct morphological descriptions:

- genus *Varanus*
- *Varanus gouldii*
- *Varanus indicus*
- *Varanus prasinus*
- *Varanus salvator*
- *Varanus spiny*
- *Varanus timorensis*

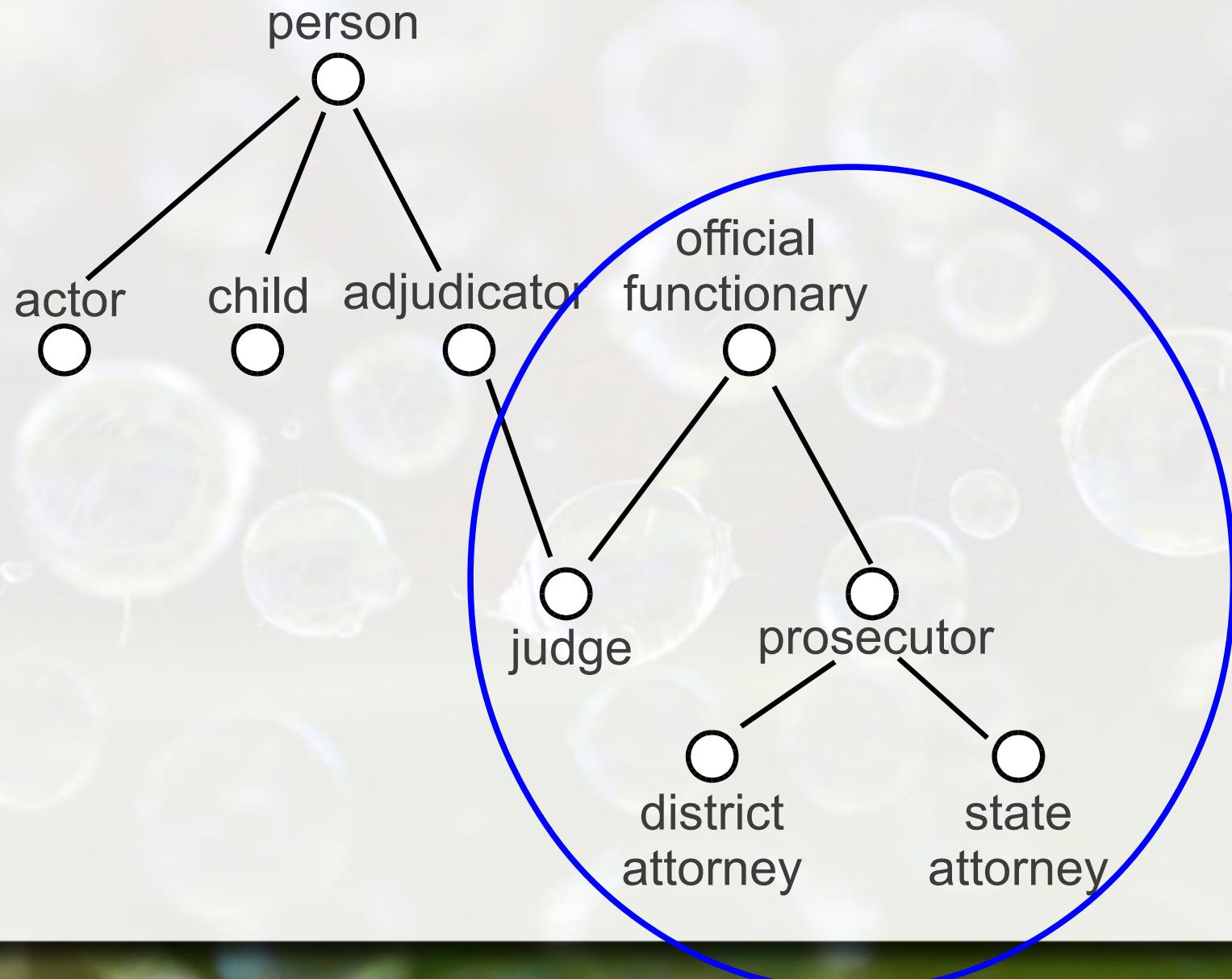


# Evaluating Similarity

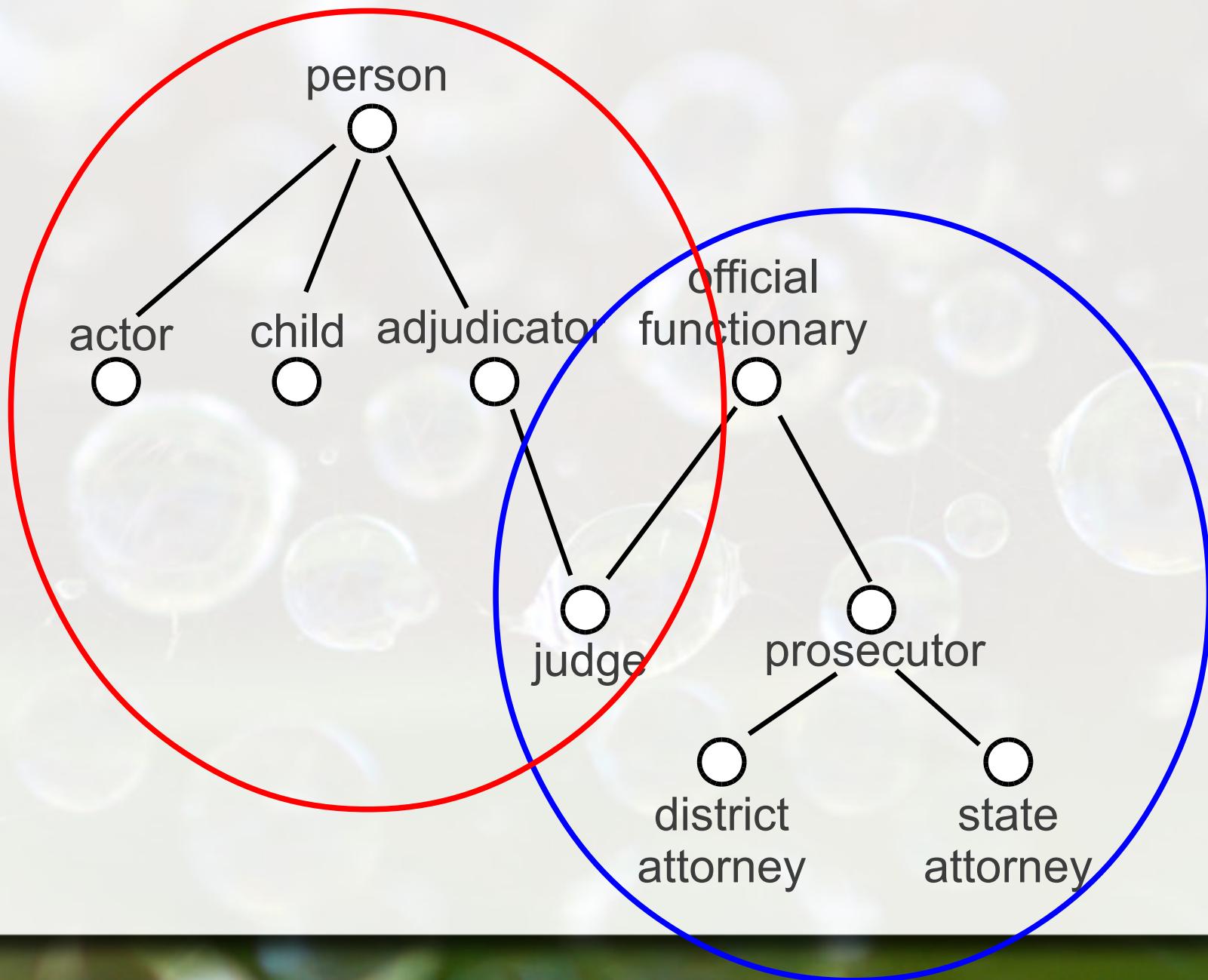
# Size of the Shared Universe



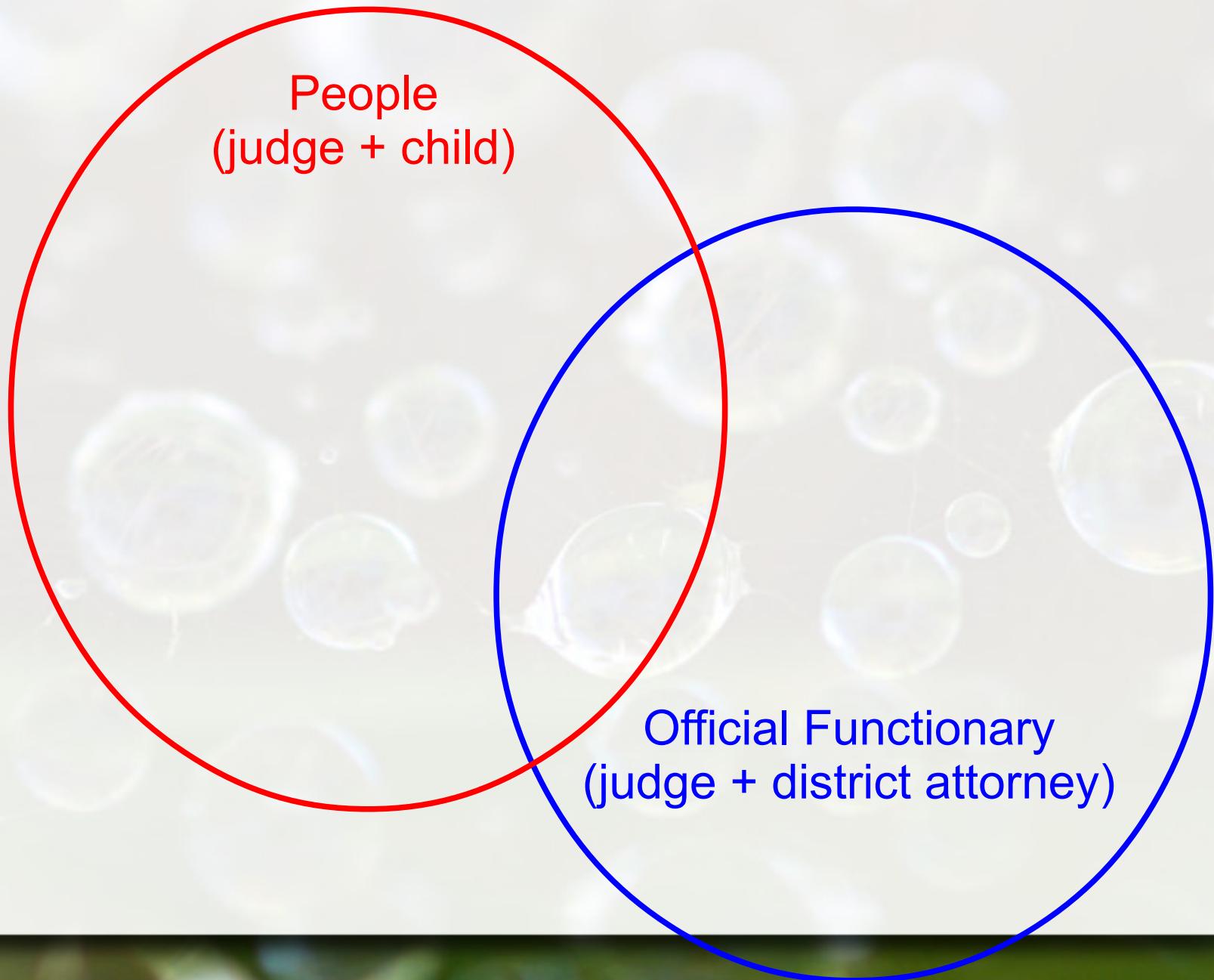
# Size of the Shared Universe



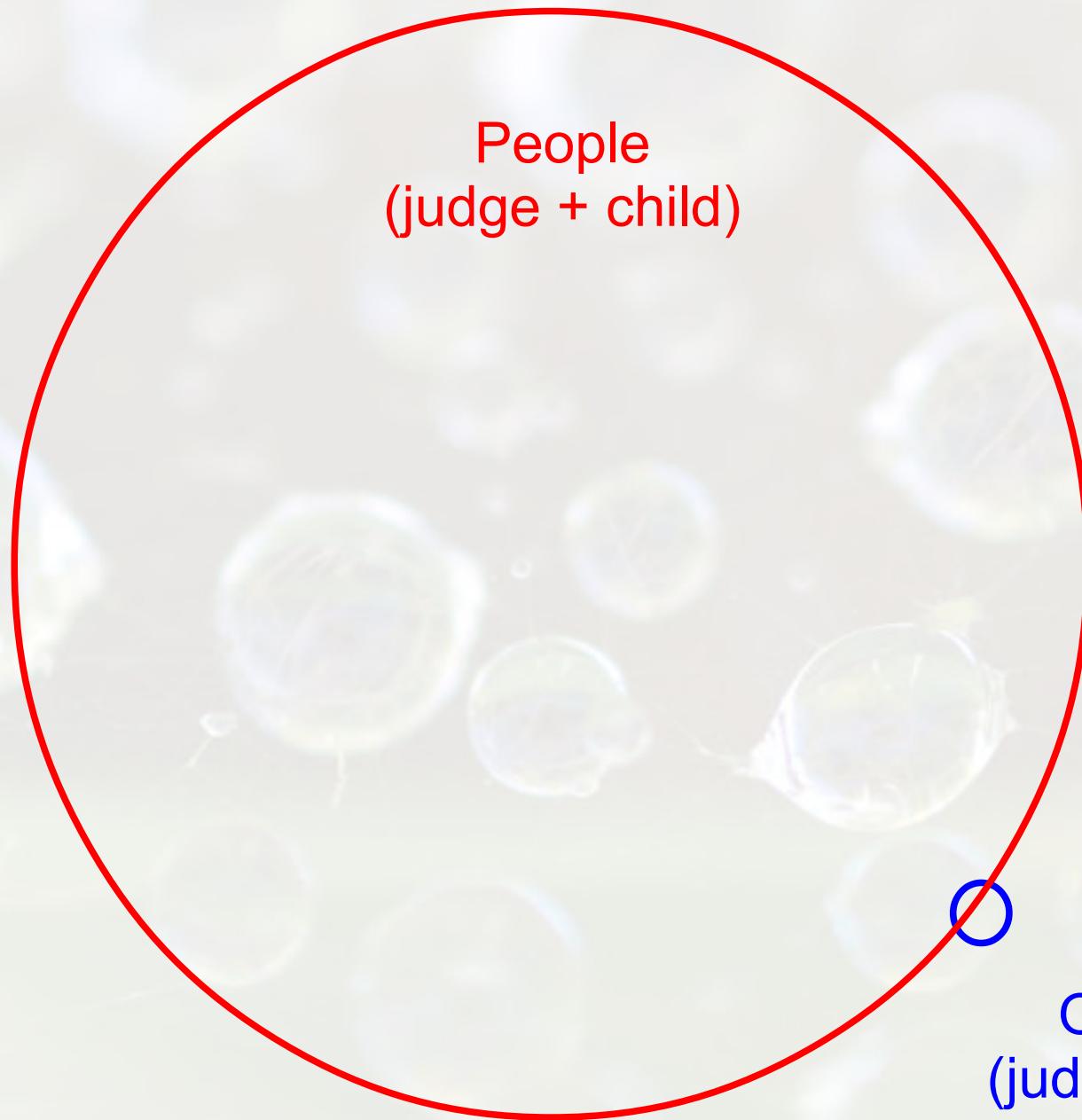
# Size of the Shared Universe



# Size of the Shared Universe

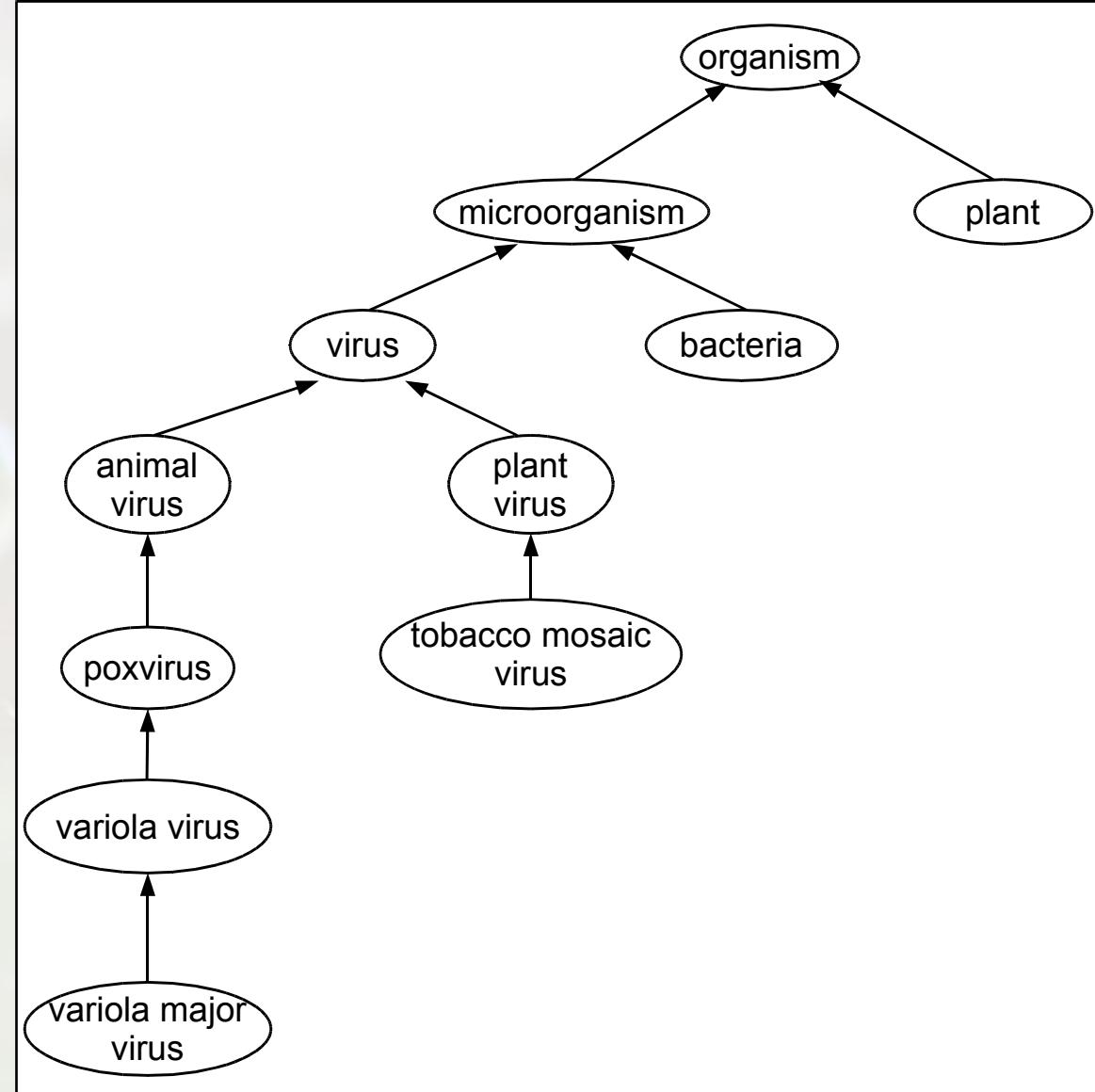


# Size of the Shared Universe



# Similarity

- What is more similar of a tobacco mosaic virus?
  - a) variola major virus
  - b) plant



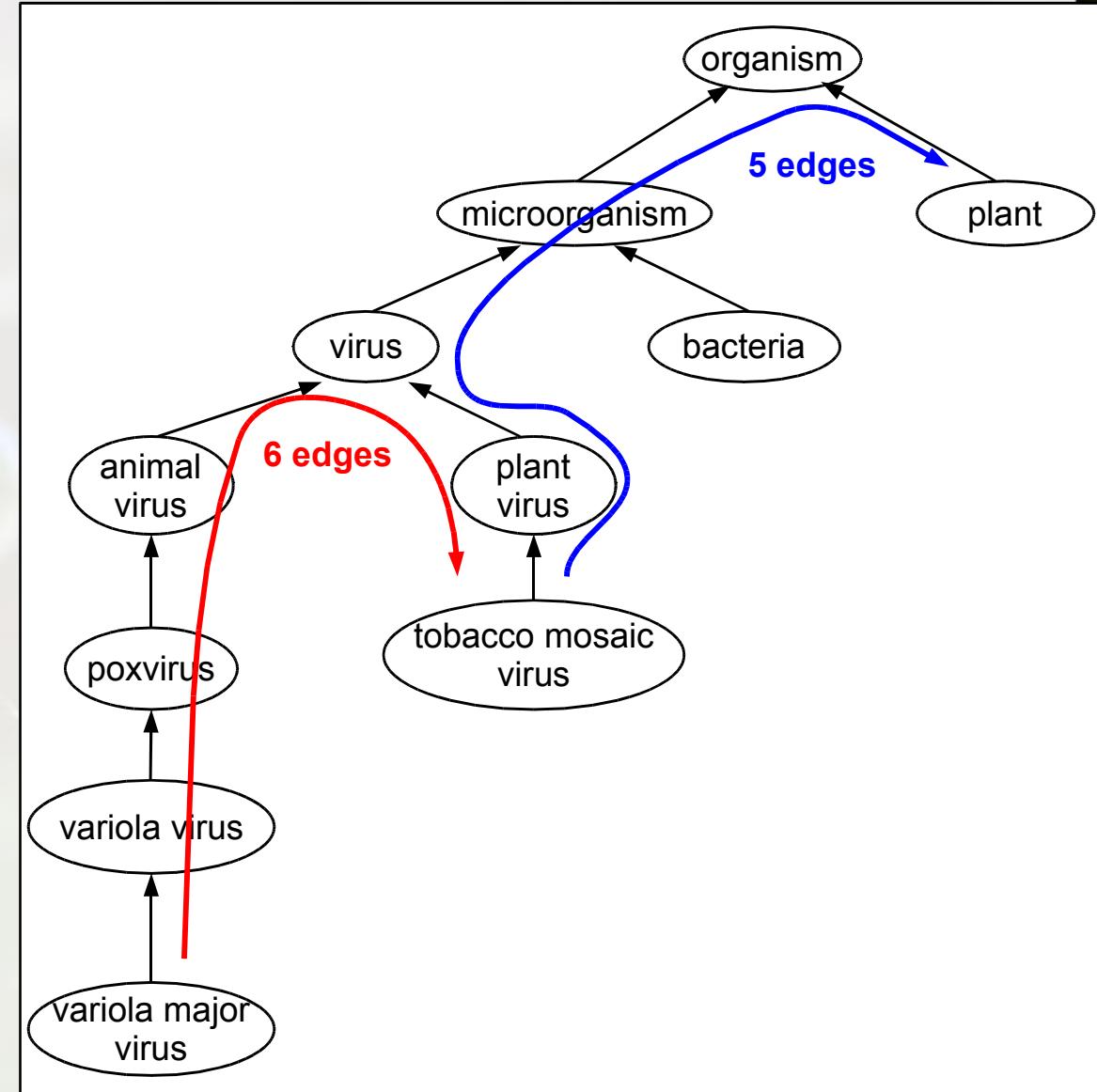
# Similarity

## Minimum path length

- What is more similar of a tobacco mosaic virus?

a) variola major  
virus

b) plant (?)

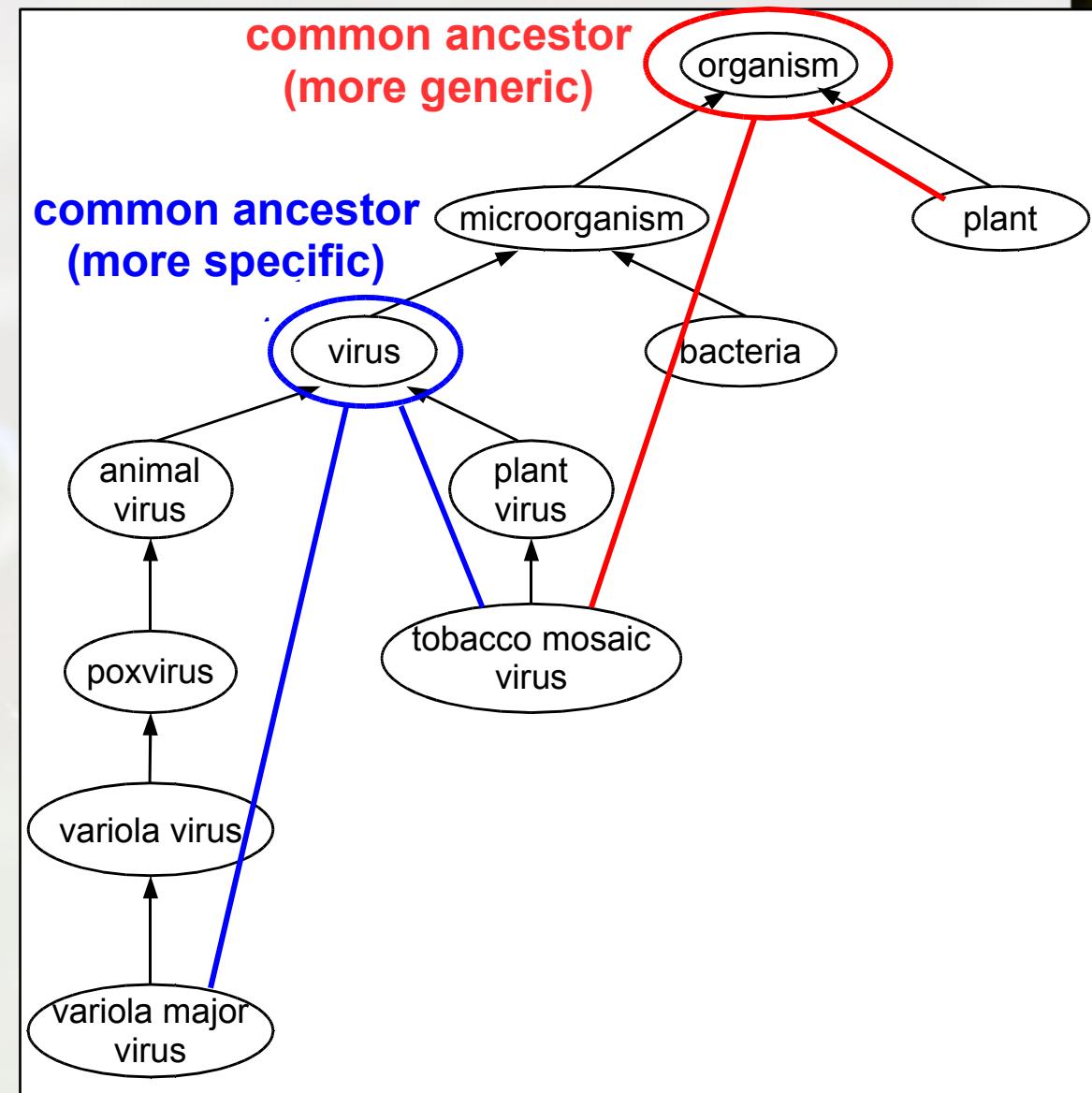


# Similarity

## More specialized common ancestor

- What is more similar of a tobacco mosaic virus?

- a) variola major virus
- b) plant

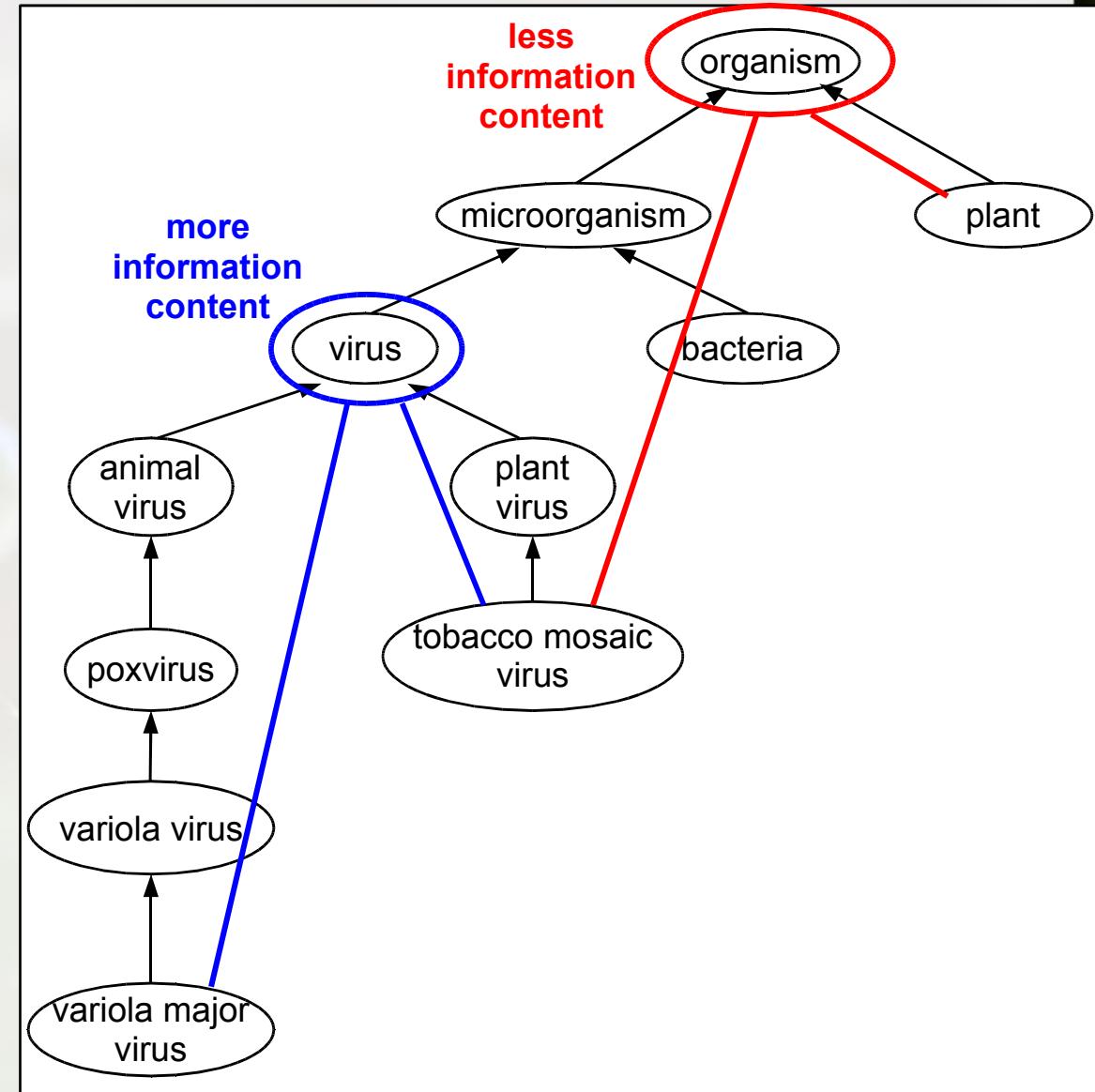


# Similarity

## Maximum information content

- What is more similar of a tobacco mosaic virus?

- a) variola major virus
- b) plant

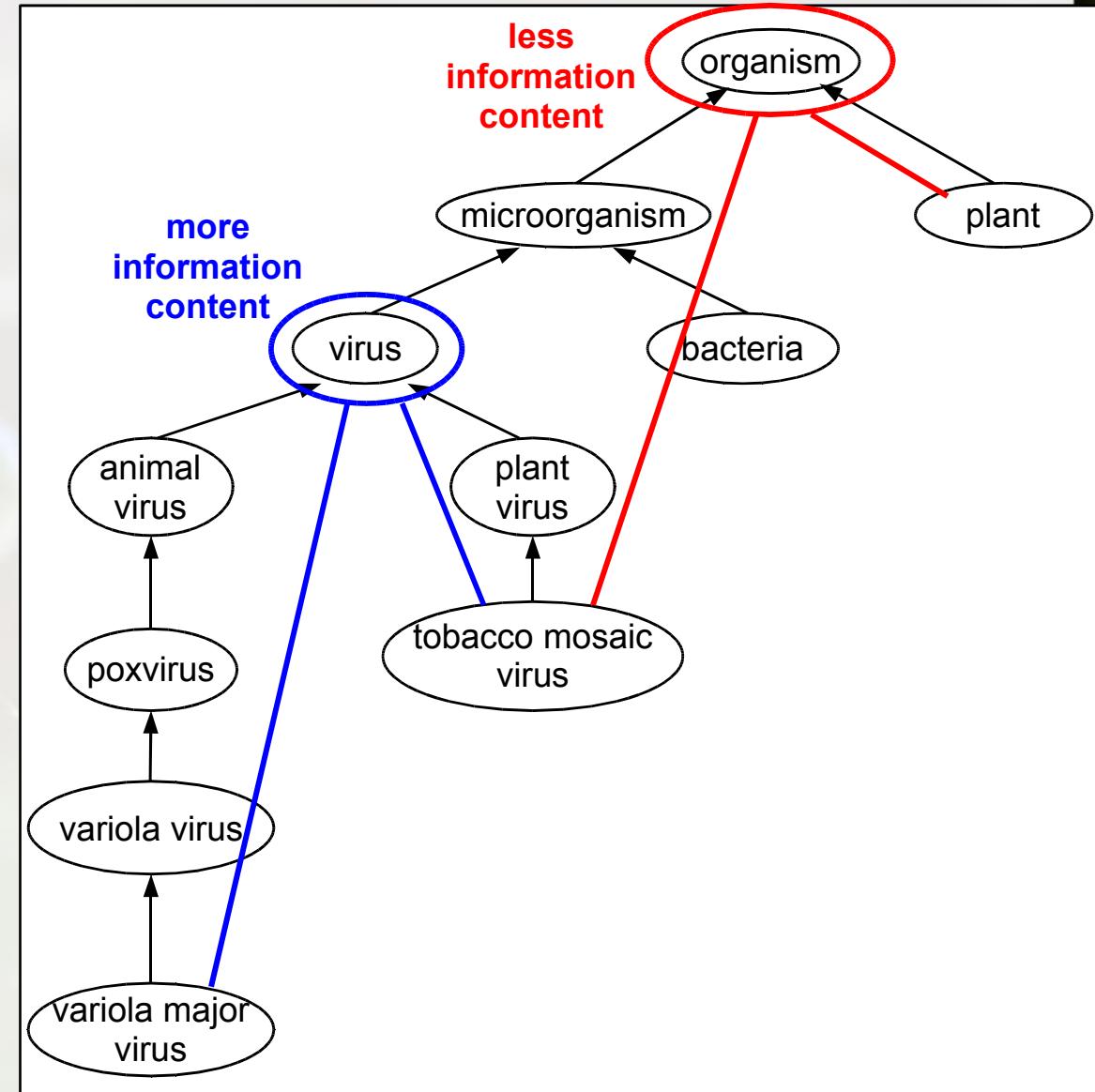


# Similarity

## Maximum information content

- What is more similar of a tobacco mosaic virus?

- a) variola major virus
- b) plant



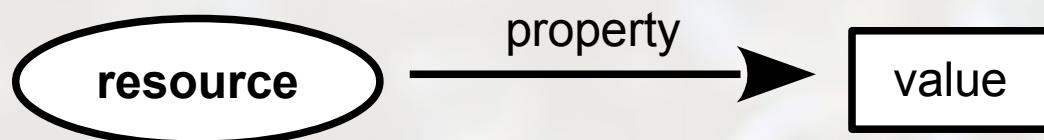
# Graph Databases

# Graph Model

# Graph Model RDF Graph

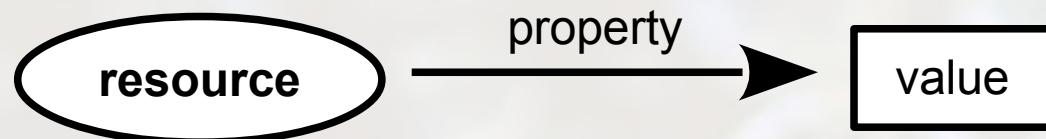
# RDF Graph

- Triple model: resource, property, value

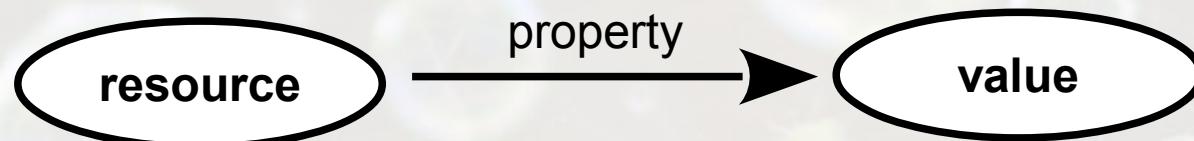


# RDF Graph

- Triple model: resource, property, value

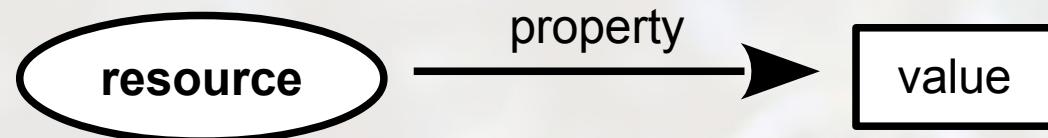


- The value can be another resource

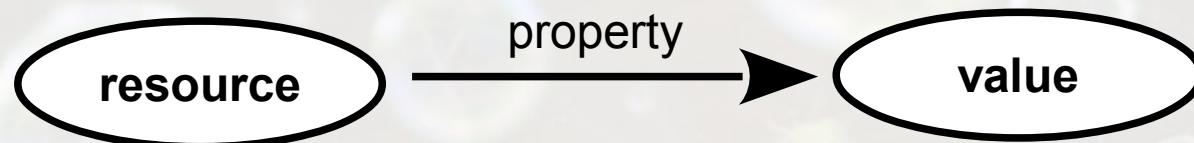


# RDF Graph

- Triple model: resource, property, value



- The value can be another resource



- ... described by a property,value

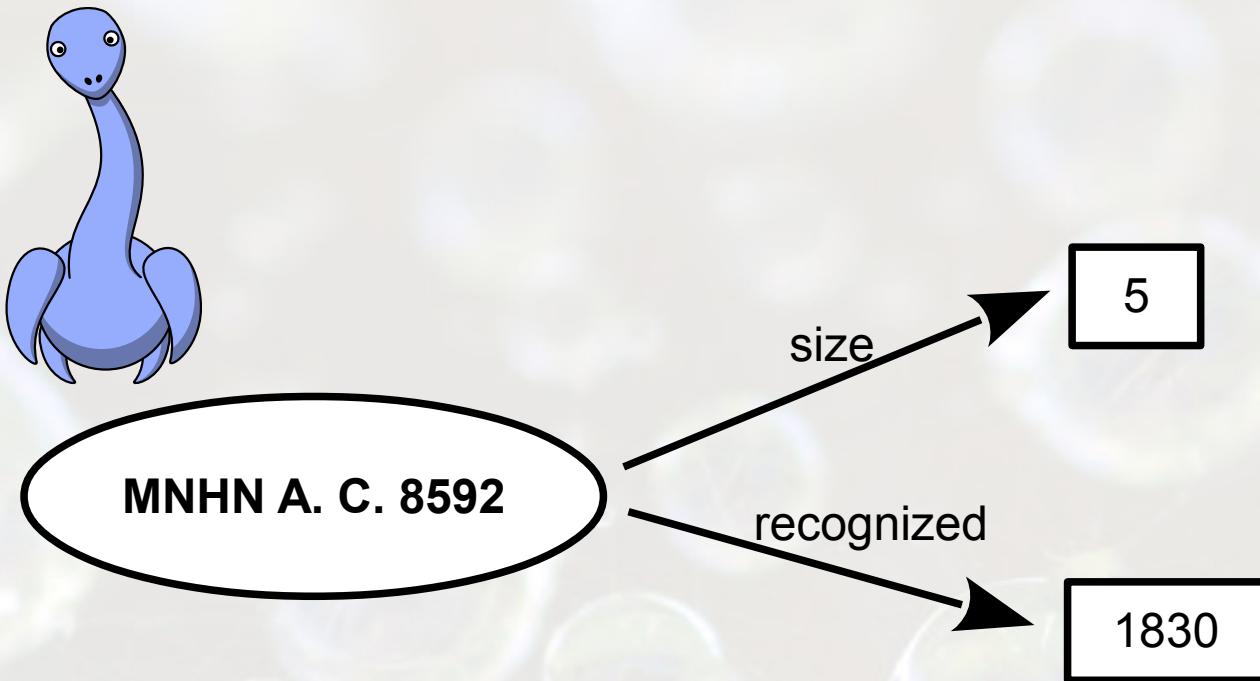


# Plesiosaurus dolichodeirus in RDF



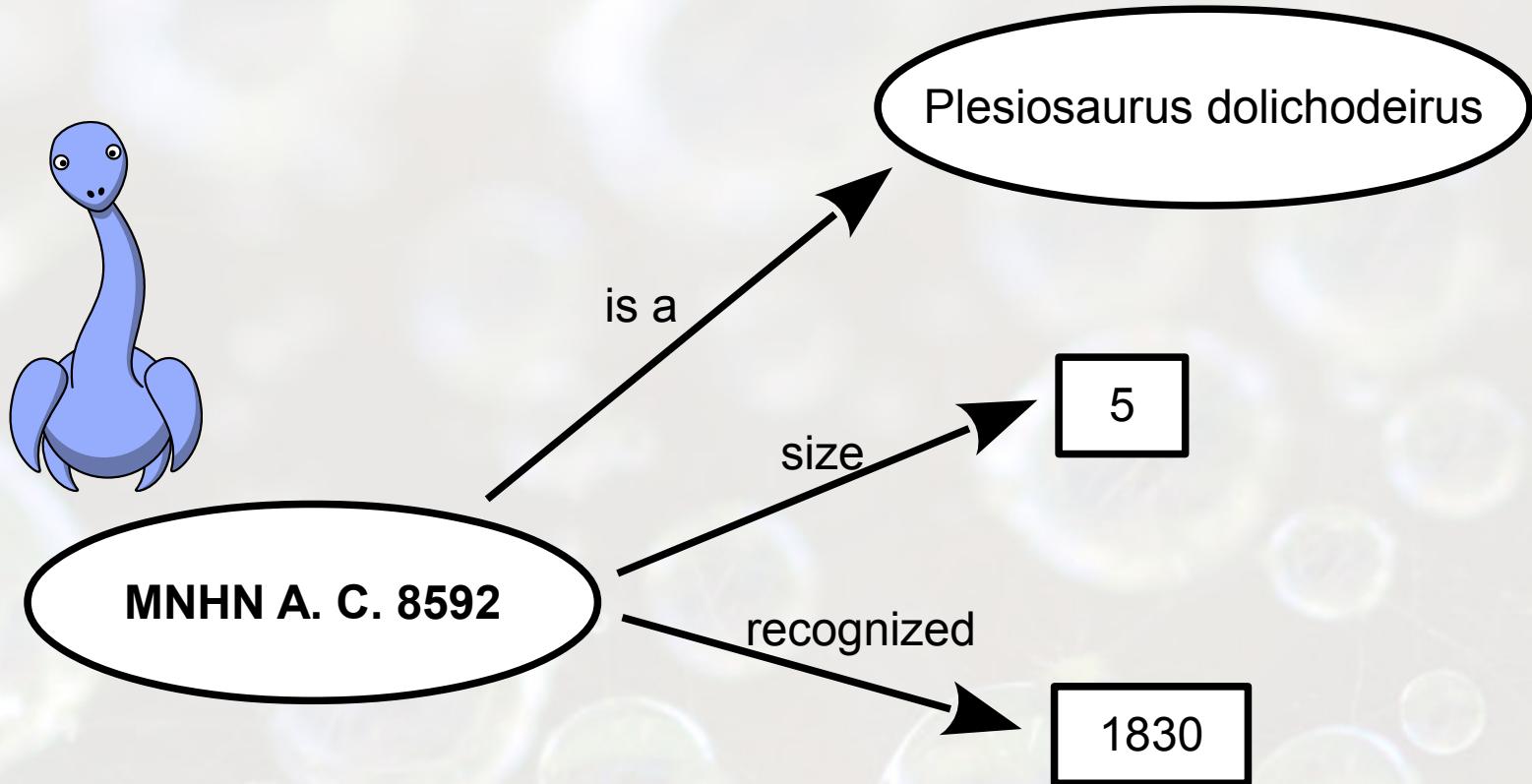
- Triple: <MNHN A. C. 8592>, size, 5

# Plesiosaurus dolichodeirus in RDF



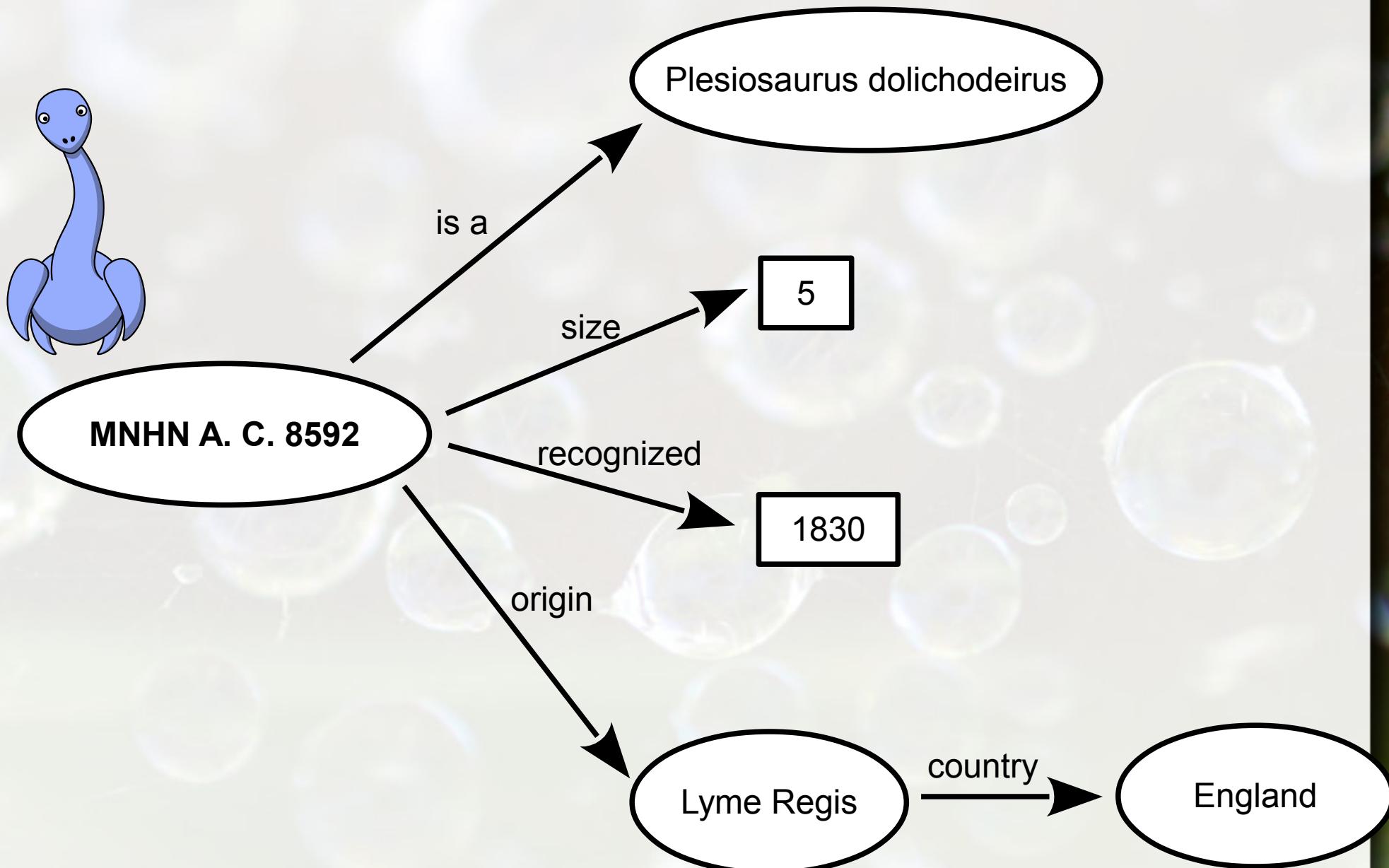
- Triple: <MNHN A. C. 8592>, size, 5
- Triple: <MNHN A. C. 8592>, recognized, 1830

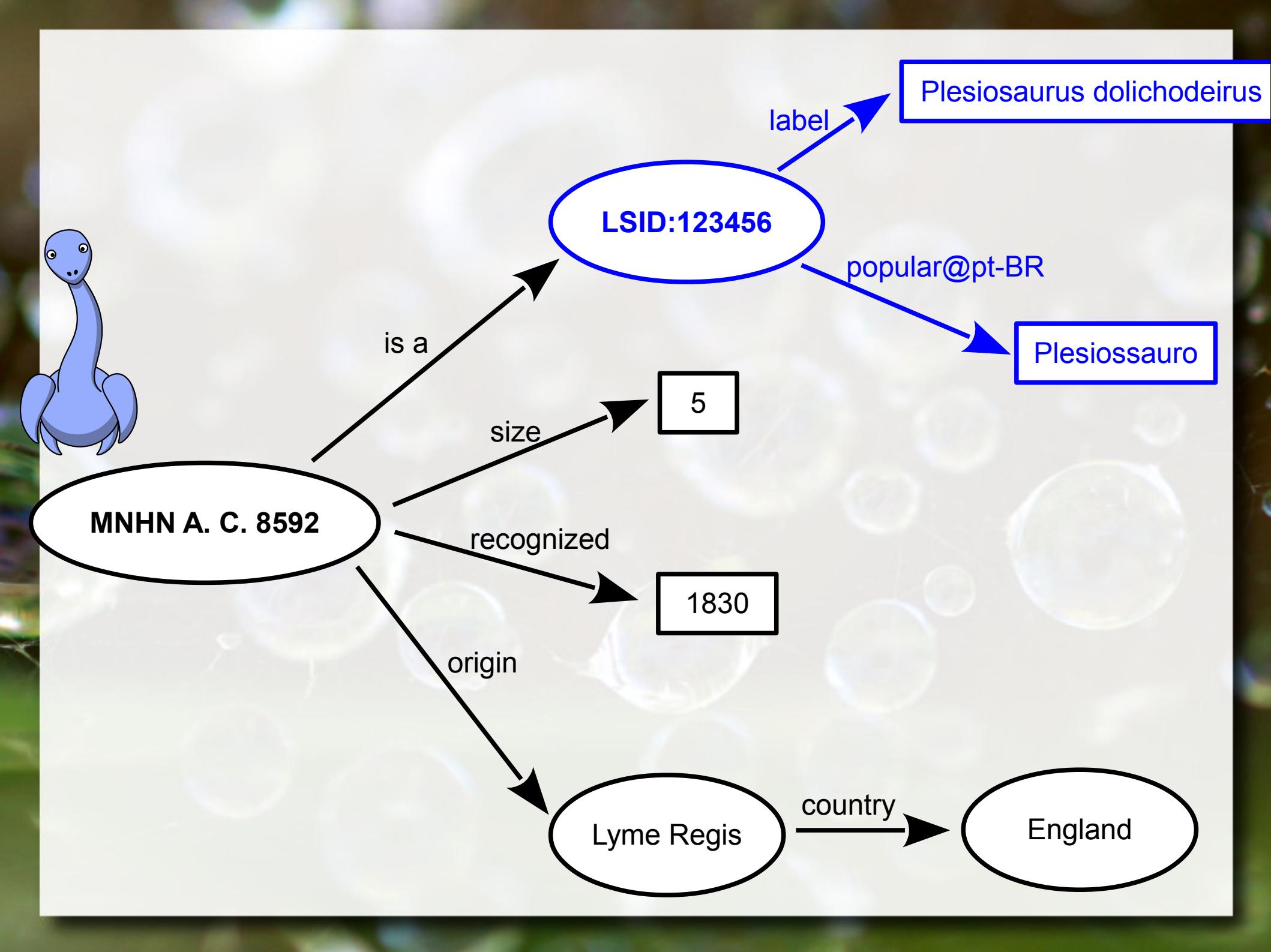
# Plesiosaurus dolichodeirus in RDF



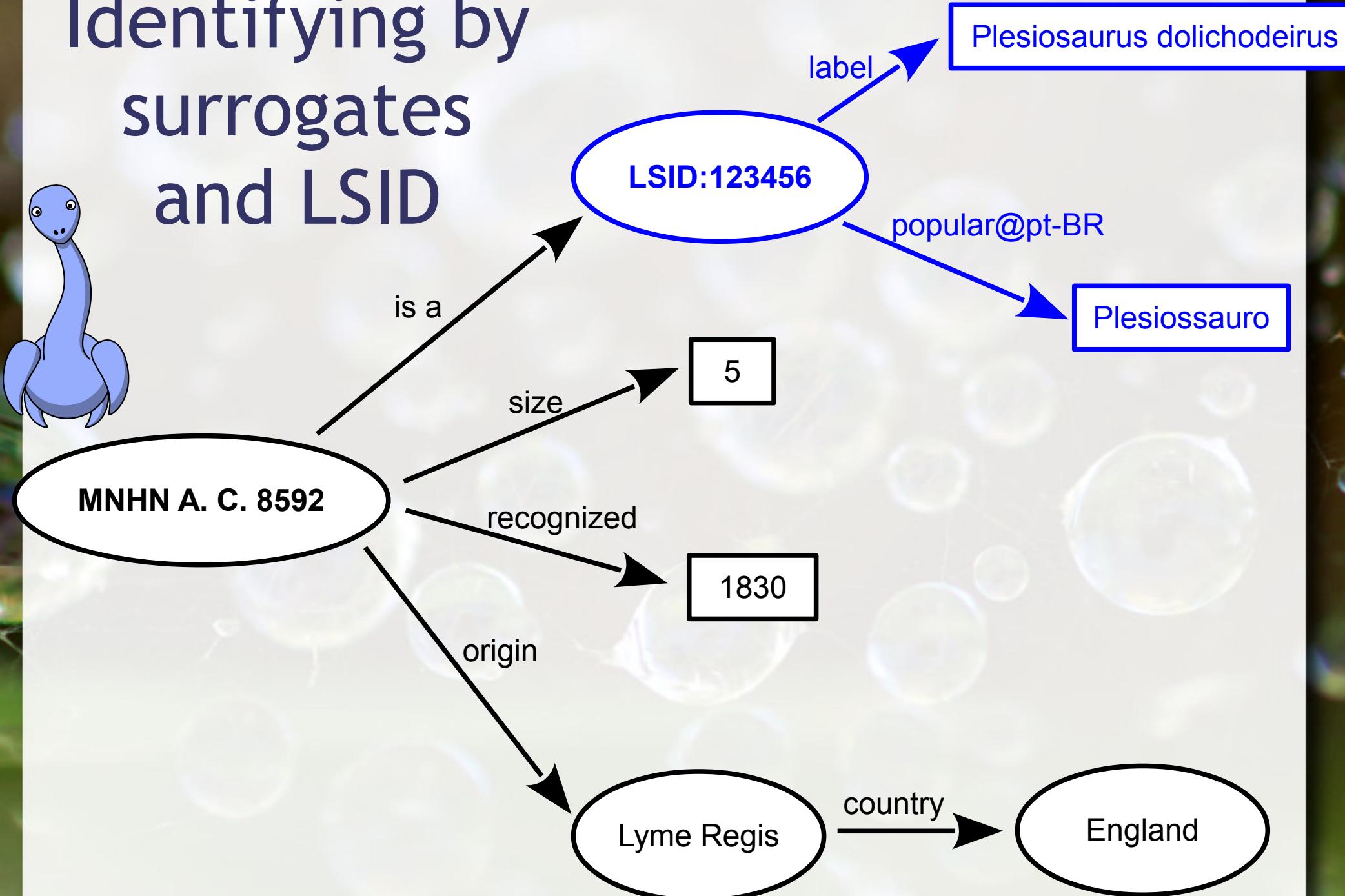
- <MHN A. C. 8592>, is\_a, <Plesiosaurus dolichodeirus>

# Plesiosaurus dolichodeirus in RDF

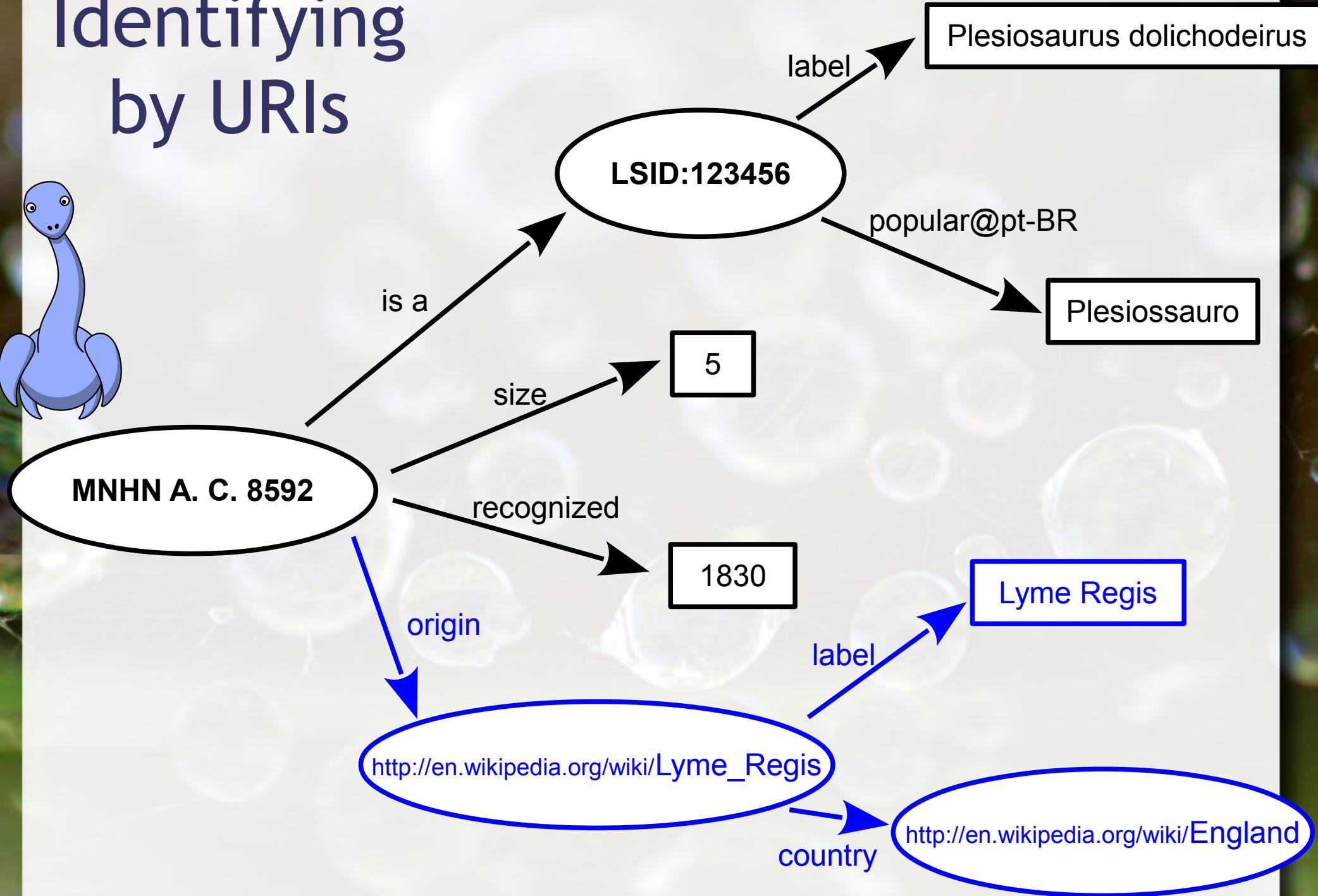
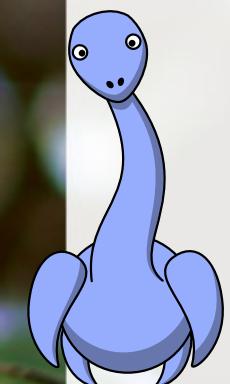




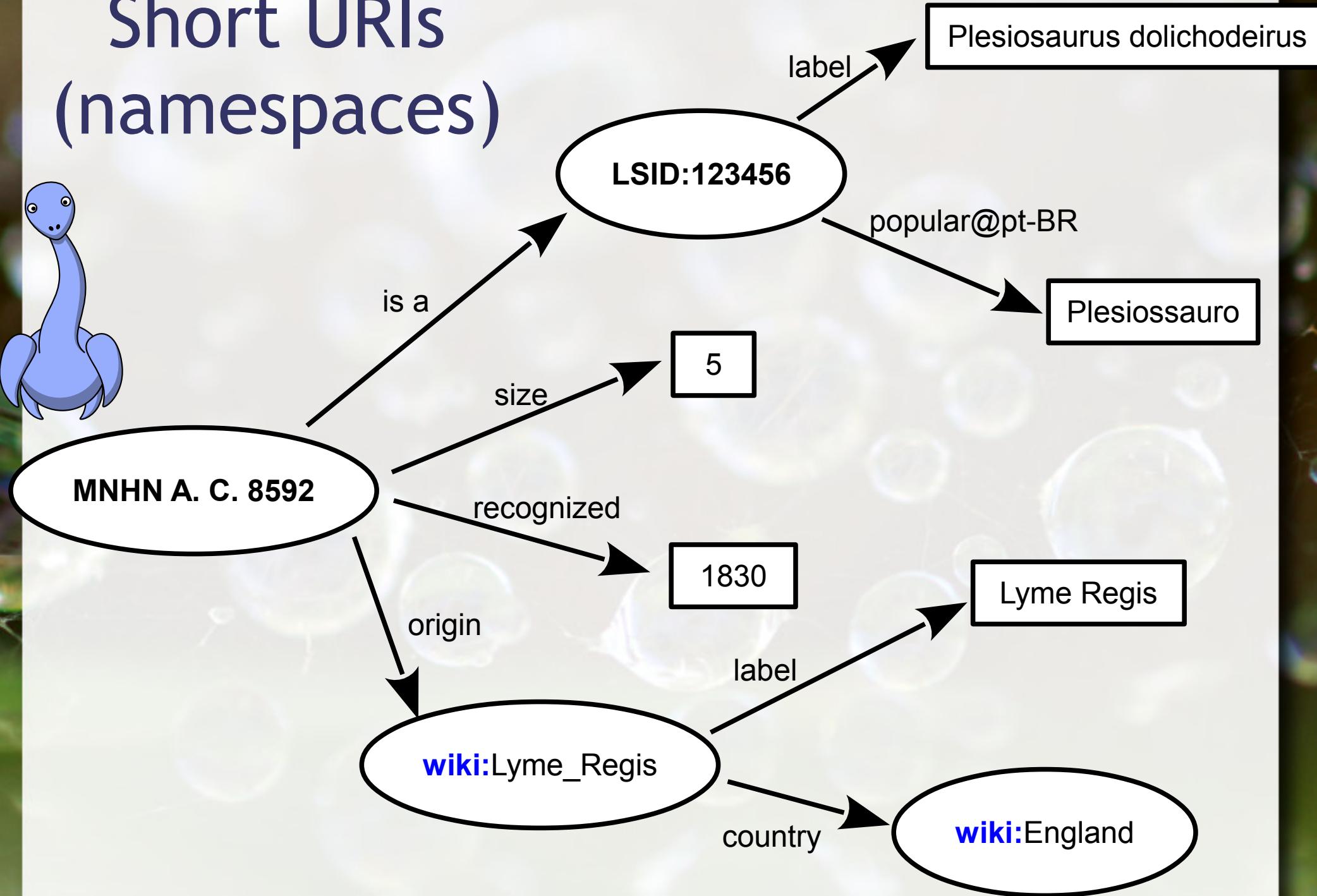
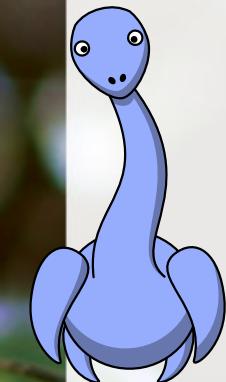
# Identifying by surrogates and LSID



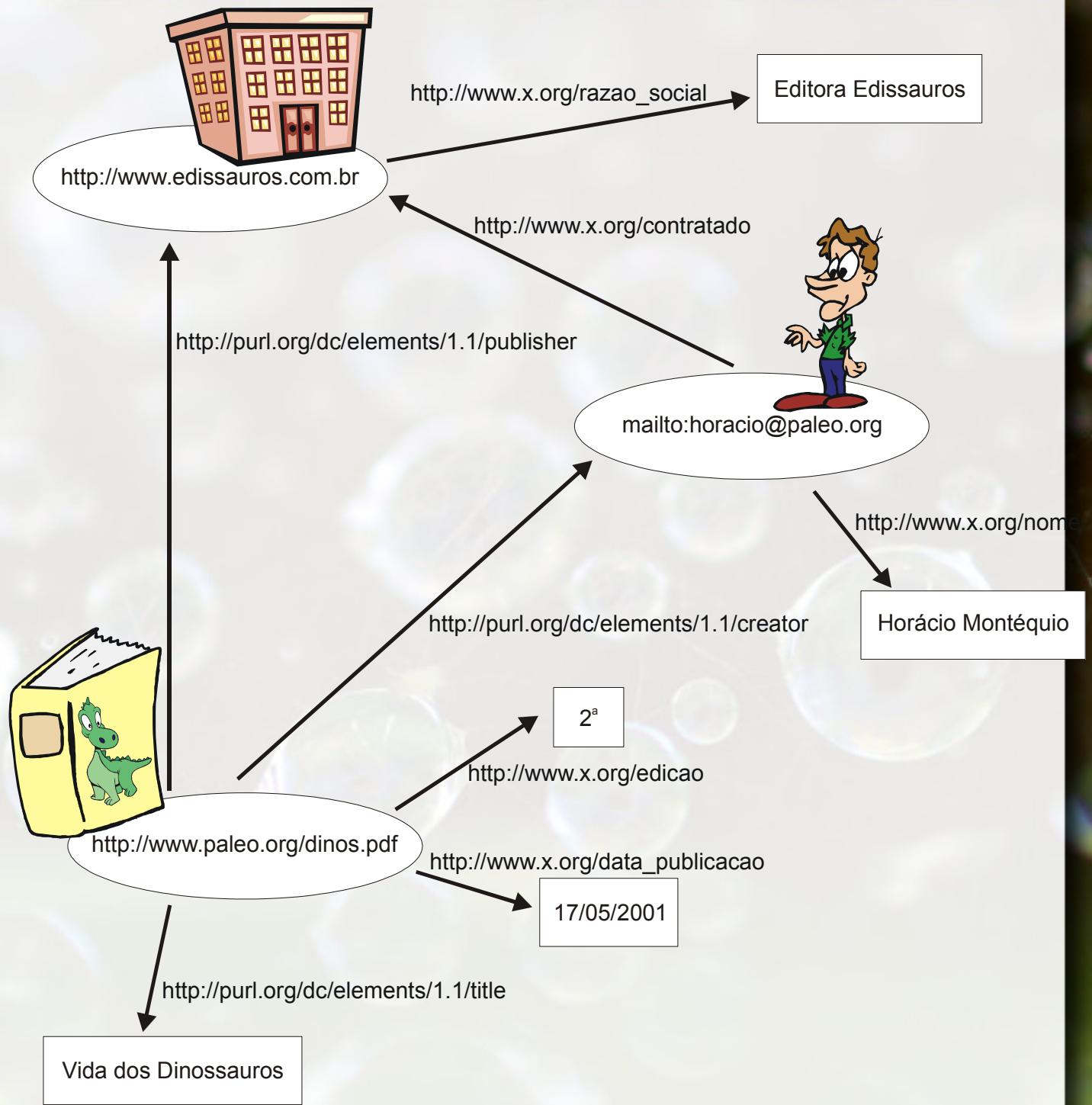
# Identifying by URIs



# Short URIs (namespaces)

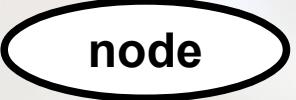


# RDF Graph

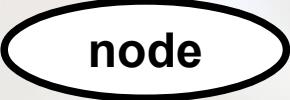


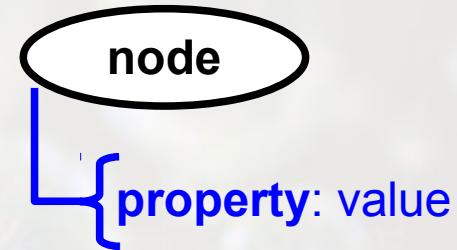
# Graph Model Property Graph

# Property Graph

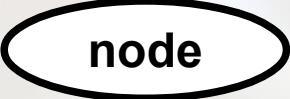
- (node) 

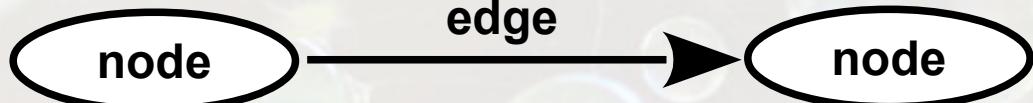
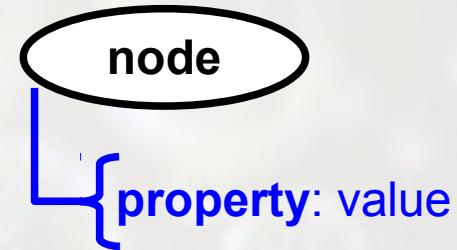
# Property Graph

- (node) 
  - Nodes can have properties  
`(node { property: value })`



# Property Graph

- (node)
  - Nodes can have properties  
`(node { property: value })`
- (node)-[edge]->(node)

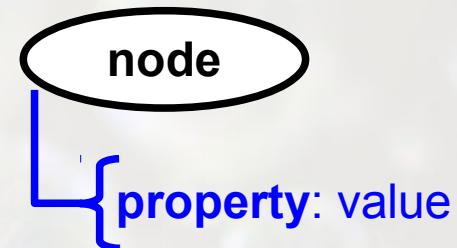


# Property Graph

- (node)



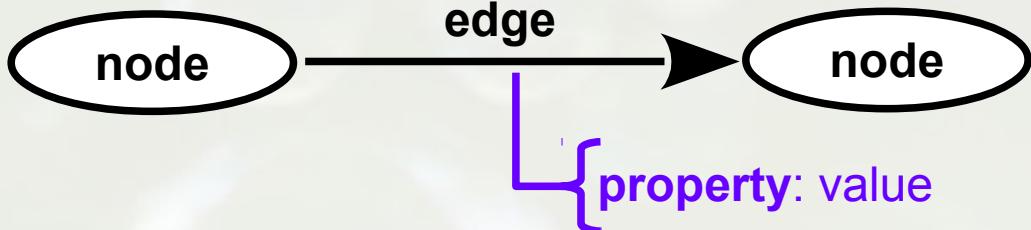
- Nodes can have properties  
`(node { property: value })`



- (node)-[edge]->(node)



- Edges can have properties  
`(node)-[edge {property:value}]->(node)`



# Plesiosaurus in a Property Graph



MNHN A. C. 8592

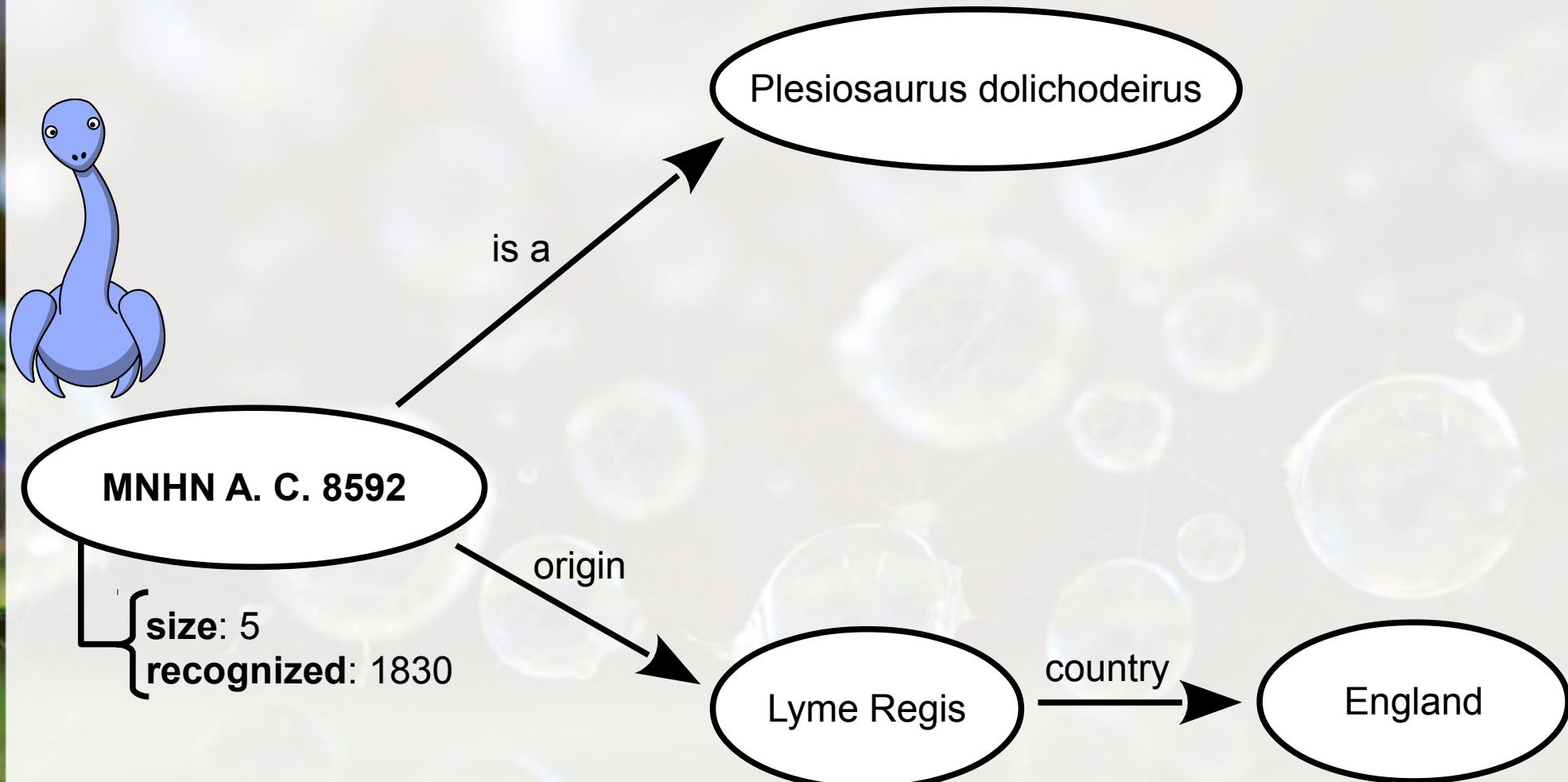
# Plesiosaurus in a Property Graph



MNHN A. C. 8592

{  
size: 5  
recognized: 1830

# Plesiosaurus in a Property Graph



# LSID Surrogate



MNHN A. C. 8592

{  
size: 5  
recognized: 1830

is a

LSID:123456

{  
label: Plesiosaurus dolichodeirus  
popular@pt-BR: Plesiossauro

origin

Lyme Regis

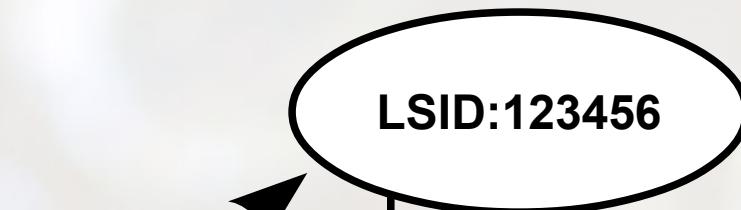
country →

England

# URI Identifiers



{  
size: 5  
recognized: 1830



{  
label: Plesiosaurus dolichodeirus  
popular@pt-BR: Plesiossauro

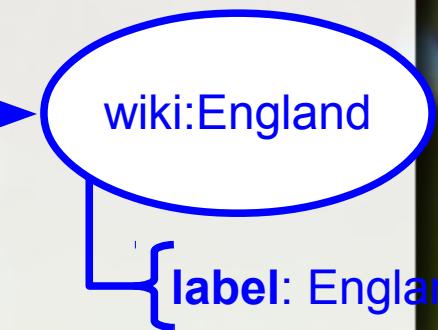
is a

origin



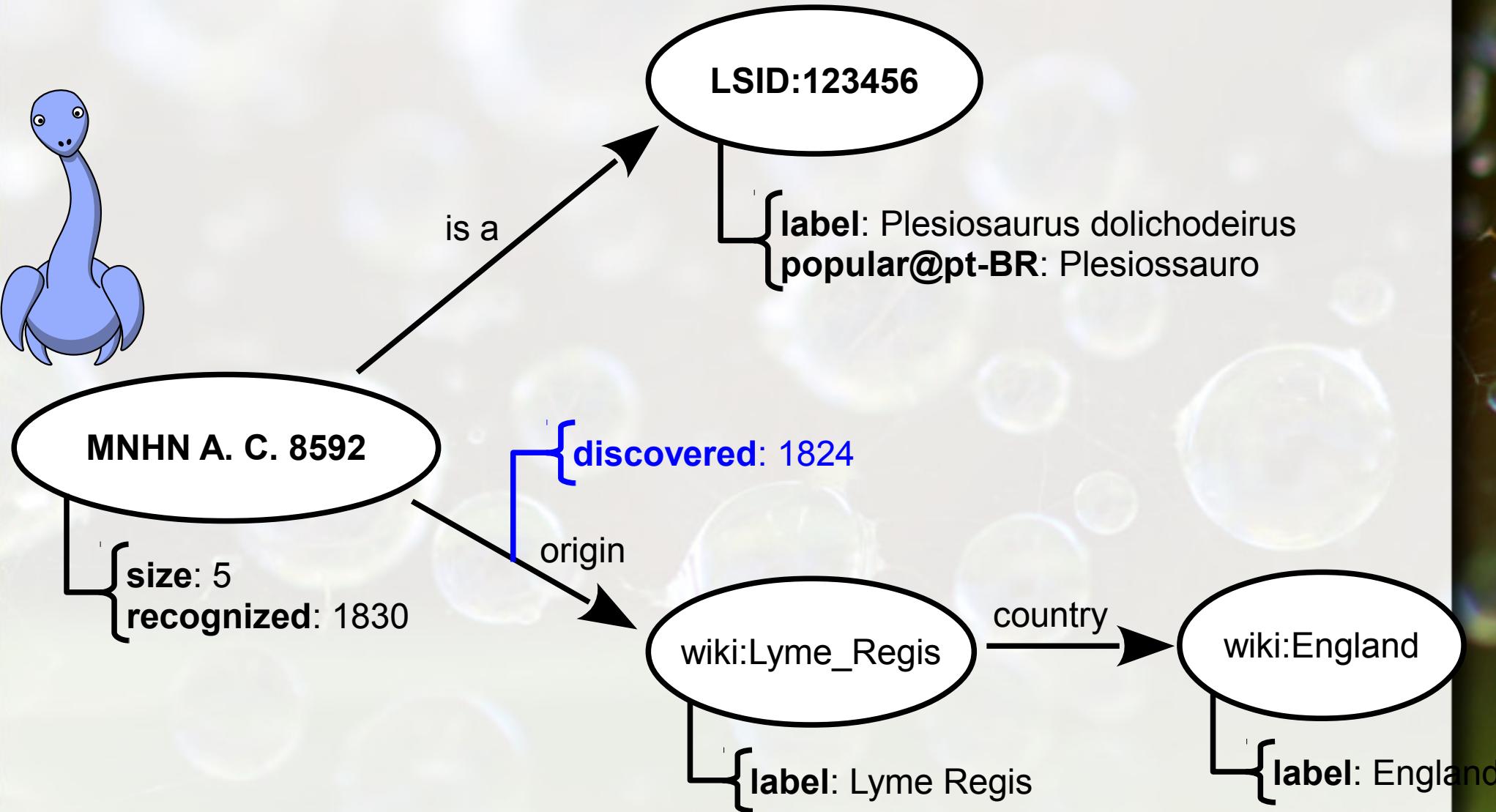
{  
label: Lyme Regis

country

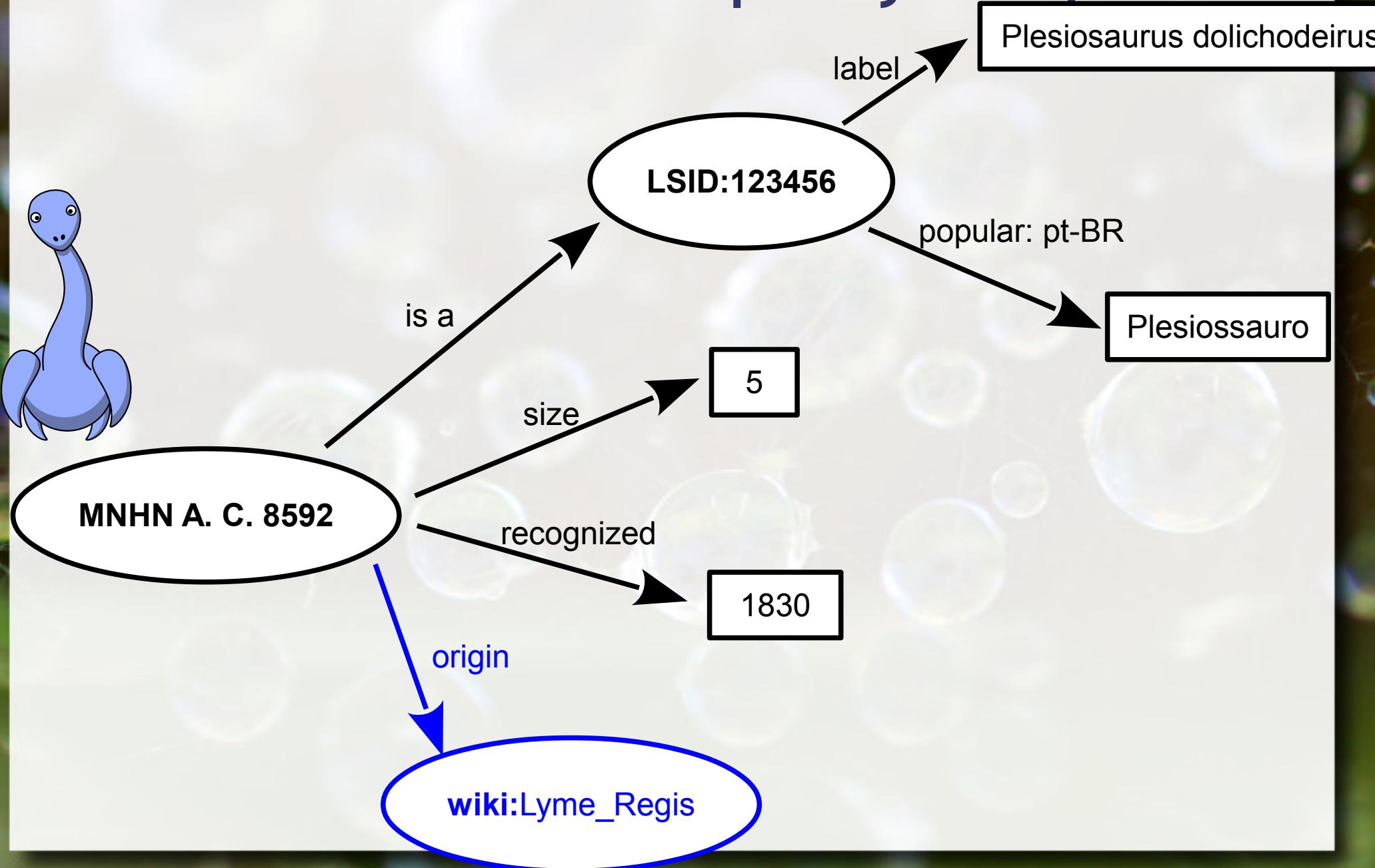


{  
label: England

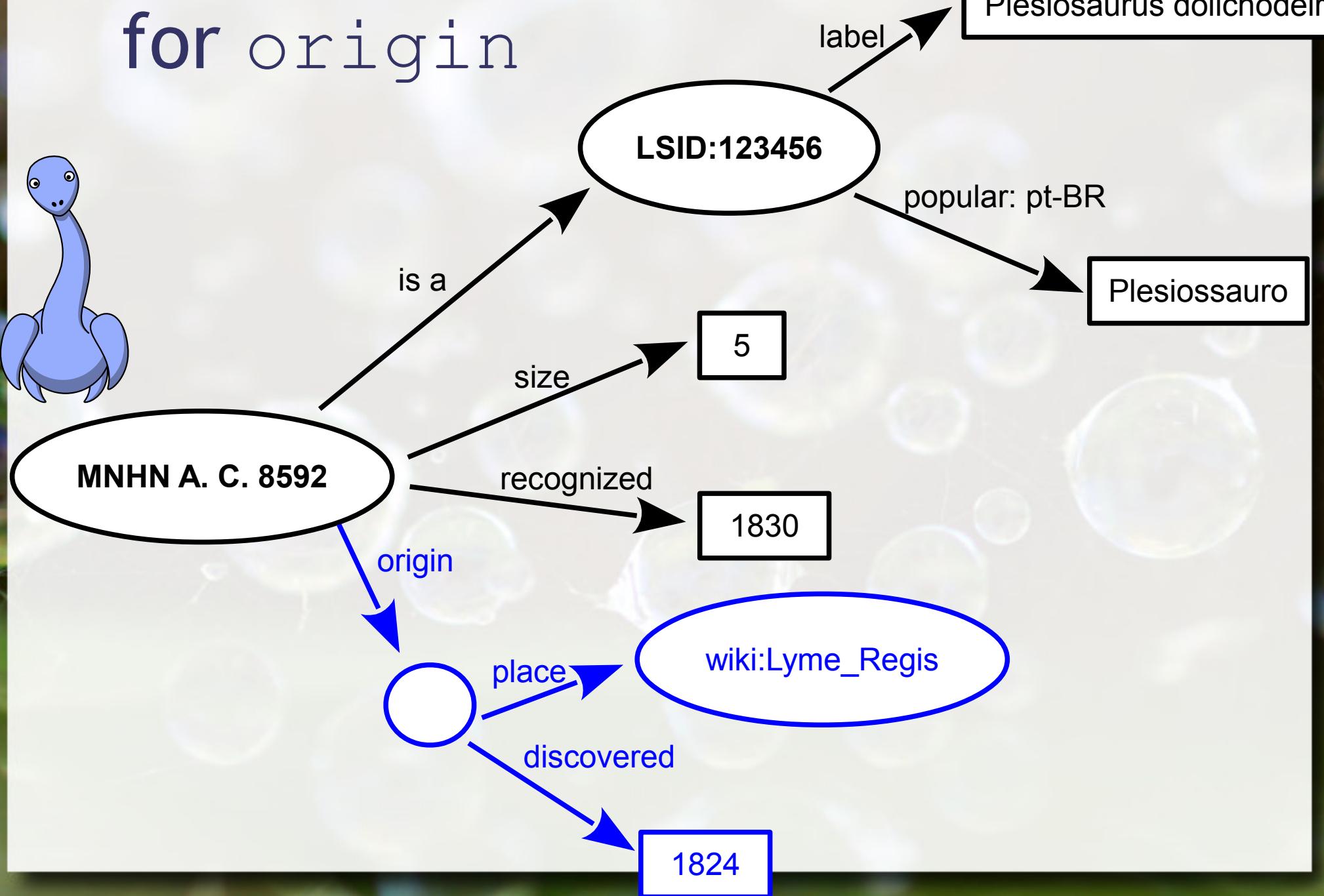
# Property in the origin edge



# Back to the Property Graph



# Two properties for origin



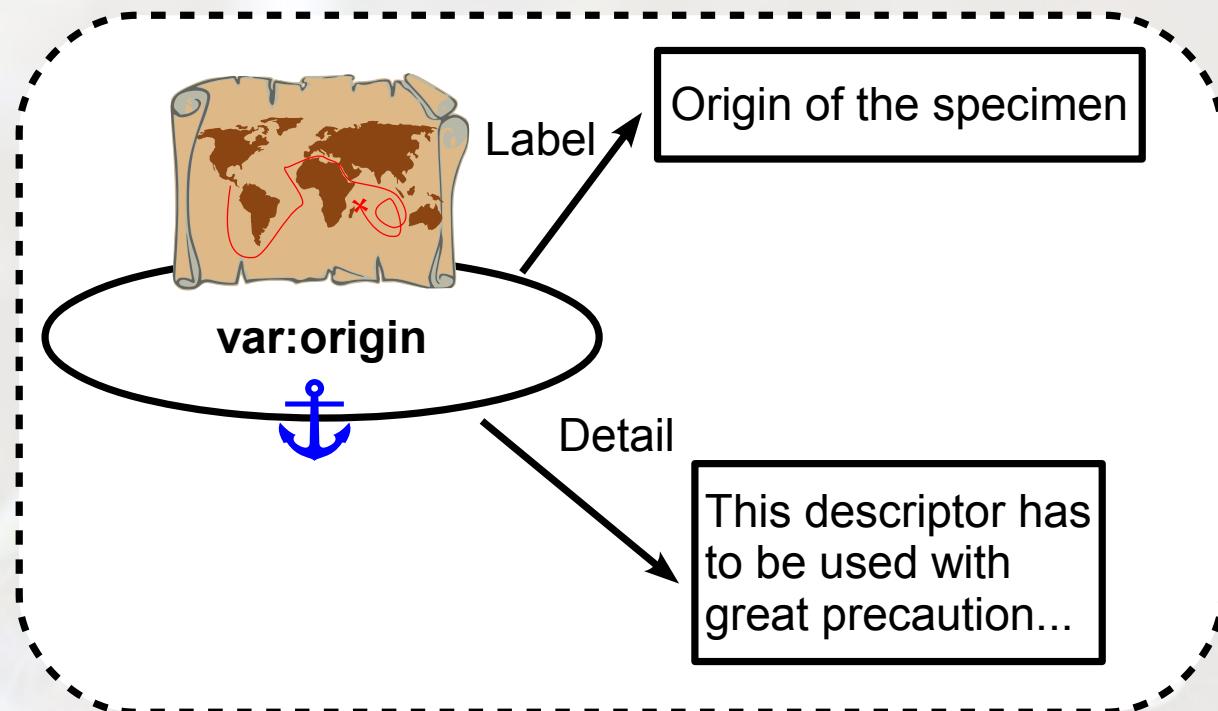
# References

- GOMES-JR, L. C.; JENSEN, R.; SANTANCHÈ, A. (2013) Towards Query Model Integration: Topology-aware, IR-inspired Metrics for Declarative Graph Querying, II Workshop on Querying Graph Structured Data 2013, Genoa.

# André Santanchè

<http://www.ic.unicamp.br/~santanche>

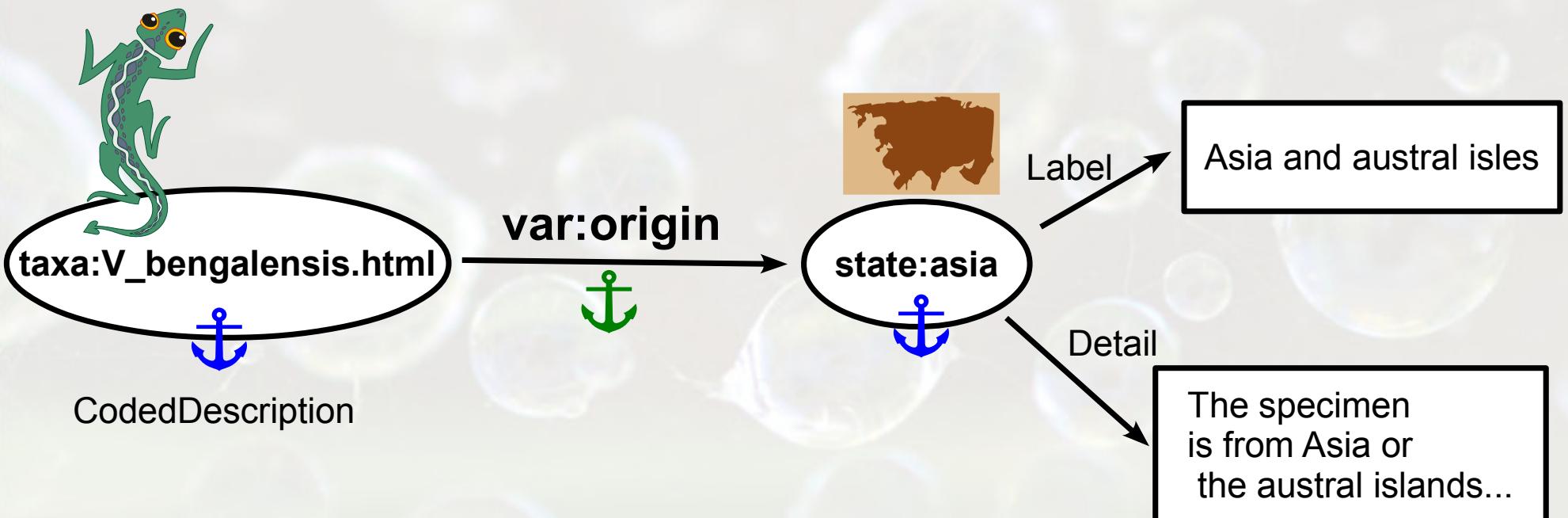
# Property



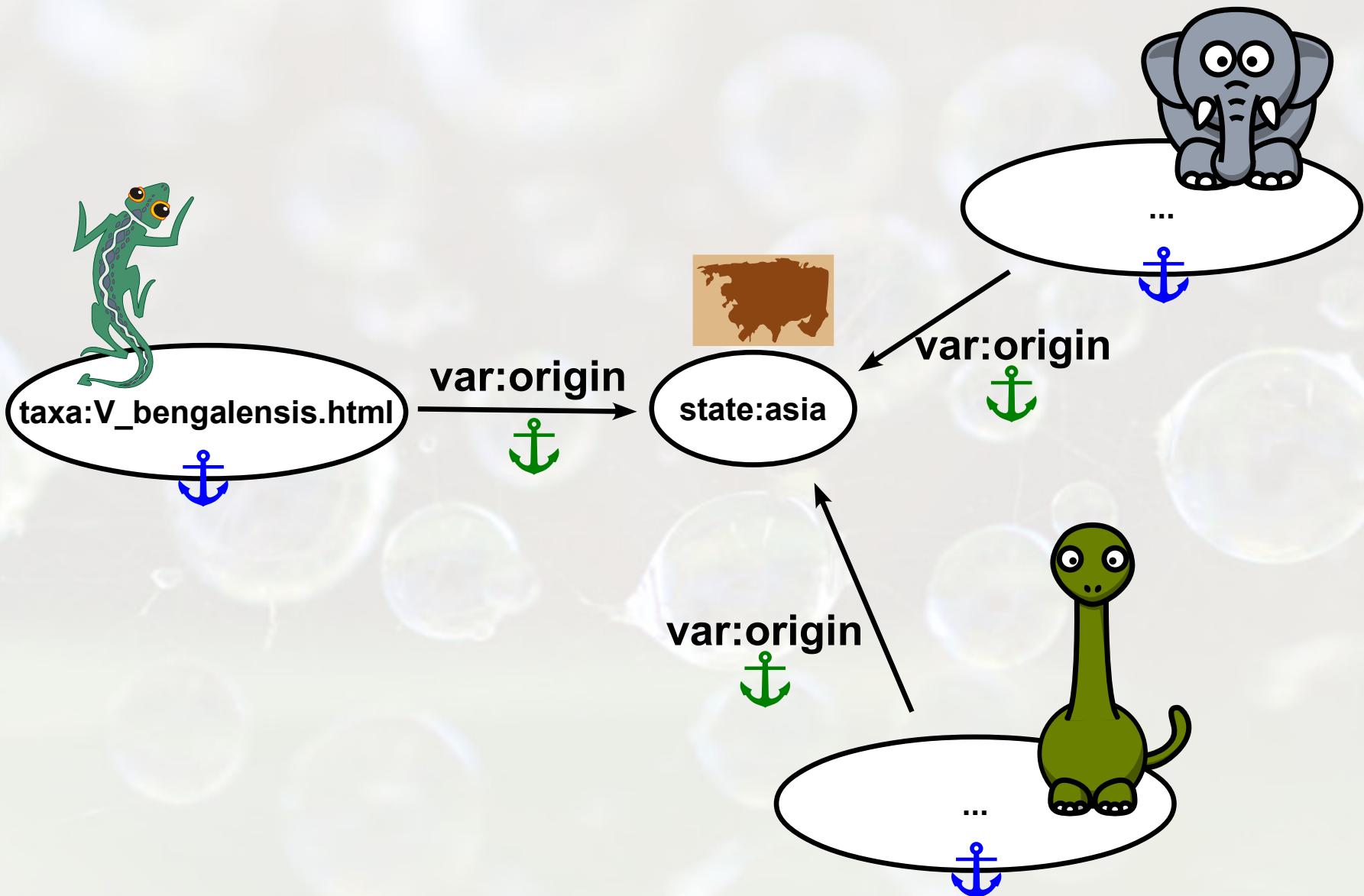
property

**var:origin**  
anchor icon

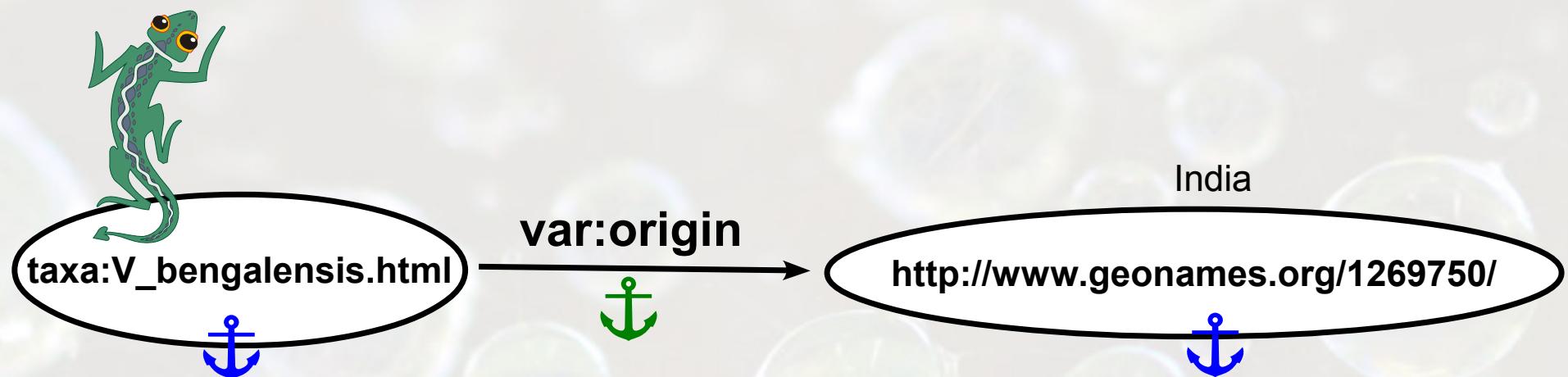
# Using the Property



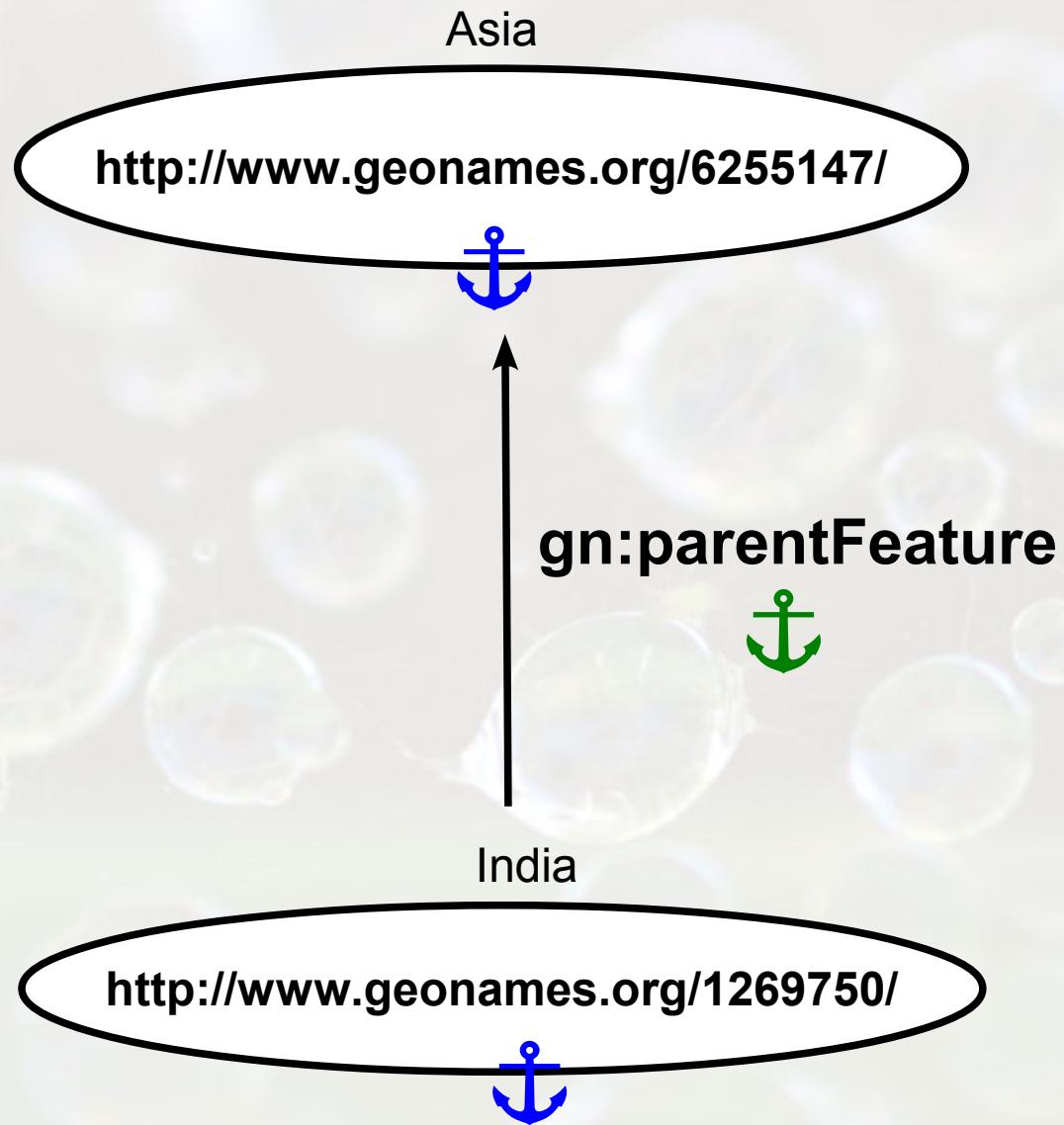
# Connected Graph



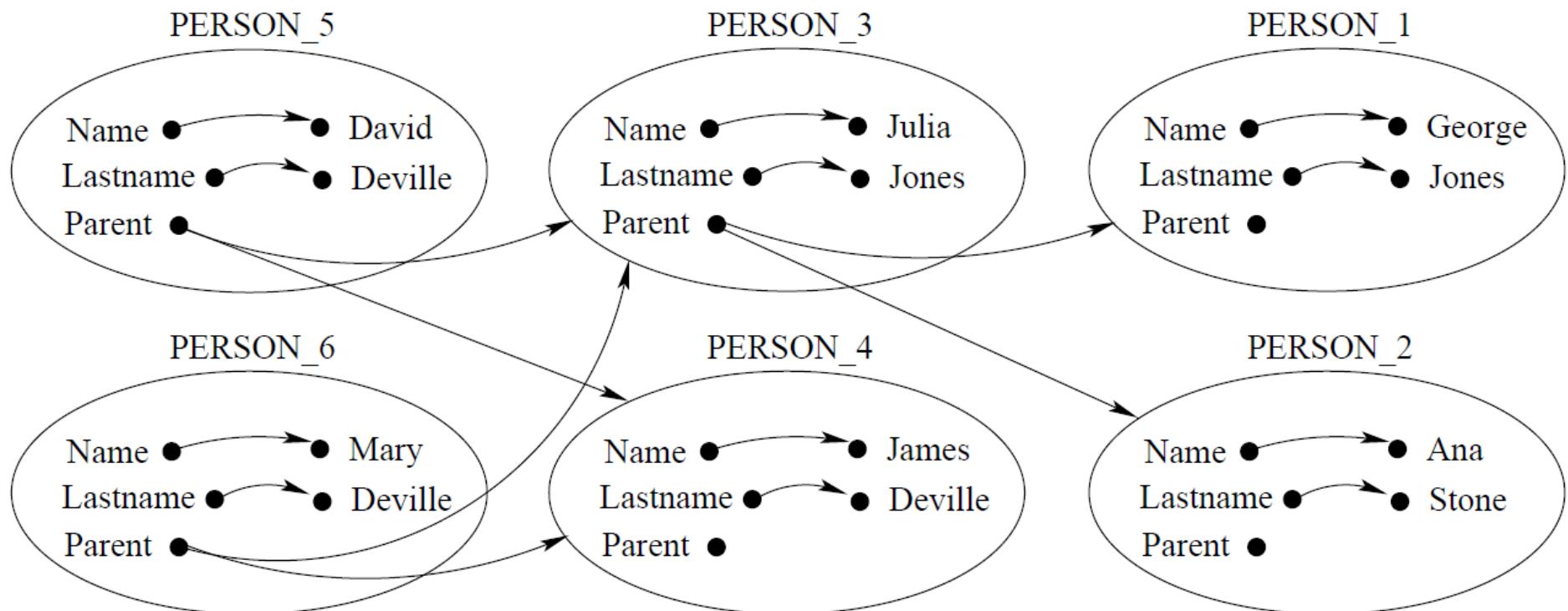
# Origin in GeoNames



# Geo Tree



# Hypergraph



(Jaudete Daltio, 2013)

# Licença

- Estes slides são concedidos sob uma Licença Creative Commons. Sob as seguintes condições: Atribuição, Uso Não-Comercial e Compartilhamento pela mesma Licença.
- Mais detalhes sobre a referida licença Creative Commons veja no link:  
<http://creativecommons.org/licenses/by-nc-sa/3.0/>
- Fotografia de capa e fundos: web-drops por Jeremy Hiebert [ <http://www.flickr.com/photos/jeremyhiebert/> ] disponível em <http://www.flickr.com/photos/jeremyhiebert/6081389428/>