# Scalable OWL 2 Reasoning for Linked Data

Aidan Hogan & Jeff Z. Pan

**Reasoning Web Summer School 2011**

*Day 2, 14:-00–15:30*

# The *Web of Data*!

August 2007

November 2007

February 2008

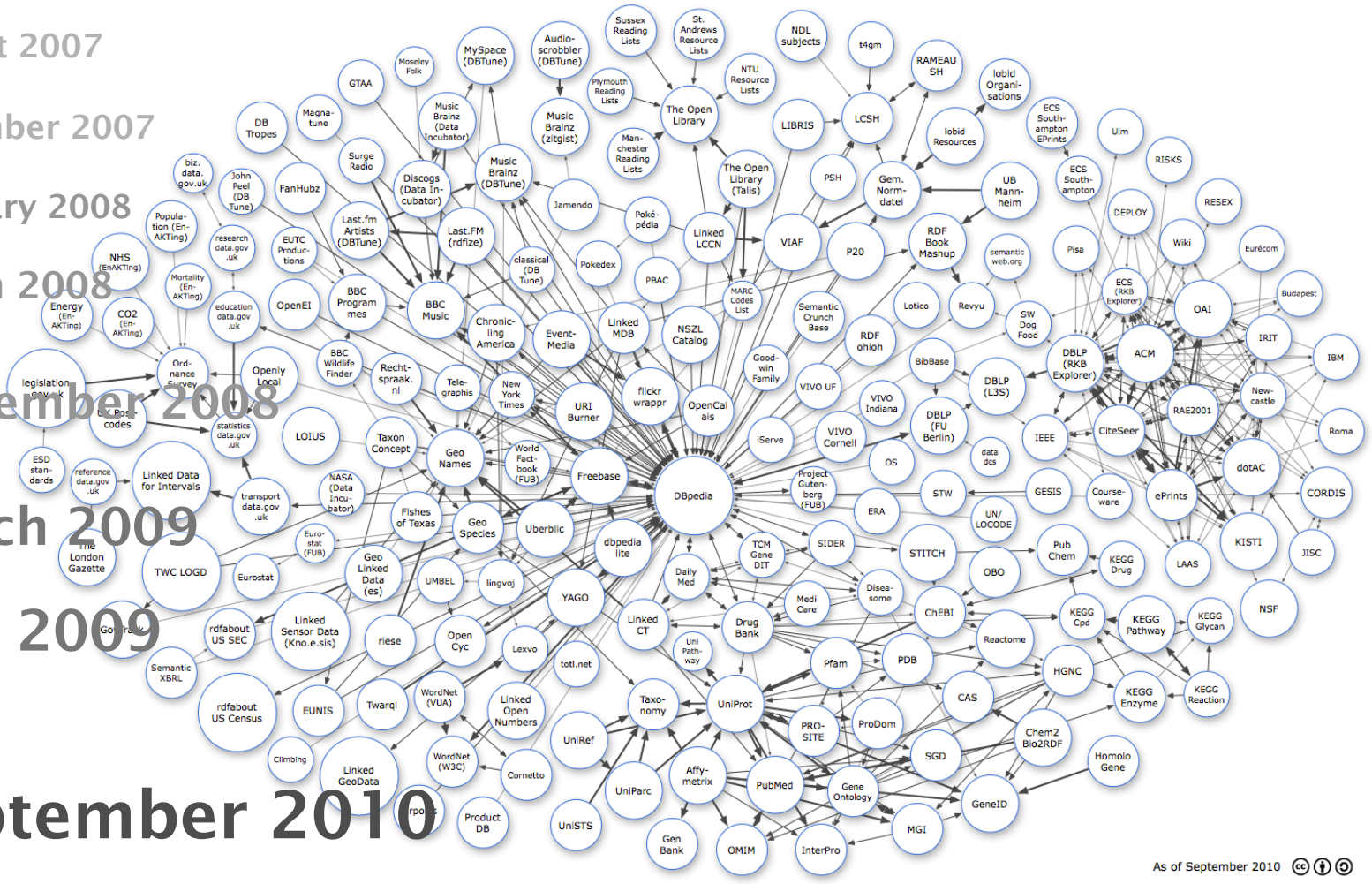
March 2008

September 2008

March 2009

July 2009

September 2010

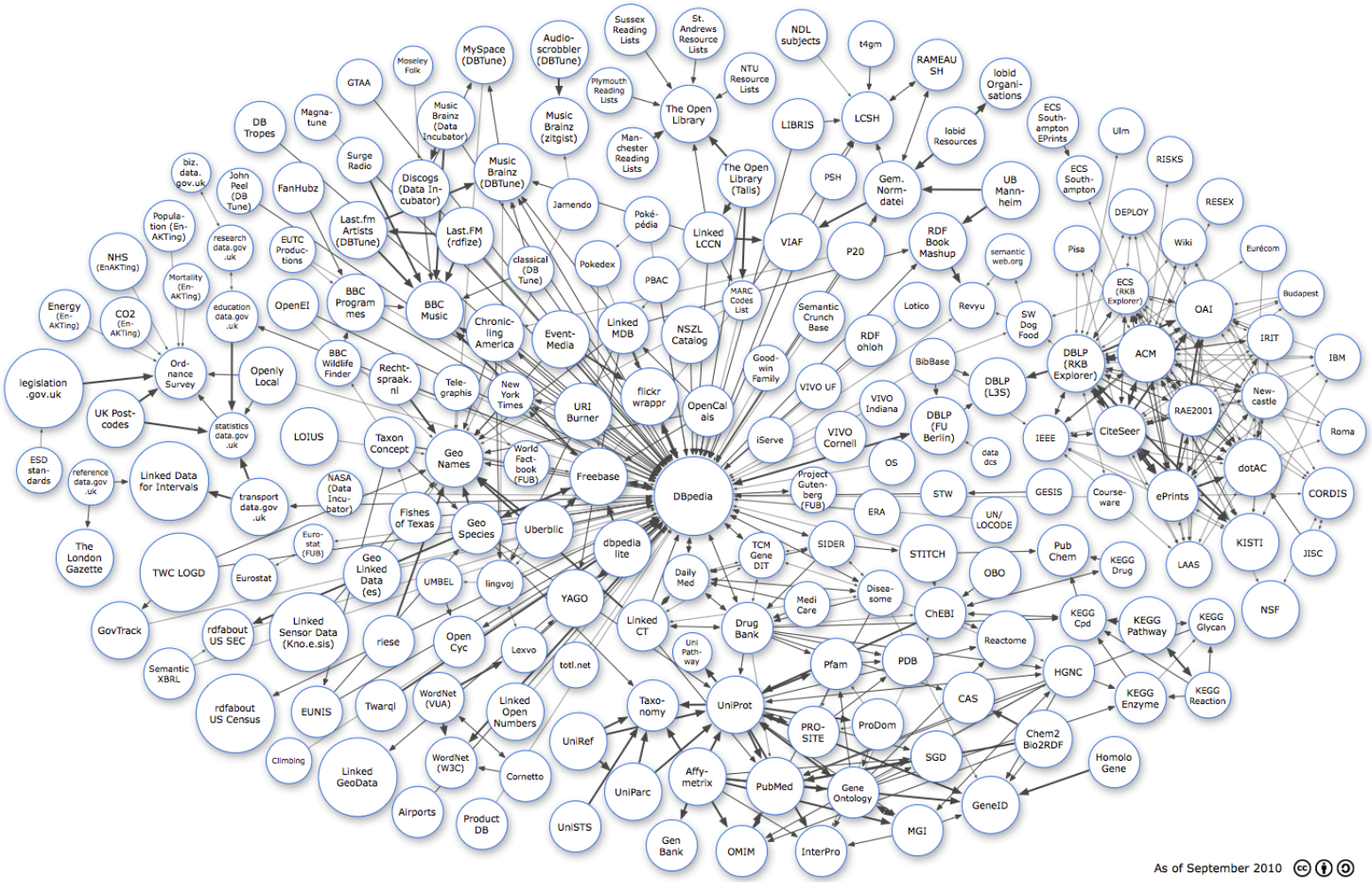



Images from: http://lod-cloud.net/ **Cyganiak, Jentzsch**



NUI Galway OÉ Gaillimh · sfi Science Foundation Ireland · DERI · University of Aberdeen 1495

# The *Web of Data*!




...reasoning...



NUI Galway OÉ Gaillimh

sfi Science Foundation Ireland

DERI

UNIVERSITY OF ABERDEEN

…data integration use-case for…

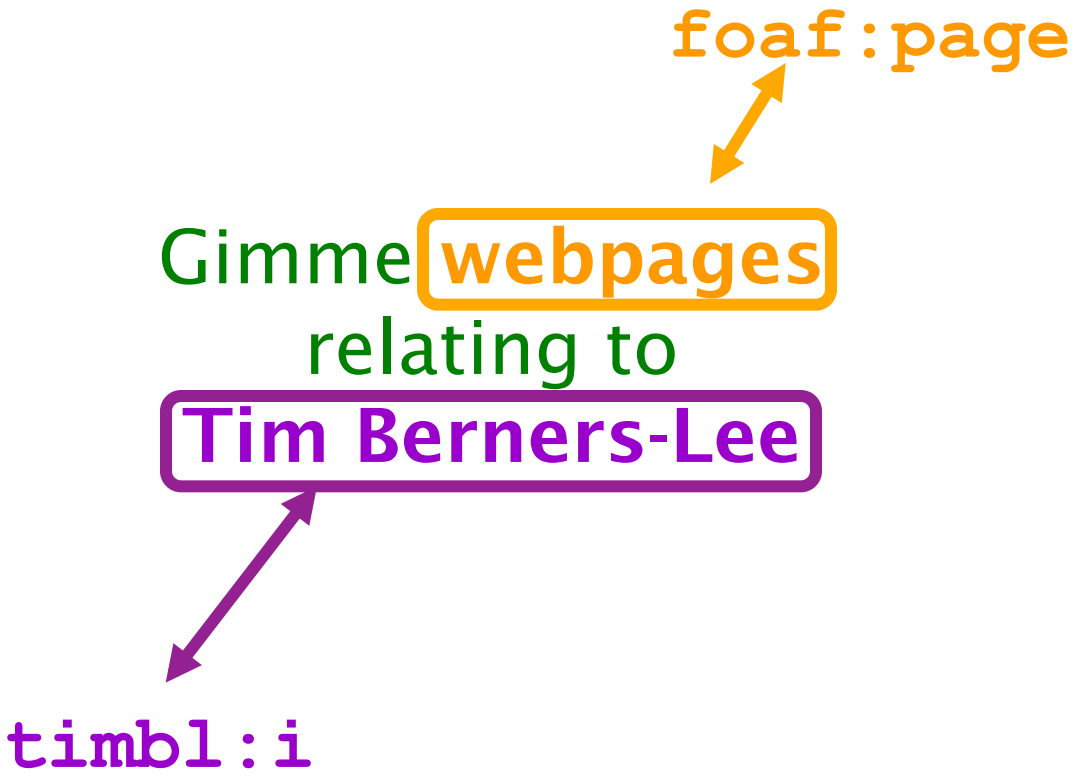
# LINKED DATA REASONING

…so what's **The Problem**?…

…*Heterogenity*

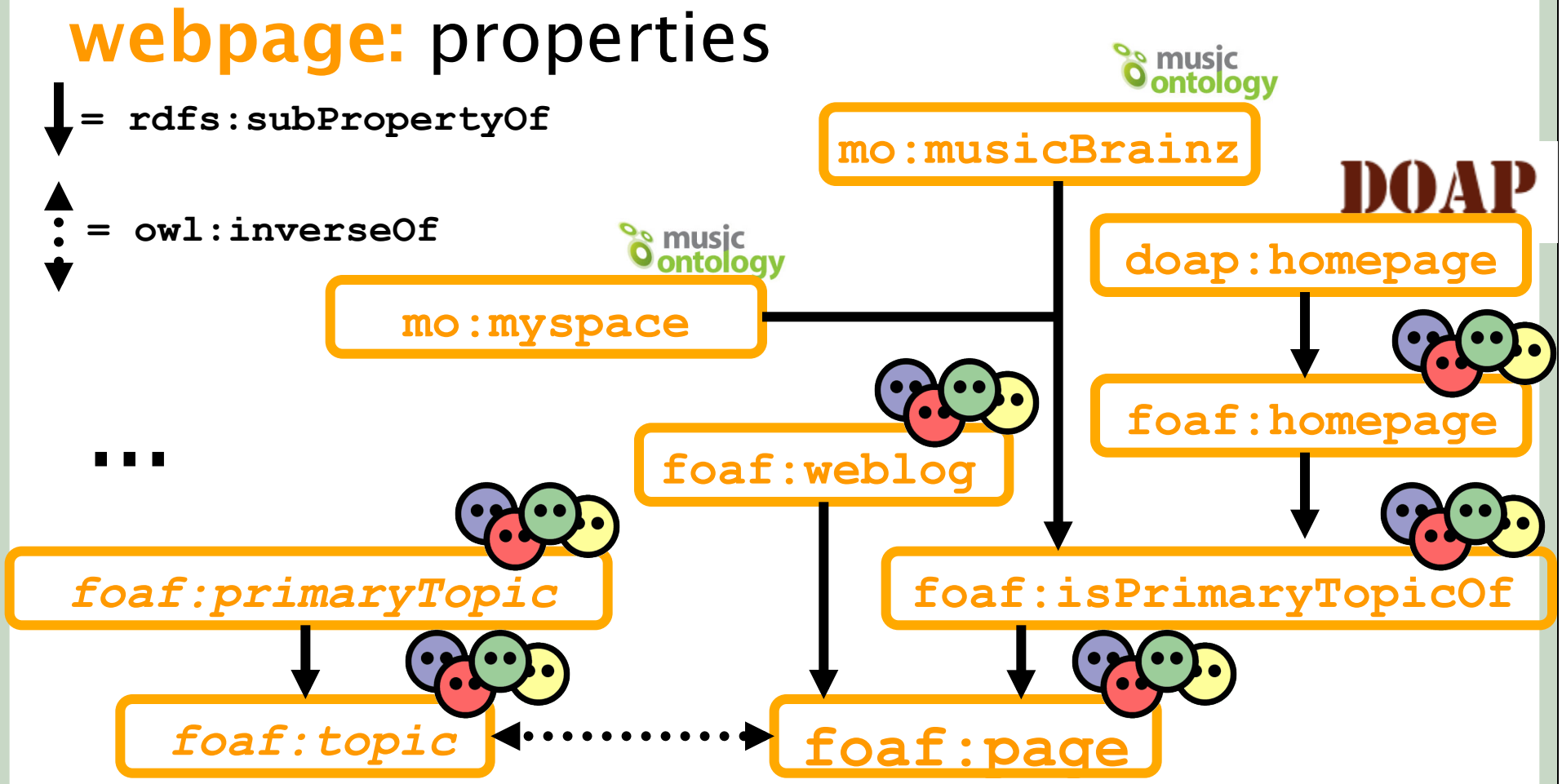
…***need to integrate data from different sources***

# Take **Query Answering**… (e.g. SPARQL)




```
timbl:i foaf:page ?pages .
```

# Hetereogenity in ***schema***…






NUI Galway OÉ Gaillimh · sfi Science Foundation Ireland · DERI · UNIVERSITY OF ABERDEEN

# Linked Data, RDFS and OWL: **Linked Vocabularies**

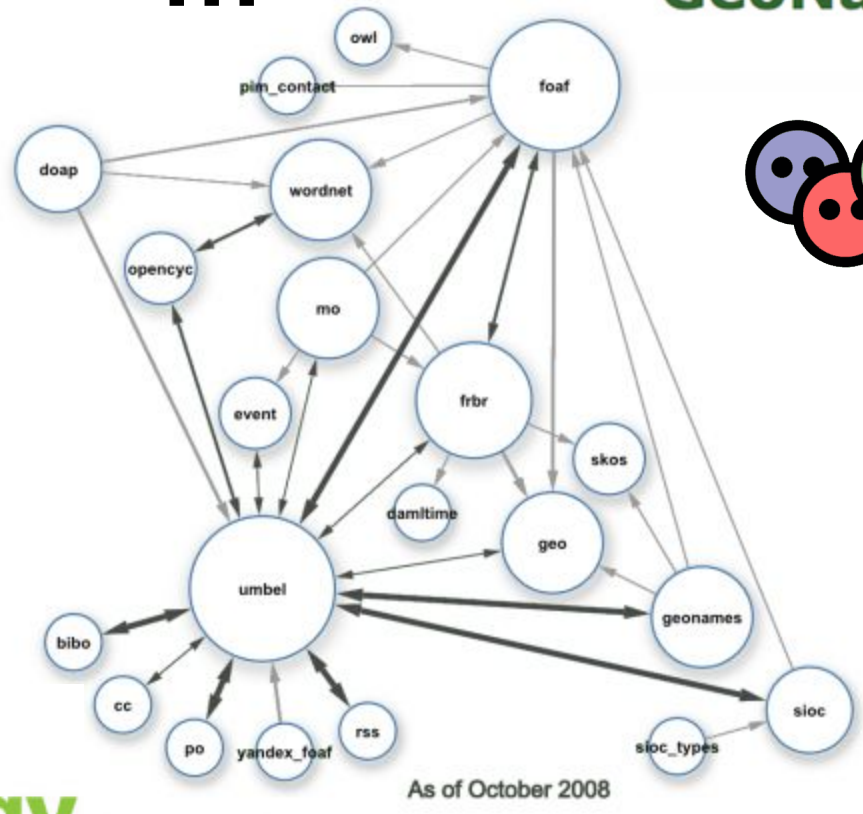
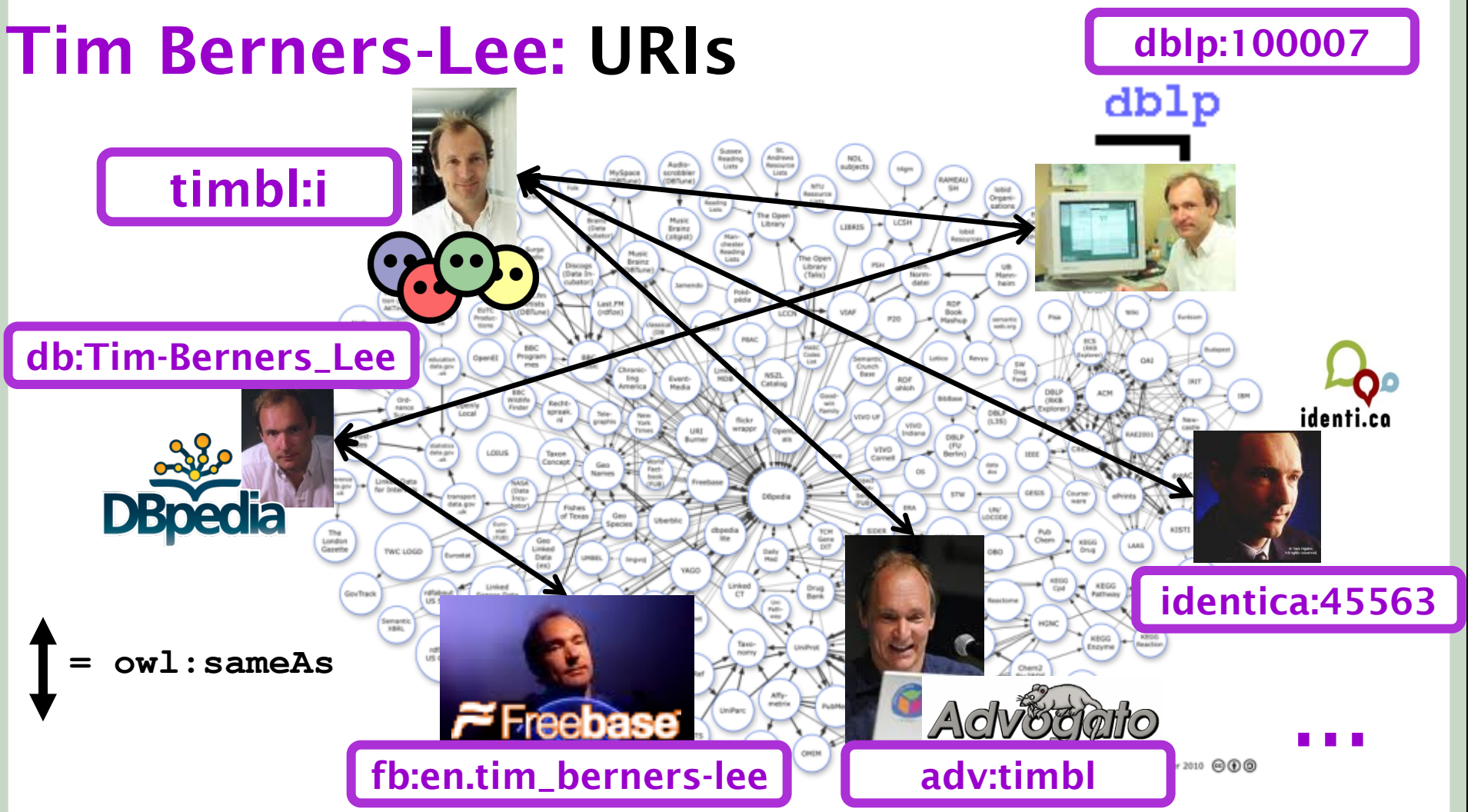



Image from http://blog.dbtune.org/public/.081005_lod_constellation_m.jpg: Giasson, Bergman

# Hetereogenity in ***naming***…

## Tim Berners-Lee: URIs



# Returning to our simple query…








DOAP

mo:myspace

foaf:primaryTopic

foaf:topic

foaf:page

doap:homepage

foaf:homepage

foaf:isPrimaryTopicOf


Gimme webpages relating to Tim Berners-Lee

identica:45563

adv:timbl

db:Tim-Berners_Lee

dblp:100007

fb:en.tim_berners-lee

timbl:i


```
timbl:i foaf:page ?pages .
```

... 7 x 6 = 42 possible patterns

# …reasoning to the rescue?


"THERE'S PROBABLY NO SEMANTIC WEB…
…NOW STOP INFERING AND GET LOD'ing"

Image from: http://www.whatreallypissesmeoff.com/hugh/

# Can we avoid reasoning?

OKKAM: What's timbl's URI?


Okay, what does that dereference to?

http://okkam.com/person/l33t


404

…

OKKAM: I'm doing my FOAF file… What's my URI?

…

**Hmm…** how about avoiding the heterogenity in the first place…



NUI Galway OÉ Gaillimh

sfi Science Foundation Ireland

DERI

1495 UNIVERSITY OF ABERDEEN

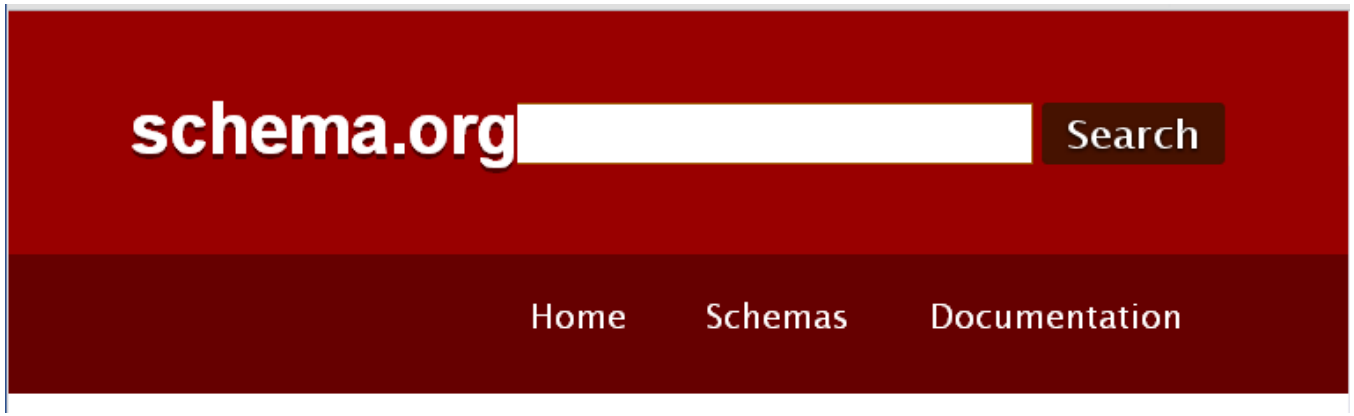
# Can we avoid reasoning?






**Err…** what about a single world model, or a centralised schema, or at least an upper ontology to get schema-level agreement started…

# Can we avoid reasoning?



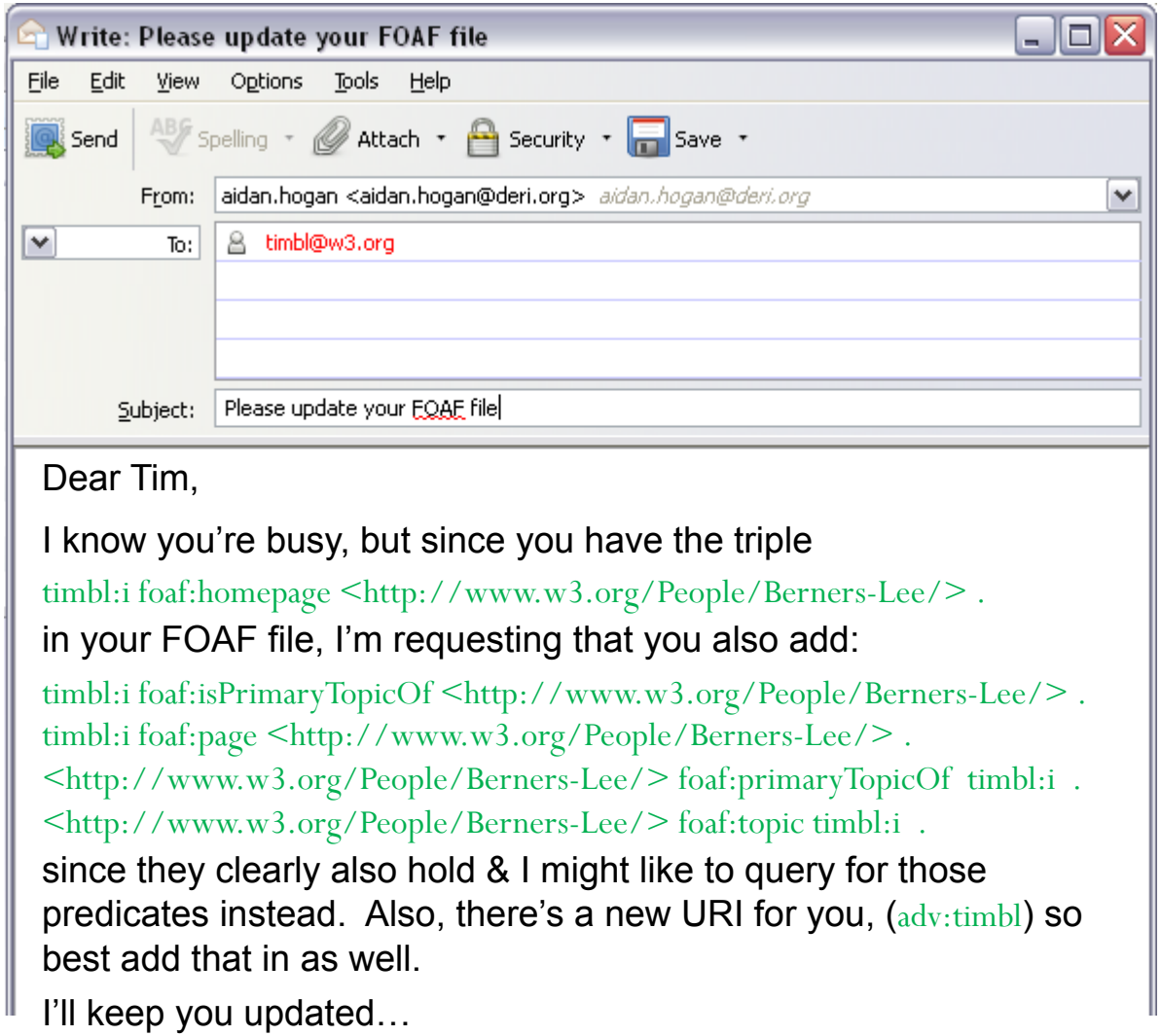

**Thing > Organization > LocalBusiness > Store**

A retail good store.

- LiquorStore
- MensClothingStore
- MobilePhoneStore
- MovieRentalStore
- MusicStore
- OfficeEquipmentStore
- OutletStore
- PawnShop
- PetStore
- ShoeStore
- SportingGoodsStore
- TireShop
- ToyStore
- WholesaleStore
- ???

# Can we avoid reasoning?

**Sure**! just email all the publishers and tell them to write data in all combinations…



# Can we avoid reasoning?

**Okay…** then just tell the users to ask the *right* query…

Query

Default Graph URI

Use only local data (including data retrieved before), but do not retrieve more

Query text

```
PREFIX …
SELECT ?page
WHERE
{ timbl:i foaf:page ?page .
UNION { identica:45563 foaf:page ?page . }
UNION { dbpedia:Berners-Lee foaf:page ?page . }
UNION { dbpedia:Tim_Berners-Lee foaf:page ?page . }
UNION { semweb:Tim_Berners-Lee foaf:page ?page . }
UNION { dblp:100007 foaf:page ?page . }
UNION { avogato:me foaf:page ?page . }
UNION { freebase:en.tim_berners-lee foaf:page ?page . }
UNION { book:Tim+Berners-Lee foaf:page ?page . }
UNION { yago:Tim_Berners-Lee foaf:page ?page . }
UNION { timbl:i foaf:homepage ?page .}
UNION { timbl:i foaf:weblog ?page .}
UNION { timbl:i foaf:isPrimaryTopicOf ?page .}
UNION { ?page foaf:primaryTopic timbl:i .}
```

## Can we avoid reasoning?

...**sure!** ...

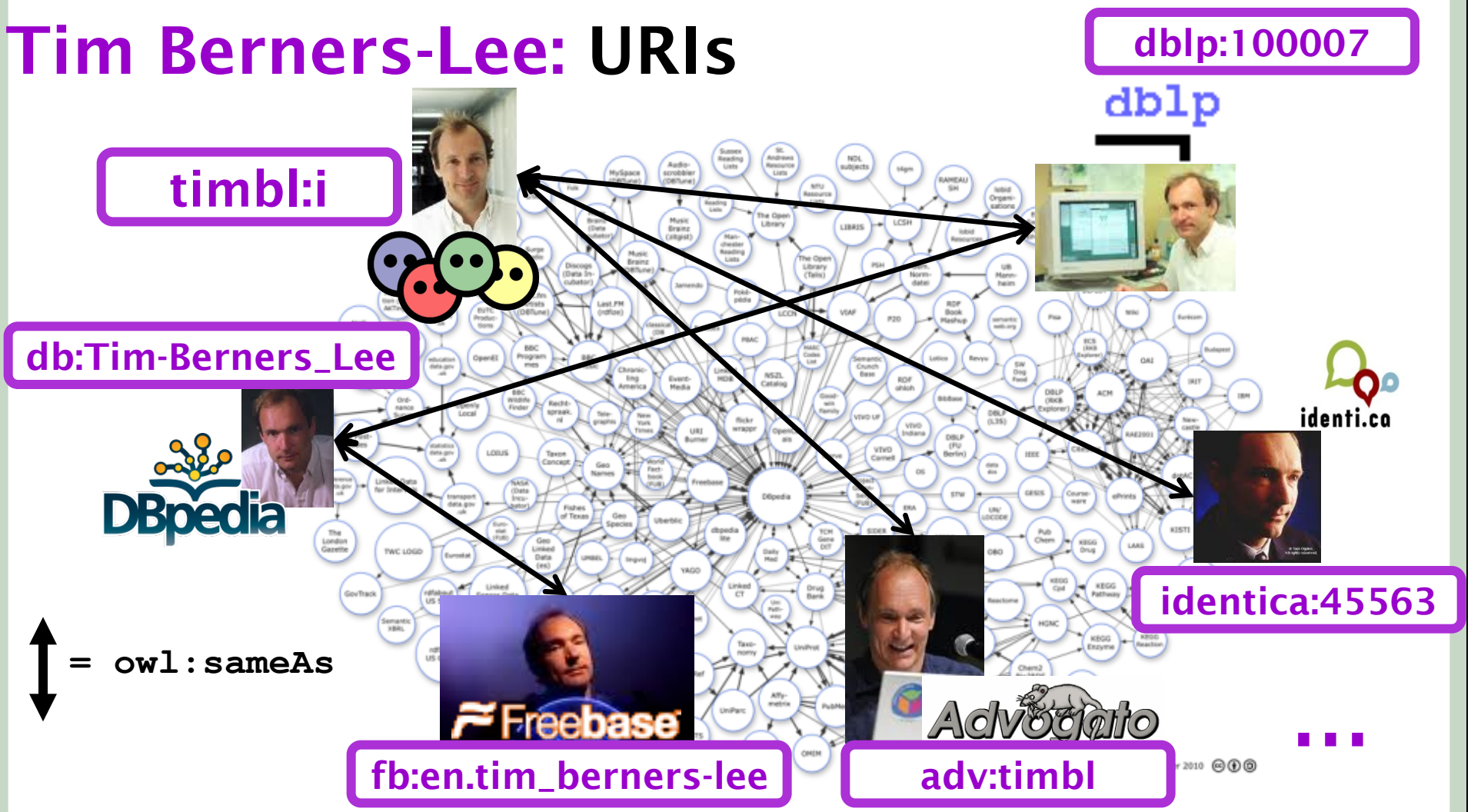
***…but it's probably better to just do reasoning…***

…what's out there…

# RDFS, OWL & LINKED DATA

# Hetereogenity in ***naming***…

## Tim Berners-Lee: URIs



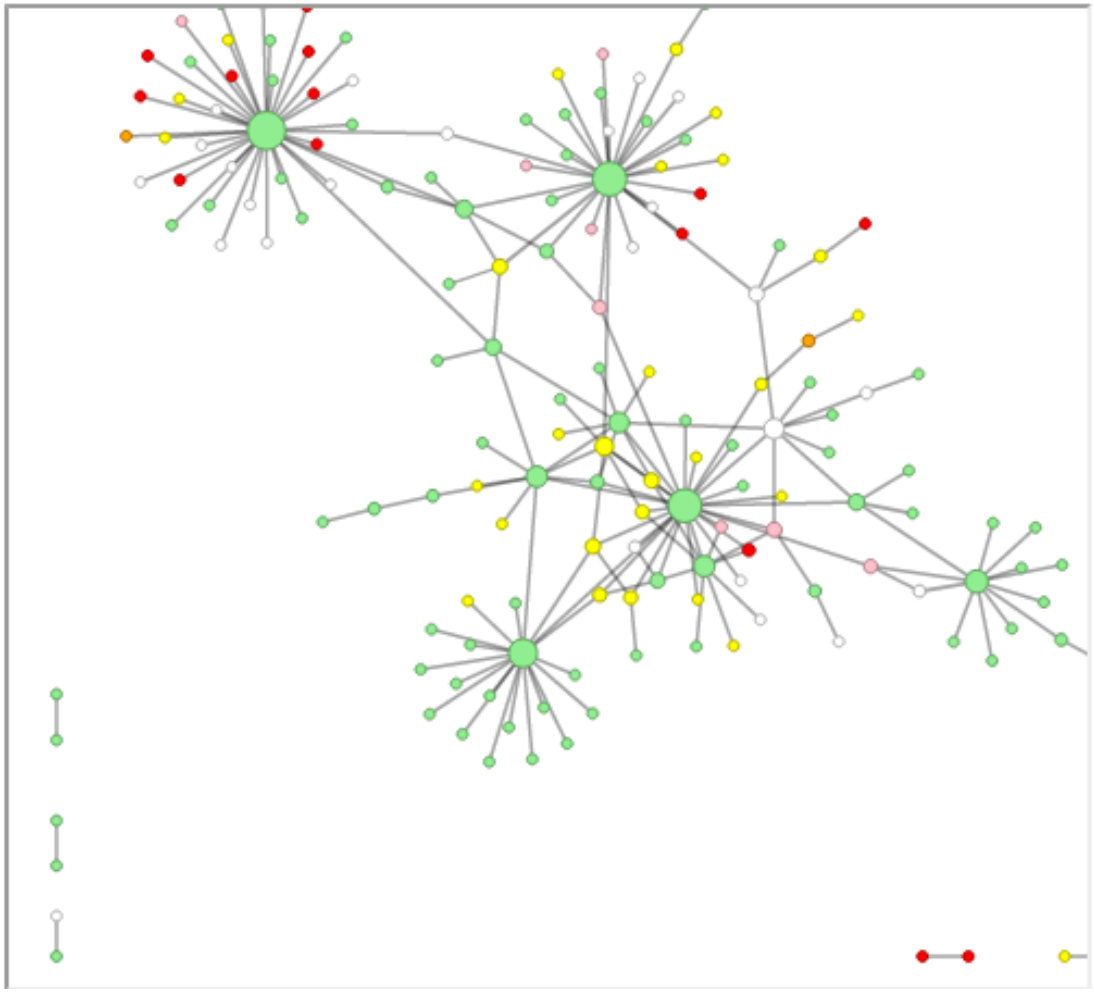



NUI Galway OÉ Gaillimh · sfi Science Foundation Ireland · DERI · 1495 University of Aberdeen

"*By common agreement, Linked Data publishers use the link type* [**owl:sameAs**] *to state that two URI aliases refer to the same resource.*"

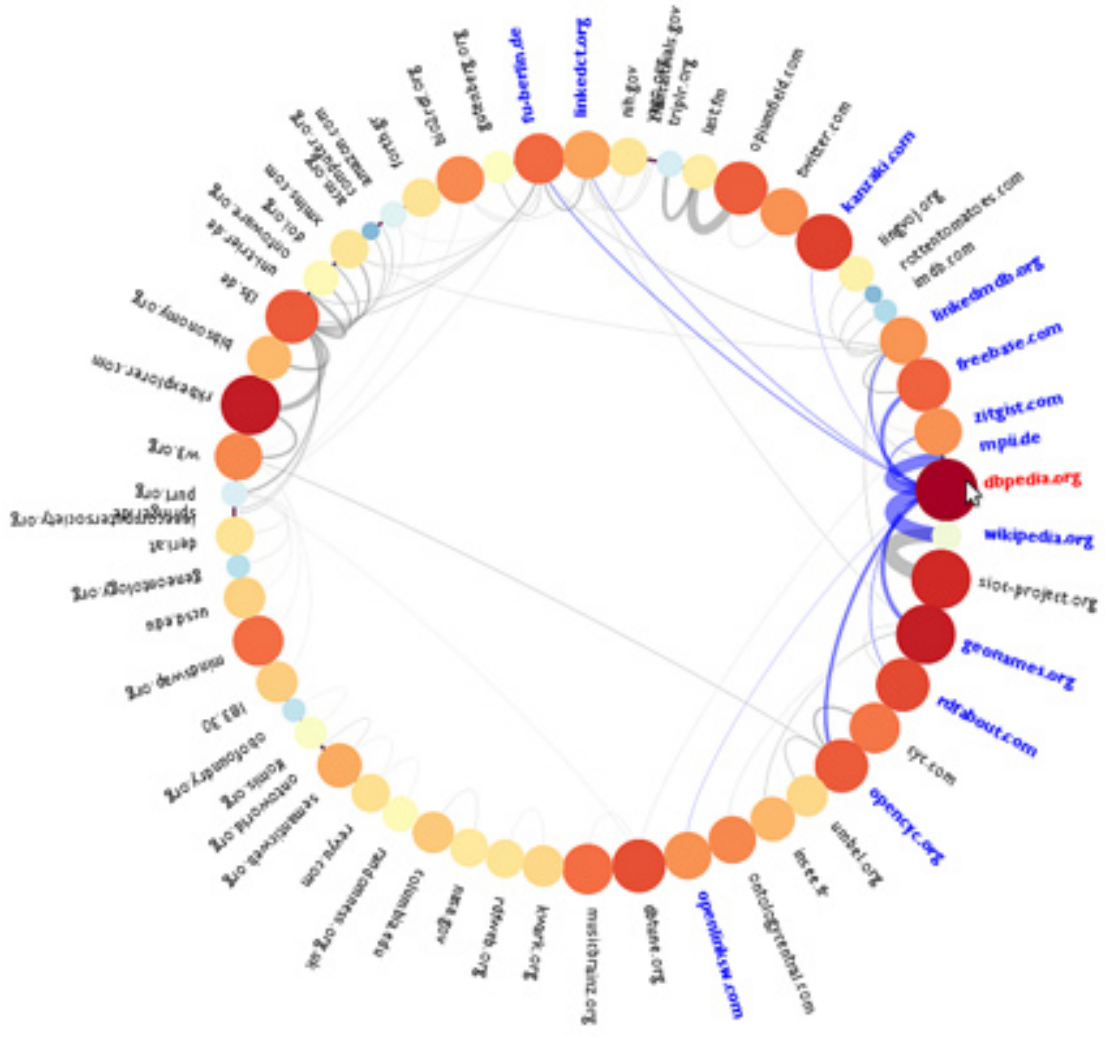
**- Heath & Bizer.** *Linked Data: Evolving the Web into a Global Data Space.* Morgan & Claypool. 2011.

# owl:sameAs linkage (i)



Interactive http://inkdroid.org/empirical-cloud/ ; E. Summers

# owl:sameAs linkage (ii)




Interactive http://gromgull.net/2010/01/swball/swball.svg ; G.A. Grimnes



NUI Galway OÉ Gaillimh · sfi Science Foundation Ireland · DERI · UNIVERSITY OF ABERDEEN

# owl:sameAs quality

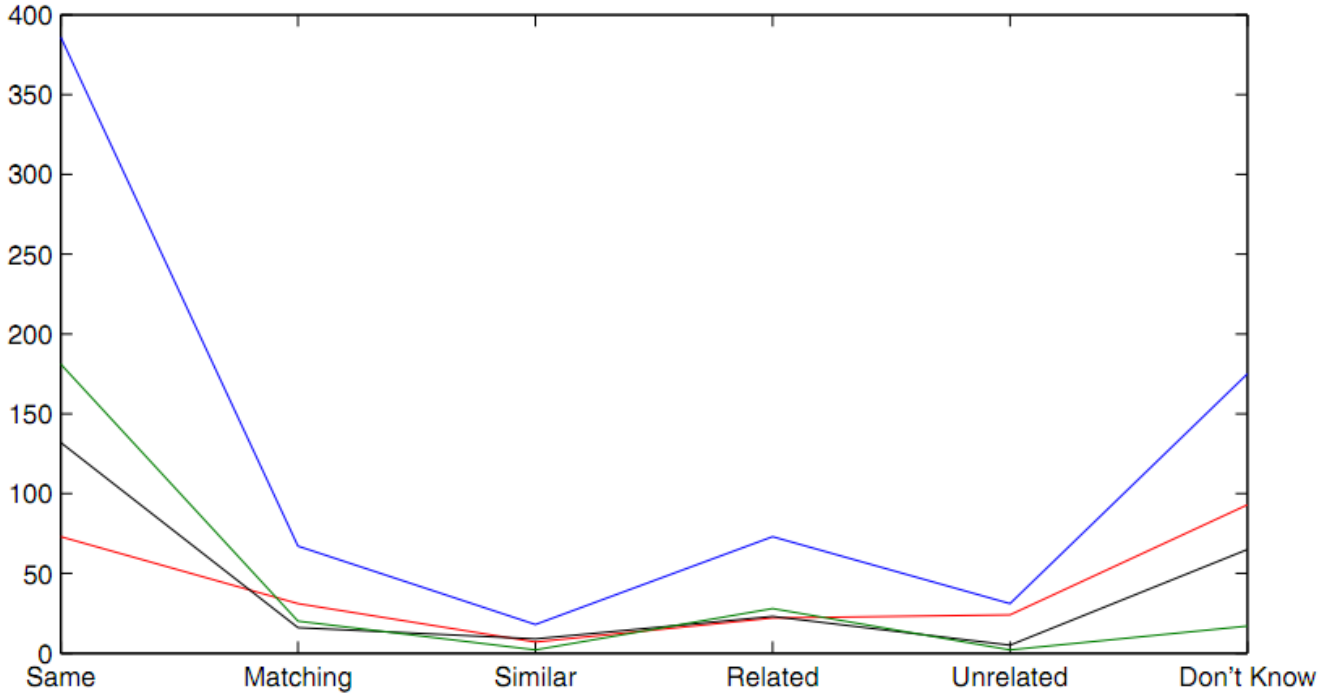
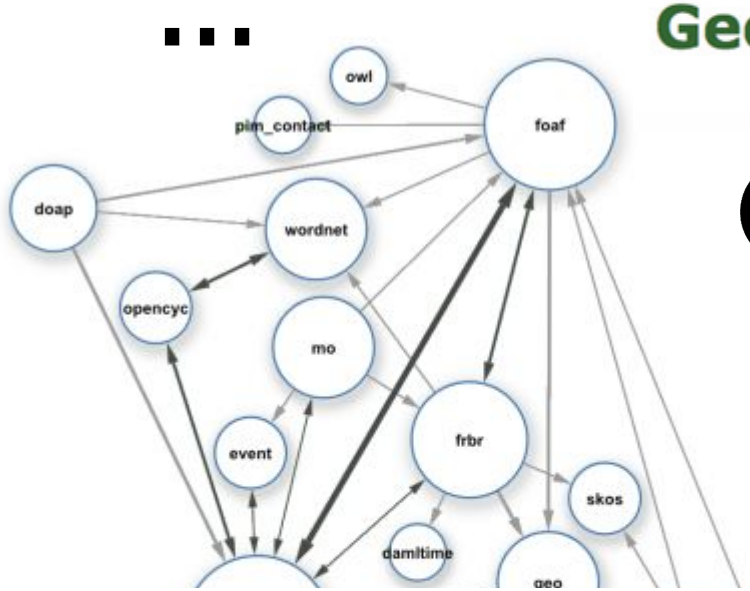



**Fig. 4.** Number of category assignments per judge. Total across all judges blue, each individual judge is red (1), black (2), and green (3). *Y*-axis is their frequency in the data-set.

- **Halpin et al.** *When owl: sameAs Isn't the Same: An Analysis of Identity in Linked Data.* ISWC, 2010.

# (Linked) Vocabularies Overview




- **Formalised using RDFS and OWL standards introduced yesterday**
  - ***(Typically OWL Full)***


Image from http://blog.dbtune.org/public/.081005_lod_constellation_m.jpg: **Giasson, Bergman**

"*The Web of Data takes a two-fold approach to dealing with heterogeneous data representation.*

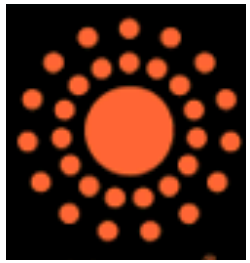
"*On the one hand side, it tries to avoid heterogeneity by advocating the* ***reuse of terms from widely deployed vocabularies****.* […] *a set of vocabularies for describing common things like people, places or projects has emerged in the Linked Data community.*

"*On the other hand,* […] *a Linked Data application which discovers some data* […] *using a previously unknown vocabulary should be able to find all meta-information that it requires to translate the data into a representation that it understands and can process.* **[Vocabularies provide]** ***RDFS and OWL definition of terms****,* **[each]** ***vocabulary term links to its own definition****,* […] ***mappings*** **[are provided]** ***between terms from different vocabularies****."*

**- Heath & Bizer.** *Linked Data: Evolving the Web into a Global Data Space.* Morgan & Claypool. 2011.

# (Linked) Vocabularies: *Dublin Core (DC)*

- Dublin **Core**
- **Models terms for *personal information***



| | |
|---|---|
| Properties in the */terms/* namespace | abstract, accessRights, accrualMethod, accrualPeriodicity, accrualPolicy, alternative, audience, available, bibliographicCitation, conformsTo, contributor, coverage, created, creator, date, dateAccepted, dateCopyrighted, dateSubmitted, description, educationLevel, extent, format, hasFormat, hasPart, hasVersion, identifier, instructionalMethod, isFormatOf, isPartOf, isReferencedBy, isReplacedBy, isRequiredBy, issued, isVersionOf, language, license, mediator, medium, modified, provenance, publisher, references, relation, replaces, requires, rights, rightsHolder, source, spatial, subject, tableOfContents, temporal, title, type, valid |
| Properties in the legacy */elements/1.1/* namespace | contributor, coverage, creator, date, description, format, identifier, language, publisher, relation, rights, source, subject, title, type |
| Vocabulary Encoding Schemes | DCMIType, DDC, IMT, LCC, LCSH, MESH, NLM, TGN, UDC |
| Syntax Encoding Schemes | Box, ISO3166, ISO639-2, ISO639-3, Period, Point, RFC1766, RFC3066, RFC4646, RFC5646, URI, W3CDTF |
| Classes | Agent, AgentClass, BibliographicResource, FileFormat, Frequency, Jurisdiction, LicenseDocument, LinguisticSystem, Location, LocationPeriodOrJurisdiction, MediaType, MediaTypeOrExtent, MethodOfAccrual, MethodOfInstruction, PeriodOfTime, PhysicalMedium, PhysicalResource, Policy, ProvenanceStatement, RightsStatement, SizeOrDuration, Standard |

Table from http://dublincore.org/documents/dcmi-terms/

# (Linked) Vocabularies: *FOAF*

- **F**riend **O**f **A** **F**riend
- **Models terms for *personal information***

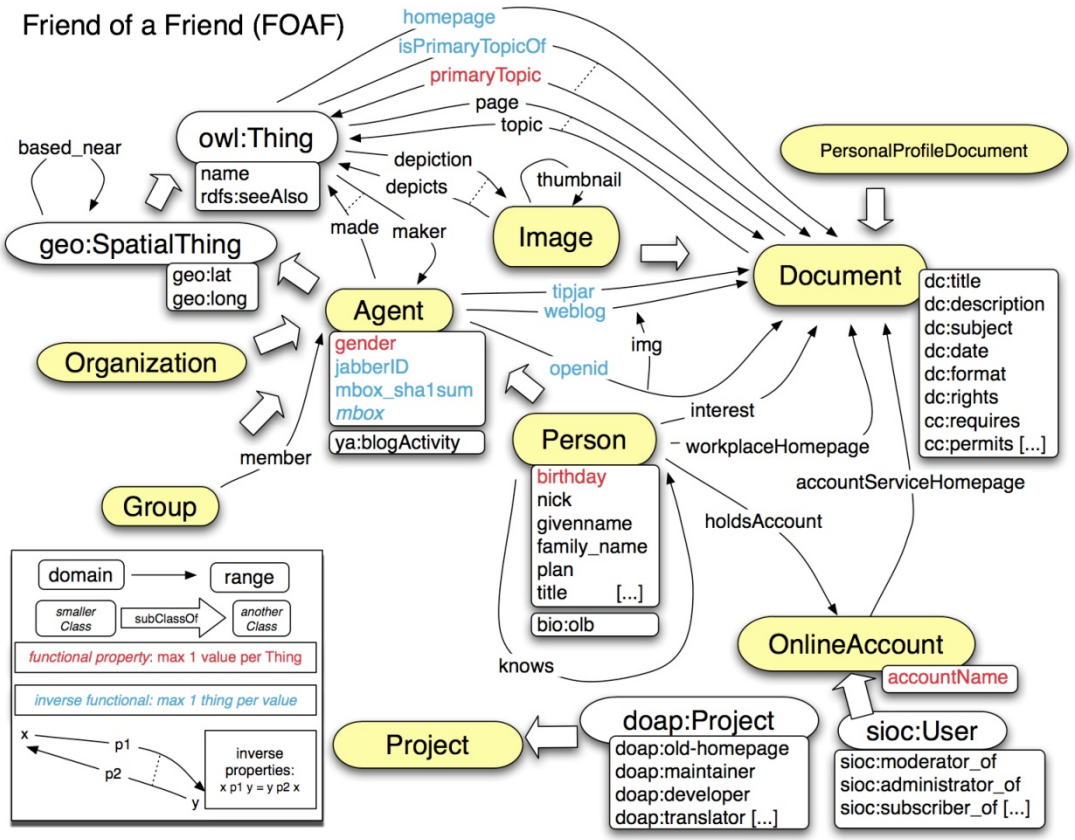



Image from http://www.deri.ie/fileadmin/images/blog/ : **Breslin**

# (Linked) Vocabularies: *SIOC*

- **S**emantically **I**nterlinked **O**nline **C**ommunities
- **Models terms for *online communities and presence***

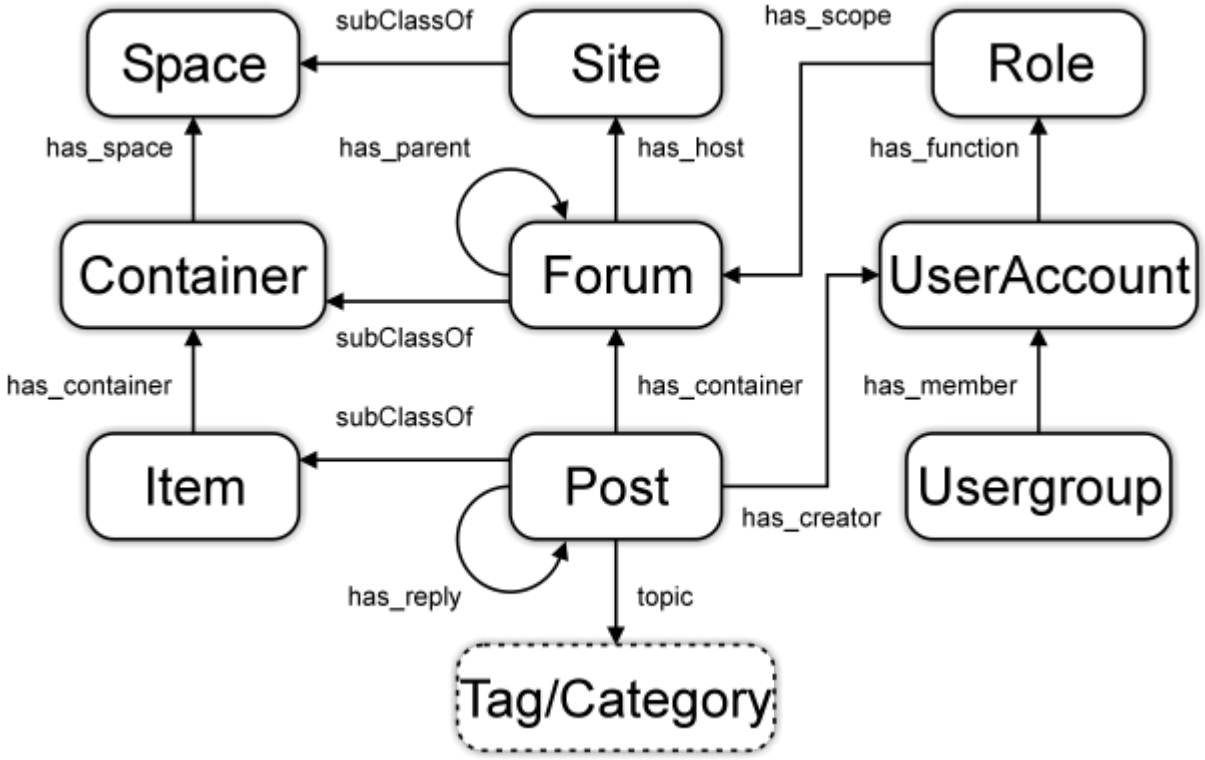


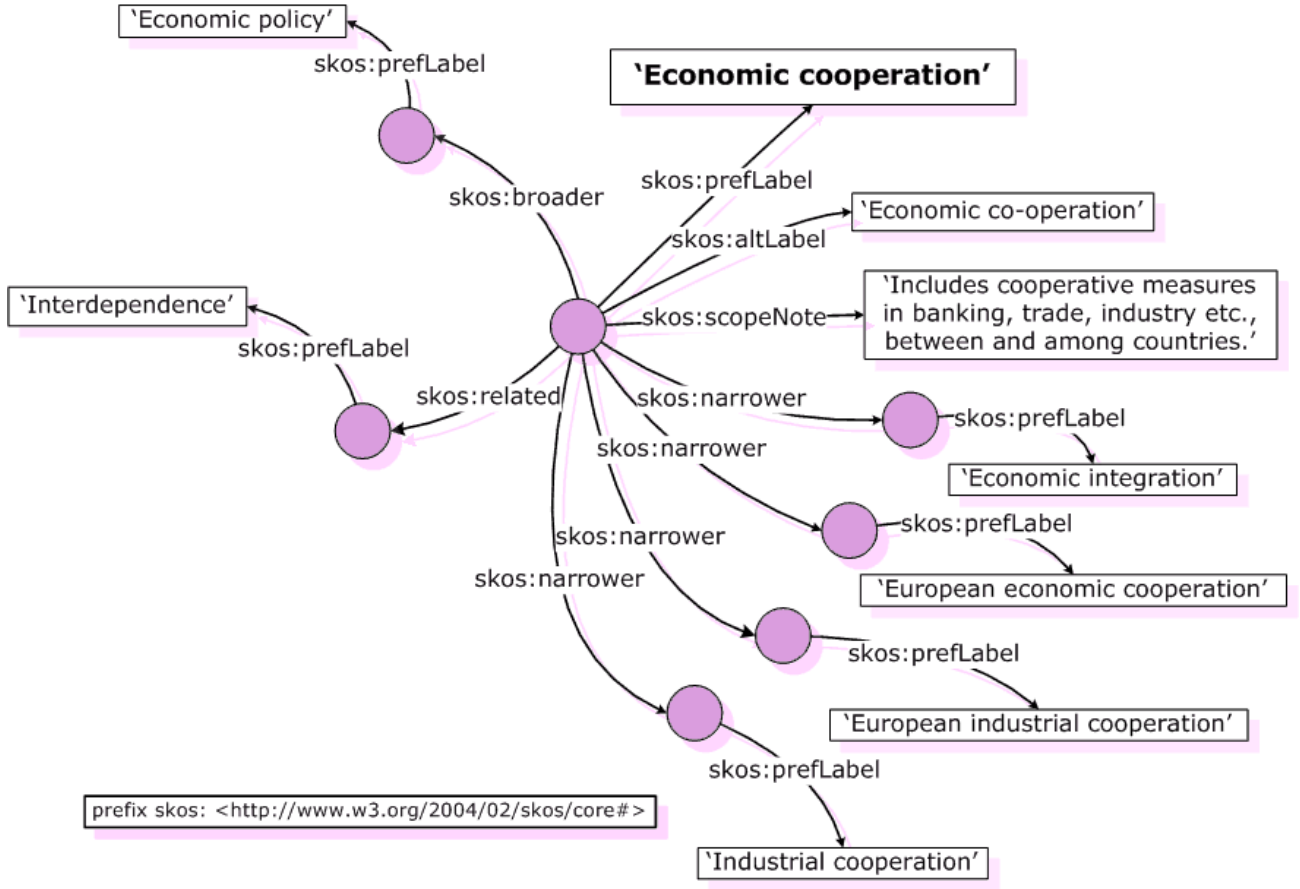

Image from http://rdfs.org/sioc/spec/ : **Bojārs, Breslin et al.**



NUI Galway OÉ Gaillimh | sfi | DERI | UNIVERSITY OF ABERDEEN

# (Linked) Vocabularies: *SKOS*

- **S**imple **K**nowledge **O**rganization **S**ystem
- **Metavocabulary for *concepts schemes***




W3C Recommendation

Image from http://www.w3.org/TR/swbp-skos-core-guide : **Miles, Brickley**

# (Linked) Vocabularies: *FOAF+SIOC+SKOS*

- **Example of how vocabularies can interleave**

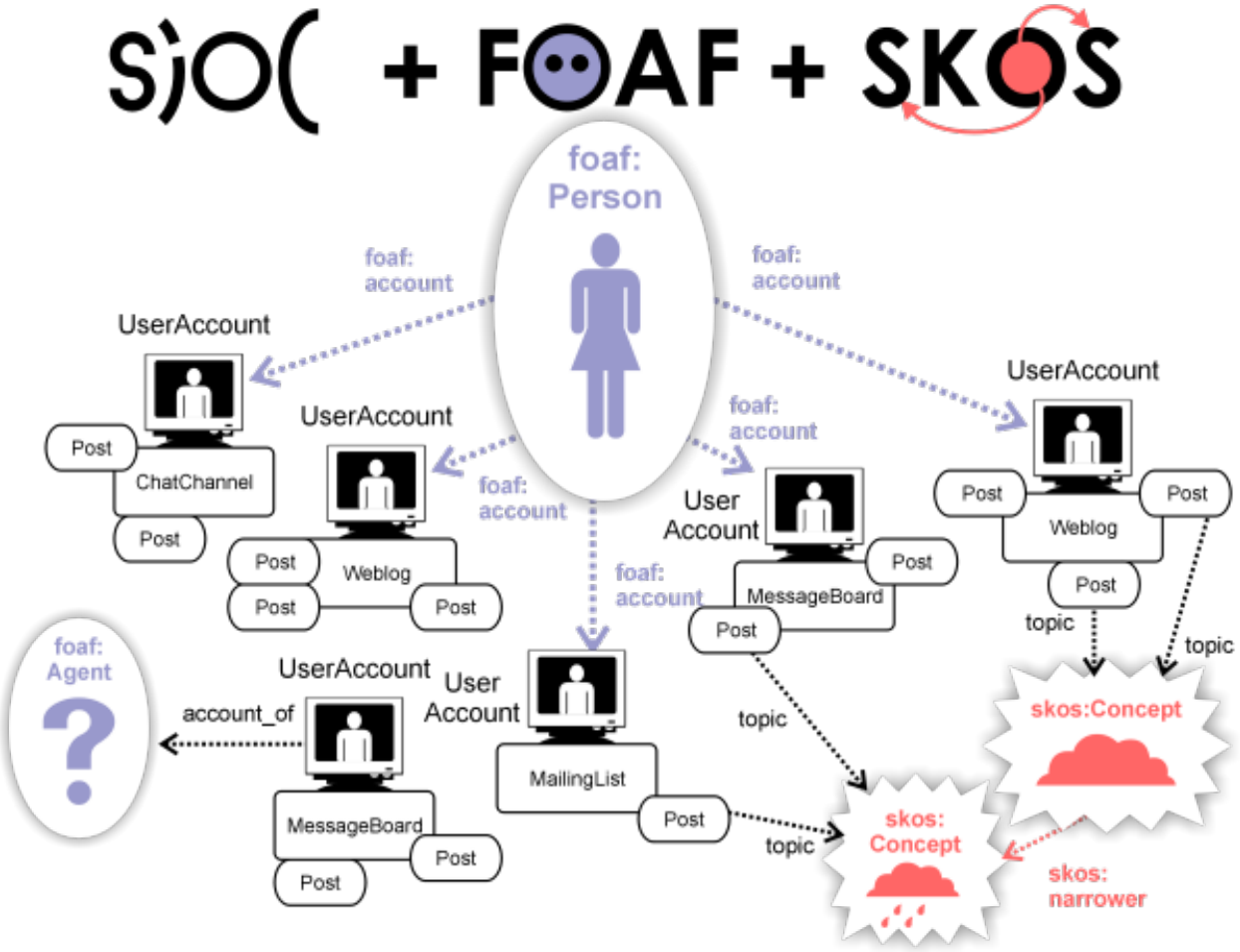



Image from http://sioc-project.org/node/158; **Breslin**

# (Linked) Vocabularies: *DOAP*

DOAP

- **D**escription **O**f **A** **P**roject
- **Models terms for *projects*** *(research, software, etc.)*

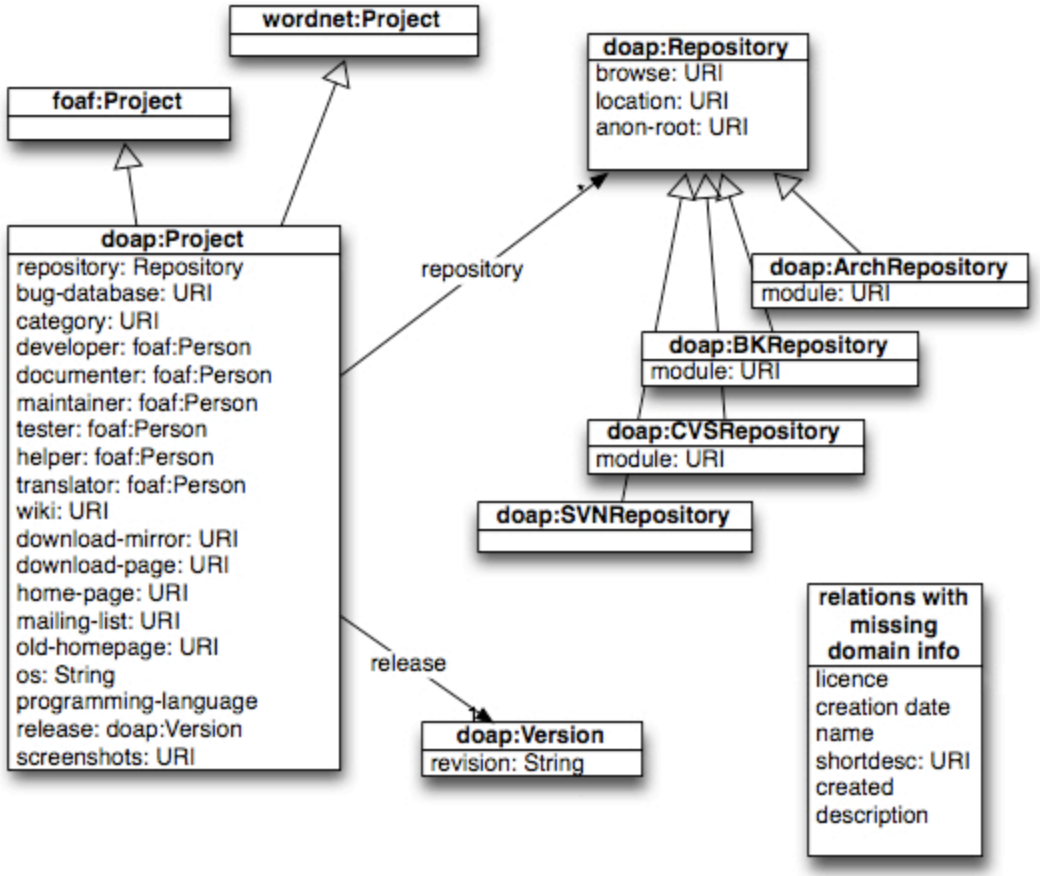



Image from http://code.google.com/p/baetle/wiki/DoapOntology ; **Breslin**

# (Linked) Vocabularies: *Music Ontology*

- **Models terms for *music artists, songs, albums etc.***
  - ***(very detailed)***

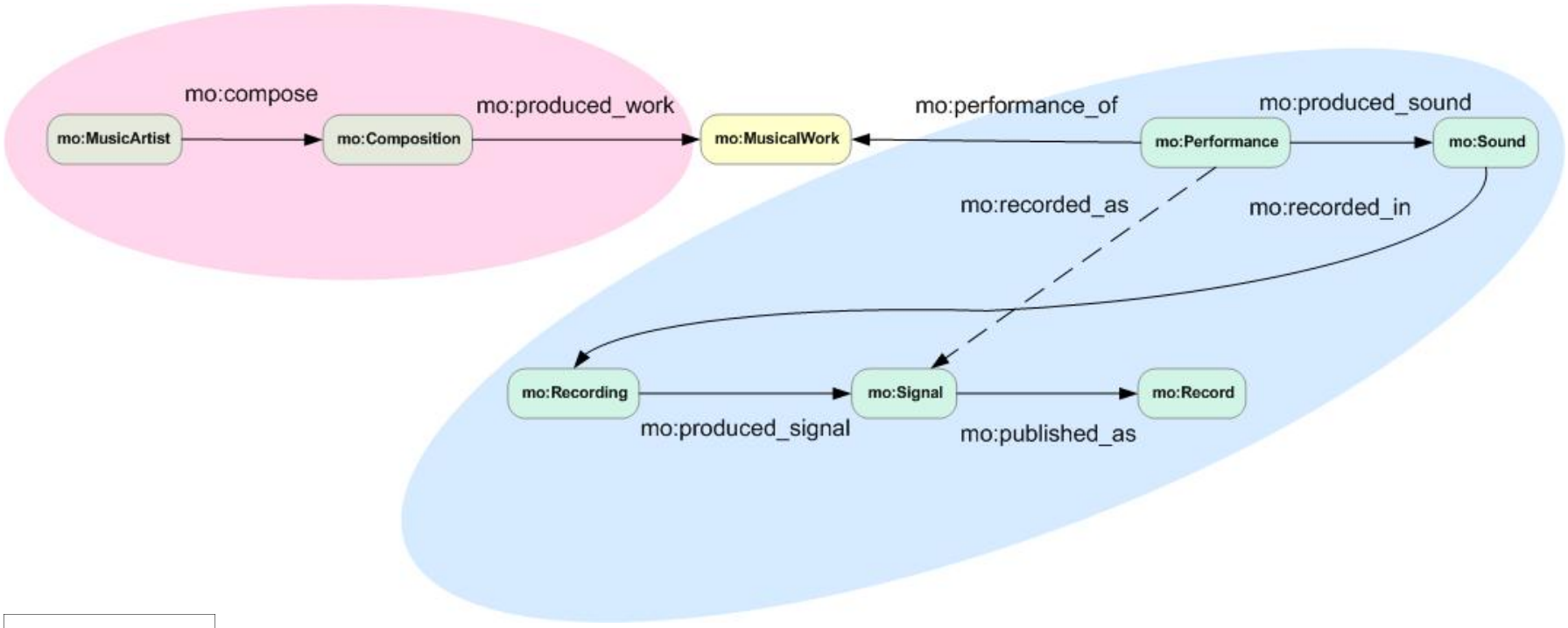



Image from http://musicontology.com/; **Raimond, Giasson**

# (Linked) Vocabularies: *GoodRelations (i)*

- **Models terms for *e-commerce, products, offerings etc.***

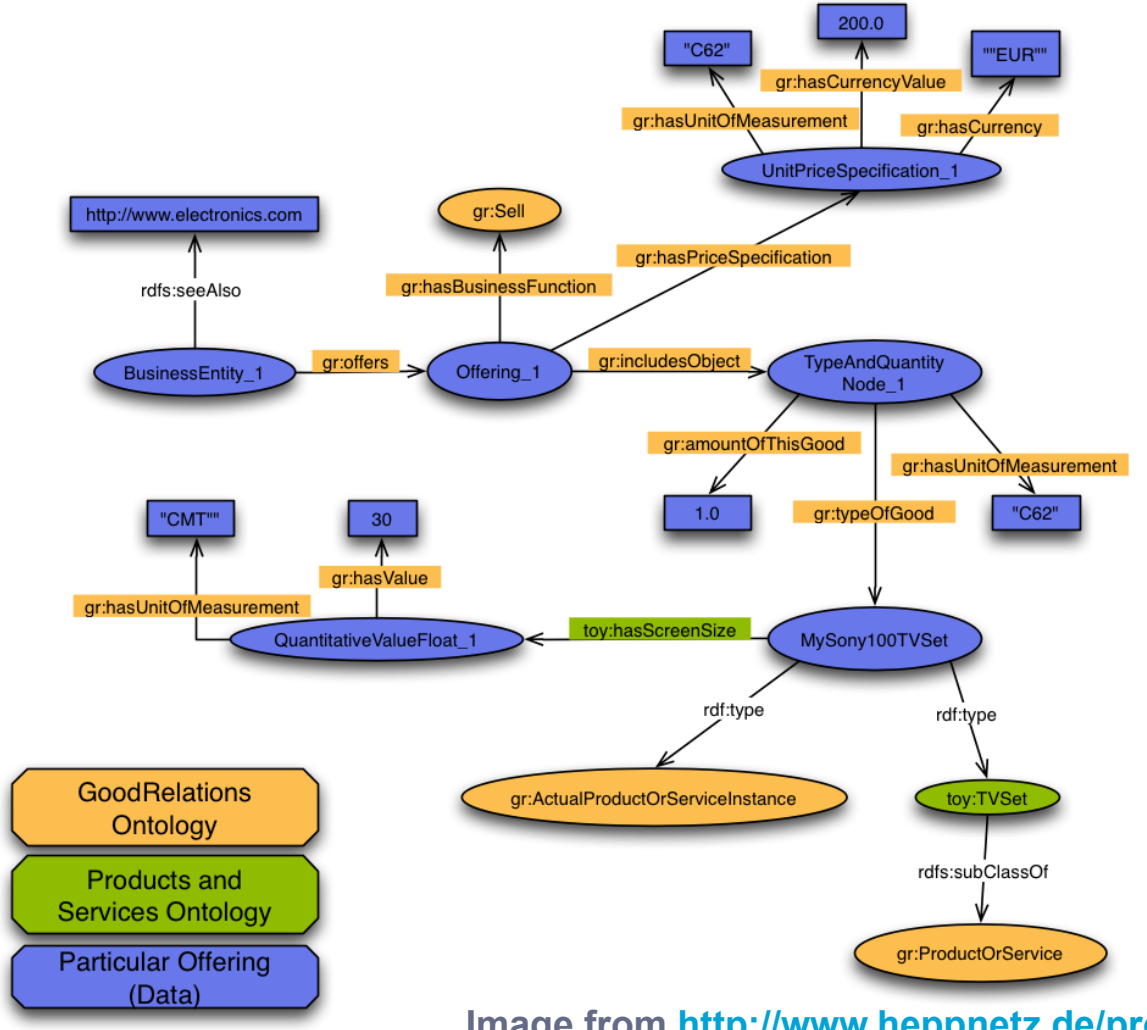



Image from http://www.heppnetz.de/projects/goodrelations/primer/; **Hepp**

# (Linked) Vocabularies: *GoodRelations (ii)*

- **Models terms for *e-commerce, products, offerings etc.***

GoodRelations
The Web Ontology for E-Commerce

Shopping results for **One-Touch Gold BBQ (22.5-in.): Blue by Weber**

**Weber** 75001 **Blue One-Touch 22.5" Blue** Stainless Steel Charcoal **Gold** Gr +1
$159.70 - Build.com

**Weber One Touch** 751001 **22.5" One-Touch Gold** Charcoal Grill with ... +1
$129.00 - AJ Madison

**Weber One Touch** 758001 **22.5" One-Touch Gold** Charcoal Grill with ... +1
$149.00 - AJ Madison

Google™

BEST BUY®

O'REILLY®

Y! ™

mart

overstock.com®

Sears

# (Linked) Vocabularies: *DBpedia*

- **Classes and properties for Wikipedia export**
  - **Cross-domain**
  - *272 classes*
  - *1,300 properties*
  - **(Too big to show)**
- **Used to model structured info-boxes in Wikipedia**

```
{{Infobox Town AT |
  name = Innsbruck |
  image_coa =  InnsbruckWappen.png |
  image_map = Karte-tirol-I.png |
  state = [[Tyrol]] |
  regbzk = [[Statutory city]] |
  population = 117,342 |
  population_as_of = 2006 |
  pop_dens = 1,119 |
  area = 104.91 |
  elevation = 574 |
  lat_deg = 47 |
  lat_min = 16 |
  lat_hem = N |
  lon_deg = 11 |
  lon_min = 23 |
  lon_hem = E |
  postal_code = 6010-6080 |
  area_code = 0512 |
  licence = I |
  mayor = Hilde Zach |
  website = [http://innsbruck.at] |
}}
```

**Innsbruck**

| | |
|---|---|
| Country | Austria |
| State | Tyrol |
| Administrative region | Statutory city |
| Population | 117,342 (2006) |
| Area | 104.91 km² |
| Population density | 1,119 /km² |
| Elevation | 574 m |
| Coordinates | 47°16' N 11°23' E |
| Postal code | 6010-6080 |
| Area code | 0512 |
| Licence plate code | I |
| Mayor | Hilde Zach |
| Website | www.innsbruck.at |

## About: Innsbruck

An Entity of Type : city, from Named Graph : http://dbpedia.org, within Data Space : dbpedia.org

Innsbruck is the capital city of the federal state of Tyrol in western Austria. It is located in the Inn Valley at the junction with the Wipptal, which provides access to the Brenner Pass, some 30 kilometers (19 mi) south of Innsbruck.

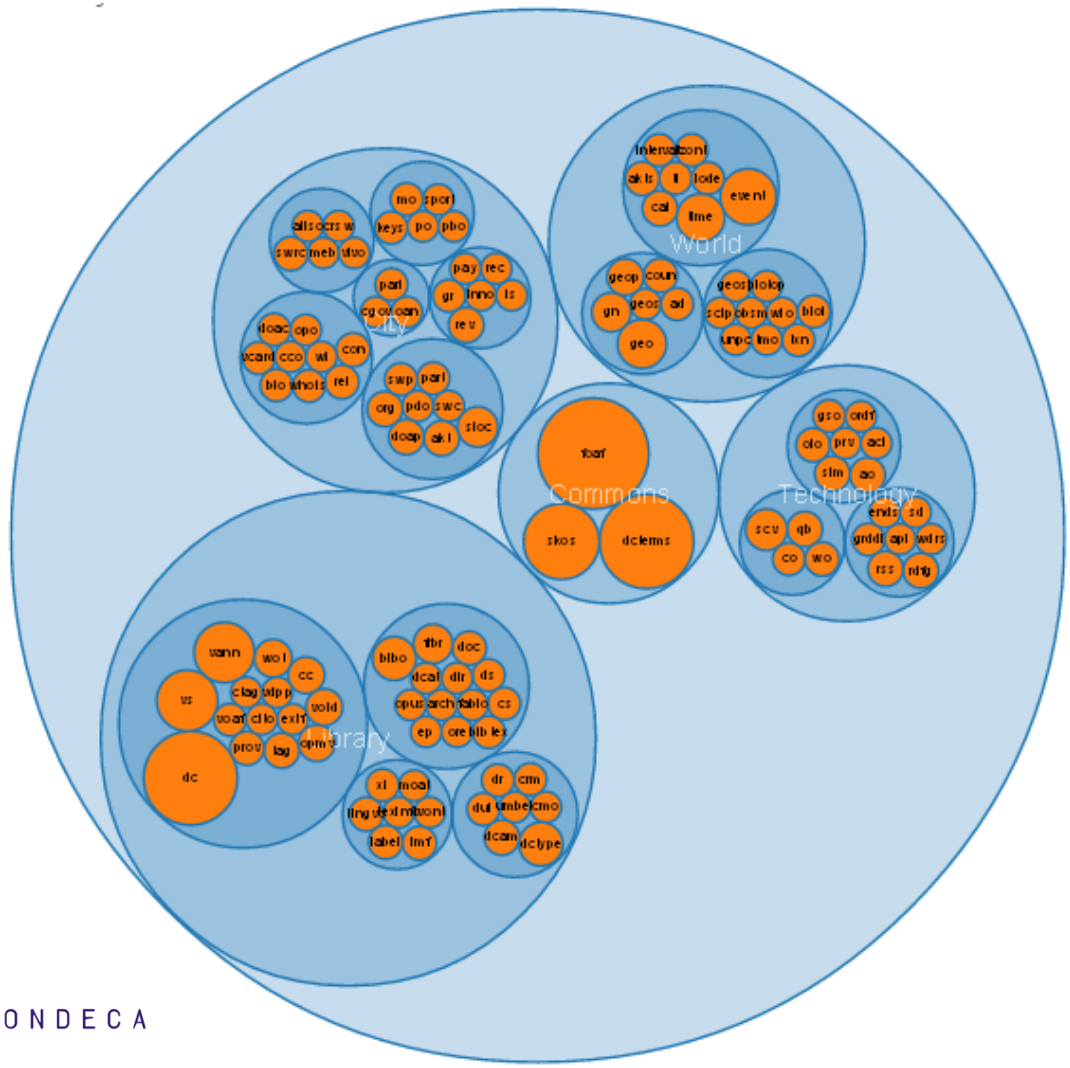
| Property | Value |
|---|---|
| dbpedia-owl:PopulatedPlace/populationDensity | 1119.0 |
| dbpedia-owl:abstract | Innsbruck ist die Landeshauptstadt des Bundeslandes Tirol Transit-Strecke Brenner (Auto- und Eisenbahn) nach Südti (Brücke über den Inn). Innsbruck ist mit 118.082 (Stand 1. und Salzburg die fünftgrößte Stadt Österreichs, im Ballung dazu kommen ca. 30.000 Studenten und andere Nebenwo Nächtigungen von Städtetouristen.<br>Innsbruck is the capital city of the federal state of Tyrol in the junction with the Wipptal, which provides access to the |

See http://wiki.dbpedia.org/

# **Linked Data vocabs**: top 15 RDFS/OWL features

| # | Axiom | Rank(Σ) | RDFS | Horst | O2R |
|---|---|---|---|---|---|
| 1. | `rdfs:subClassOf` | 0.295 | ✓ | ✓ | ✓ |
| 2. | `rdfs:range` | 0.294 | ✓ | ✓ | ✓ |
| 3. | `rdfs:domain` | 0.292 | ✓ | ✓ | ✓ |
| 4. | `rdfs:subPropertyOf` | 0.090 | ✓ | ✓ | ✓ |
| 5. | `owl:FunctionalProperty` | 0.063 | ✘ | ✓ | ✓ |
| 6. | `owl:disjointWith` | 0.049 | ✘ | ✘ | ✓ |
| 7. | `owl:inverseOf` | 0.047 | ✘ | ✓ | ✓ |
| 8. | `owl:unionOf` | 0.035 | ✘ | ✘ | ~ |
| 9. | `owl:SymmetricProperty` | 0.033 | ✘ | ✓ | ✓ |
| 10. | `owl:TransitiveProperty` | 0.030 | ✘ | ✓ | ✓ |
| 11. | `owl:equivalentClass` | 0.021 | ✘ | ✓ | ✓ |
| 12. | `owl:InverseFunctionalProperty` | 0.030 | ✘ | ✓ | ✓ |
| 13. | `owl:equivalentProperty` | 0.030 | ✘ | ✓ | ✓ |
| 14. | `owl:someValuesFrom` | 0.030 | ✘ | ~ | ~ |
| 15. | `owl:hasValue` | 0.028 | ✘ | ✓ | ✓ |

# (Linked) Vocabularies: *Interlinkage*




Powered by MONDECA

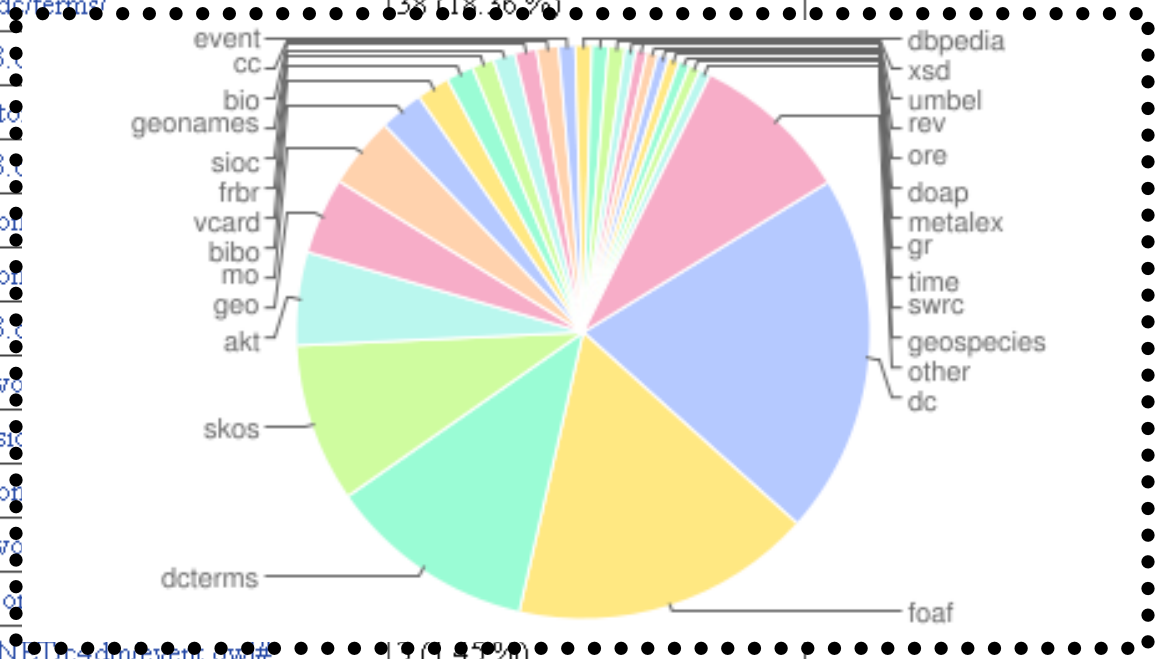
Interactive http://labs.mondeca.com/dataset/lov/; **Vatant, Vandenbussche**



NUI Galway OÉ Gaillimh · sfi Science Foundation Ireland · DERI · UNIVERSITY OF ABERDEEN

# LOD Vocabulary Usage

| Vocabulary prefix | Vocabulary link | Number of usages in data sets |
|---|---|---|
| dc | http://purl.org/dc/elements/1.1/ | 66 (31.88 %) |
| foaf | http://xmlns.com/foaf/0.1/ | 55 (26.57 %) |
| dcterms | http://purl.org/dc/terms/ | 38 (18.36 %) |
| skos | http://www.w3.o | |
| akt | http://www.akto | |
| geo | http://www.w3.o | |
| mo | http://purl.org/o | |
| bibo | http://purl.org/o | |
| vcard | http://www.w3.o | |
| frbr | http://purl.org/v | |
| sioc | http://rdfs.org/si | |
| geonames | http://www.geo | |
| bio | http://purl.org/v | |
| cc | http://creativeco | |
| event | http://purl.org/NET/c4dm/event.owl# | 3 (1.45 %) |
| dbpedia | http://dbpedia.org/resource/ | 3 (1.45 %) |
| xsd | http://www.w3.org/2001/XMLSchema# | 3 (1.45 %) |
| umbel | http://umbel.org/umbel# | 3 (1.45 %) |



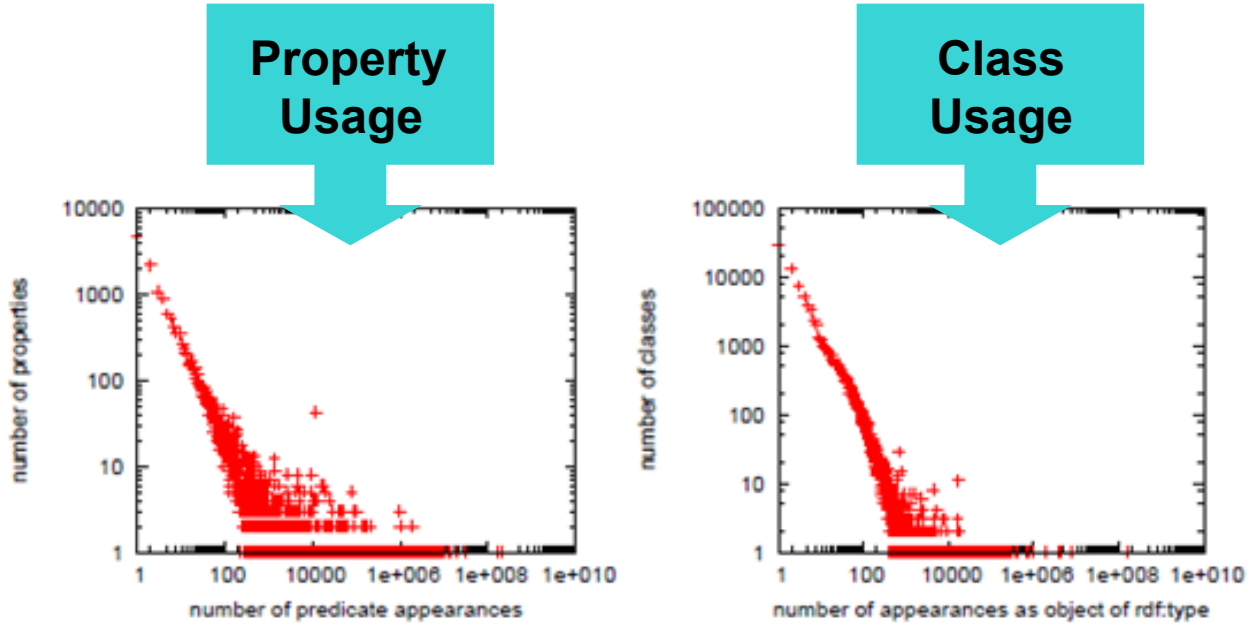

Info from http://www4.wiwiss.fu-berlin.de/lodcloud/state/ : **Bizer, Jentzsch, Cyganiak**



NUI Galway OÉ Gaillimh · sfi · DERI · UNIVERSITY OF ABERDEEN

# LOD Vocabulary Usage

- **Preferential Attachment**: ***more commonly used classes and properties are more likely to be used by others***
  - *Self-organising phenomenon/emergence*
  - *Causes power-law (long-tail) distributions...*




*log/log* scale

…who needs Linked Data reasoning?…

# LINKED DATA REASONING USE-CASE SCENARIO

# …e.g., Semantic Web Search Engine (SWSE)

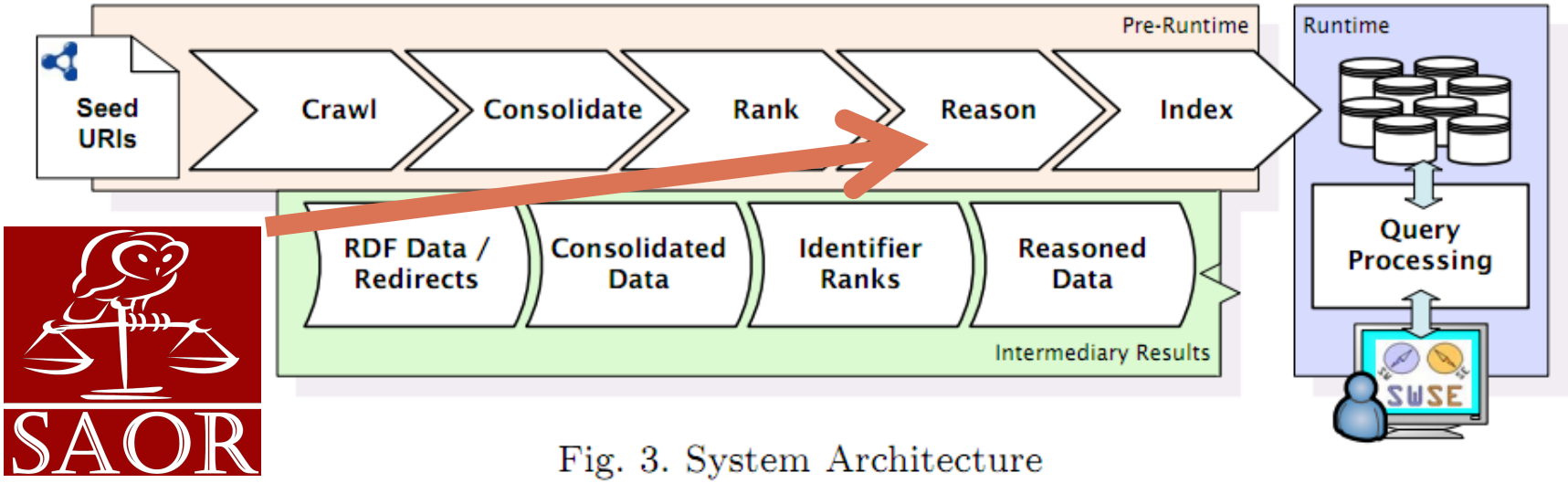



Fig. 3. System Architecture

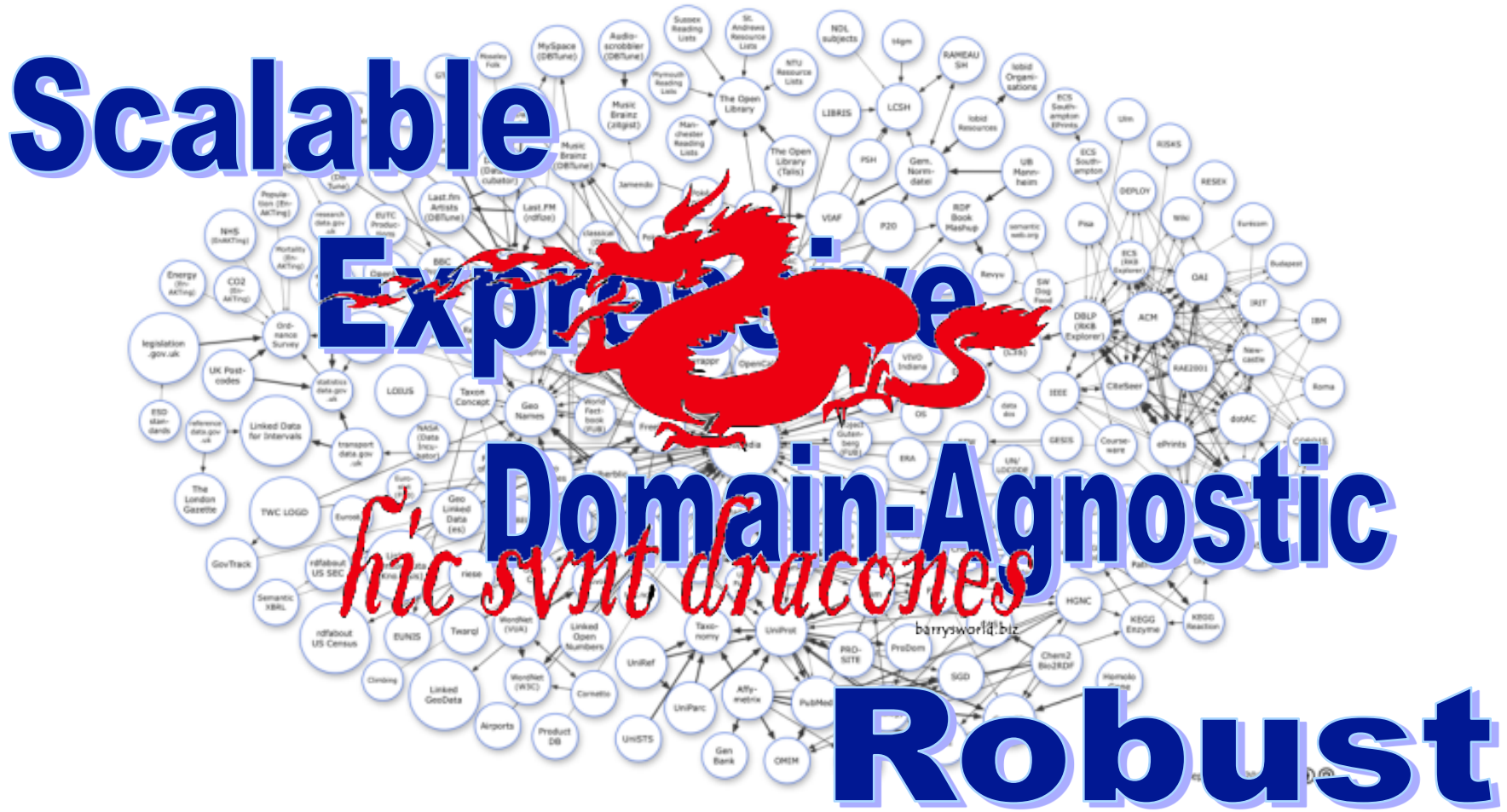
- Cyclical indexing of static Linked Data crawls
  - ***1.1 billion raw statements; 4 million RDF/XML documents; 778 pay-level-domains; open crawl***
- Want to do reasoning to integrate data

**- Hogan et al.. *Searching and Browsing Linked Data with SWSE: the Semantic Web Search Engine***
JWS (to appear), 2011.

…is it really so hard?…

# LINKED DATA REASONING CHALLENGES

# Linked Data Reasoning: **Challenges**






NUI Galway
OÉ Gaillimh

sfi Science Foundation Ireland

DERI

UNIVERSITY OF ABERDEEN

# Linked Data Reasoning: **Challenges**

## Scalability

- At least tens of billions of statements (for the moment)
  - Near linear scale!!!

## Noisy data

- Inconsistencies galore
- Publishing errors

# **Noisy Data**: Omnipotent Being

**Web data is noisy.**

***Proof:***
08445a31a78661b5c746feff39a9db6e4e2cc5cf

- sha1-sum of `'mailto:'`
- common value for `foaf:mbox_sha1sum`
  - An inverse-functional (uniquely identifying) property!!!
  - Any person who shares the same value will be considered the same

Q.E.D.

# **Noisy Data**: Redefining everything

**More proof (courtesy of http://www.eiao.net/rdf/1.0)**

```
rdf:type rdf:type owl:Property .
rdf:type rdfs:label "type"@en .
rdf:type rdfs:comment "Type of resource" .
rdf:type rdfs:domain eiao:testRun .
rdf:type rdfs:domain eiao:pageSurvey .
rdf:type rdfs:domain eiao:siteSurvey .
rdf:type rdfs:domain eiao:scenario .
rdf:type rdfs:domain eiao:rangeLocation .
rdf:type rdfs:domain eiao:startPointer .
rdf:type rdfs:domain eiao:endPointer .
rdf:type rdfs:domain eiao:header .
rdf:type rdfs:domain eiao:runs .
```

# **Noisy Data**: Inconsistency

```
w3c rdf:type foaf:Organization .
w3c rdf:type foaf:Person .
foaf:Person owl:disjointWith foaf:Organization .
```


# **Web** Reasoning: Challenges

## Challenges (Semantic Web Wikipedia Article)

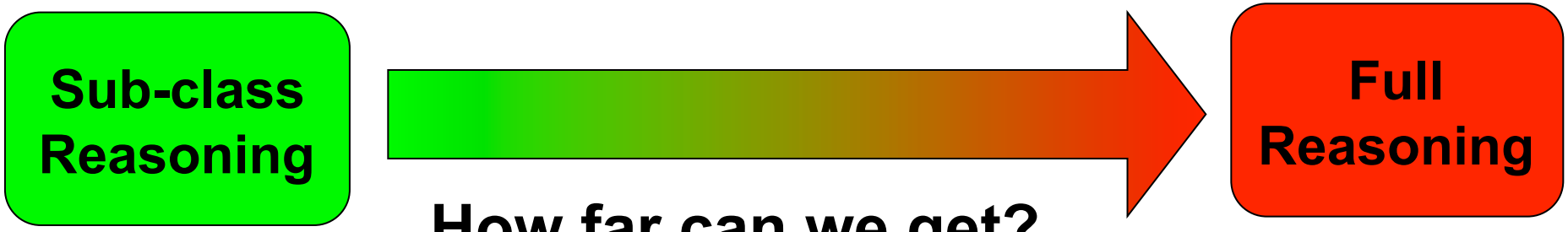
- **Some of the challenges for the Semantic Web include vastness, vagueness, uncertainty, inconsistency and deceit. Automated reasoning systems will have to deal with all of these issues in order to deliver on the promise of the Semantic Web.**
- **Vastness:** The World Wide Web contains at least 48 billion pages as of this writing (August 2, 2009). The SNOMED CT medical terminology ontology contains 370,000 class names, and existing technology has not yet been able to eliminate all semantically duplicated terms. **Any automated reasoning system will have to deal with truly huge inputs.**
- **Vagueness:** These are imprecise concepts like "young" or "tall". This arises from the vagueness of user queries, of concepts represented by content providers, of matching query terms to provider terms and of trying to combine different knowledge bases with overlapping but subtly different concepts. Fuzzy logic is the most common technique for dealing with vagueness.
- **Uncertainty:** These are precise concepts with uncertain values. For example, a patient might present a set of symptoms which correspond to a number of different distinct diagnoses each with a different probability. Probabilistic reasoning techniques are generally employed to address uncertainty.
- **Inconsistency:** These are logical contradictions which will inevitably arise during the development of large ontologies, and when ontologies from separate sources are combined. **Deductive reasoning fails catastrophically when faced with inconsistency, because "anything follows from a contradiction"**. Defeasible reasoning and paraconsistent reasoning are two techniques which can be employed to deal with inconsistency.
- **Deceit:** This is when **the producer of the information is intentionally misleading** the consumer of the information. Cryptography techniques are currently utilized to ameliorate this threat.

# So, Linked Data reasoning is not possible…

**…we should just give up.**

**Thanks for listening! Questions?**

# Scalable Reasoning: **Incomplete Reasoning**




**How far can we get?**

**…and how useful is that?**

- **When operating over Linked Data, sound but incomplete (monotonic) reasoning is better than nothing.**
  - **Open World Assumption: what's not known is unknown**
  - **Incomplete reasoning: we can know a little more**
  - **…oh yep, it's Web data…**

…we use rules for a start…

# LINKED DATA REASONING USING RULES

# RDFS and OWL 2 RL: **Entailment rules**

- **RDFS entailment rules provide sound, complete(ish) RDFS reasoning**
- **OWL 2 RL/RDF provide partial support for OWL 2 RDF-based semantics …quite expressive!**
- **Monotonic rules which are guarded**
- **Positive subset of datalog with a fixed ternary predicate**
- **Rules have cubic complexity (with trivial exceptions aside)**
  - **Due to the arity of triples (3)**

# Rules

***Body***/*Antecedent/Condition* ***Head***/*Consequent*

**IF ⇒ THEN**

***Schema***/*Terminology/ Ontological*

***Instance***/*Assertional*

```
?c1 rdfs:subClassOf ?c2 .
?x rdf:type ?c1 .
⇒ ?x rdf:type ?c2 .
```

```
foaf:Person rdfs:subClassOf foaf:Agent .
timbl:me rdf:type foaf:Person .
⇒ timbl:me rdf:type foaf:Agent .
```

# Constraint Rules (Inconsistencies)

***Body**/Antecedent/Condition* ***Head**/Consequent*

**IF ⇒ THEN**

```
?c1 owl:disjointWith ?c2 .
?x rdf:type ?c1 .
?x rdf:type ?c2 .
⇒ false
```

```
foaf:Person owl:disjointWith foaf:Organization .
w3c rdf:type foaf:Organization .
w3c rdf:type foaf:Person .
⇒ false
```

# …behind the scenes…

## …OWL 2 RDF-Based Semantics

Table 5.3: Semantic Conditions for the Vocabulary Properties

| IRI *E* | *I*(*E*) | IEXT(*I*(*E*)) |
|---|---|---|
| owl:allValuesFrom | ∈ IP | ⊆ ICEXT(*I*(owl:Restriction)) × IC |
| owl:annotatedProperty | ∈ IP | ⊆ IR × IR |
| owl:annotatedSource | ∈ IP | ⊆ IR × IR |
| owl:annotatedTarget | ∈ IP | ⊆ IR × IR |
| owl:assertionProperty | ∈ IP | ⊆ ICEXT(*I*(owl:NegativePropertyAssertion)) × IP |
| owl:backwardCompatibleWith | ∈ IOXP , ∈ IOAP | ⊆ IX × IX |
| owl:bottomDataProperty | ∈ IODP | = ∅ |
| owl:bottomObjectProperty | ∈ IP | = ∅ |
| owl:cardinality | ∈ IP | ⊆ ICEXT(*I*(owl:Restriction)) × INNI |
| rdfs:comment | ∈ IOAP | ⊆ IR × LV |
| owl:complementOf | ∈ IP | ⊆ IC × IC |

■ ■ ■

Table from http://www.w3.org/TR/owl2-rdf-based-semantics/ : **Schneider**

# …behind the scenes…
# …OWL 2 RDF-Based Semantics

- **Applicable for arbitrary RDF graphs (i.e., OWL 2 Full)**

- **Undecidable!**
  - **…but we can still do *incomplete* reasoning**
  - **…OWL 2 RL/RDF a *partial* axiomatisation of OWL 2 RDF-Based Sem.**
    - **…with known relation to Direct Semantics for OWL 2 RL profile**

# …sufficient just to see RDF transformations (…if, like me, you're not logic savy…)

## IF ⇒ THEN

$?c_1$ rdfs:subClassOf $?c_2$ .

?x rdf:type $?c_1$ .

⇒ ?x rdf:type $?c_2$ .

foaf:Person rdfs:subClassOf foaf:Agent .

timbl:me rdf:type foaf:Person .

⇒ timbl:me rdf:type foaf:Agent .

# Tableau vs. Rules

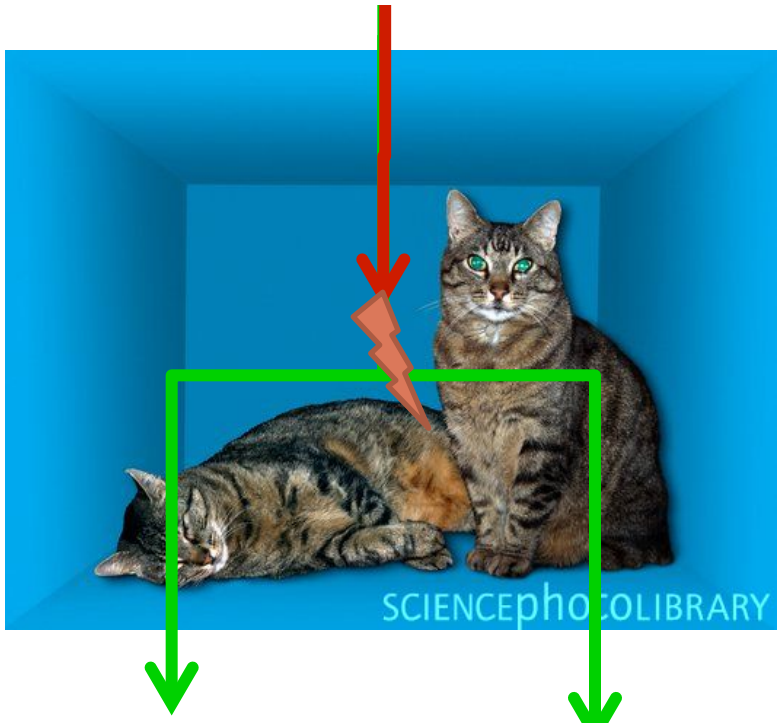
- Rules don't handle "branching cases" very well…

```
:SchrödingersCat a :QuantumCat .
:QuantumCat subClassOf [ unionOf (:Alive :Dead) ] .
```

(:QuantumCat ⊑ :Alive ⊓ :Dead)

Tableau

Rules




- …also existentials not well supported…
- You lose some expressivity with rules!
- …also some reasoning tasks…

# **Linked Data vocabs**: top 15 RDFS/OWL features

| # | Axiom | Rank(Σ) | RDFS | Horst | O2R |
|---|---|---|---|---|---|
| 1. | `rdfs:subClassOf` | 0.295 | ✓ | ✓ | ✓ |
| 2. | `rdfs:range` | 0.294 | ✓ | ✓ | ✓ |
| 3. | `rdfs:domain` | 0.292 | ✓ | ✓ | ✓ |
| 4. | `rdfs:subPropertyOf` | 0.090 | ✓ | ✓ | ✓ |
| 5. | `owl:FunctionalProperty` | 0.063 | ✘ | ✓ | ✓ |
| 6. | `owl:disjointWith` | 0.049 | ✘ | ✘ | ✓ |
| 7. | `owl:inverseOf` | 0.047 | ✘ | ✓ | ✓ |
| 8. | `owl:unionOf` | 0.035 | ✘ | ✘ | ~ |
| 9. | `owl:SymmetricProperty` | 0.033 | ✘ | ✓ | ✓ |
| 10. | `owl:TransitiveProperty` | 0.030 | ✘ | ✓ | ✓ |
| 11. | `owl:equivalentClass` | 0.021 | ✘ | ✓ | ✓ |
| 12. | `owl:InverseFunctionalProperty` | 0.030 | ✘ | ✓ | ✓ |
| 13. | `owl:equivalentProperty` | 0.030 | ✘ | ✓ | ✓ |
| 14. | `owl:someValuesFrom` | 0.030 | ✘ | ~ | ~ |
| 15. | `owl:hasValue` | 0.028 | ✘ | ✓ | ✓ |

# Rules vs. Tableau

- RDFS or OWL 2 RL/RDF rules can be applied to any RDF graph
  - Don't need to translate back into DL formulae
  - OWL 2 RL/RDF based semantics defined directly for RDF
  - Remember: a lot of Linked Data vocabs are OWL Full!
    - `Datatype`/`ObjectProperty`, `Class` declarations missing
    - `InverseFunctionalProperty` / `DatatypeProperty`
    - Extending core vocab. (e.g., `rdfs:label`)
- Entailment not based on satisfiability
  - Positive/monotonic rules
  - Inconsistencies can be "overlooked"
- Tractable! Easier to implement (for non-logicians). More intuitive (for non-logicians).
- *Jeff to talk about scalable approximations for OWL 2 DL later…*

…rules aren't a magic bullet…

# ROBUST LINKED DATA REASONING

# Authoritative Reasoning

**Consider source of schema data**

- Class/property URIs dereference to their authoritative document
  - FOAF spec authoritative for `foaf:Person` ✓
  - MY spec not authoritative for `foaf:Person` ✘
- Allow "extension" in third-party documents
  - `my:Person rdfs:subClassOf foaf:Person .` **(MY spec)** ✓
- BUT: Reduce obscure memberships
  - `foaf:Person rdfs:subClassOf my:Person .` (MY spec) ✘
- ALSO: Protect specifications
  - `foaf:knows a owl:SymmetricProperty .` (MY spec) ✘

# Authoritative Reasoning

**OWL 2 RL rule `prp-inv1`**

```
?p1 owl:inverseOf ?p2 .
?x ?p1 ?y .
⇒ ?y ?p2 ?x .
```

**OWL 2 RL rule `prp-inv2`**

```
?p1 owl:inverseOf ?p2 .
?x ?p2 ?y .
⇒ ?y ?p1 ?x .
```

**TBOX / schema:**

```
foo:doesntKnow owl:inverseOf
   foaf:knows .
```
(from `foo:`) ✗

**ABOX / instances:**

```
bar:Aidan foo:doesntKnow
    bar:Axel .
bar:Stefan foaf:knows bar:Jim .
```

**AUTHORITATIVE INFERENCE:**

```
bar:Axel foaf:knows bar:Aidan .
bar:Jim foo:doesntKnow bar:Stefan .
```

# Example non-authoritative inferences...

`foaf:Person`

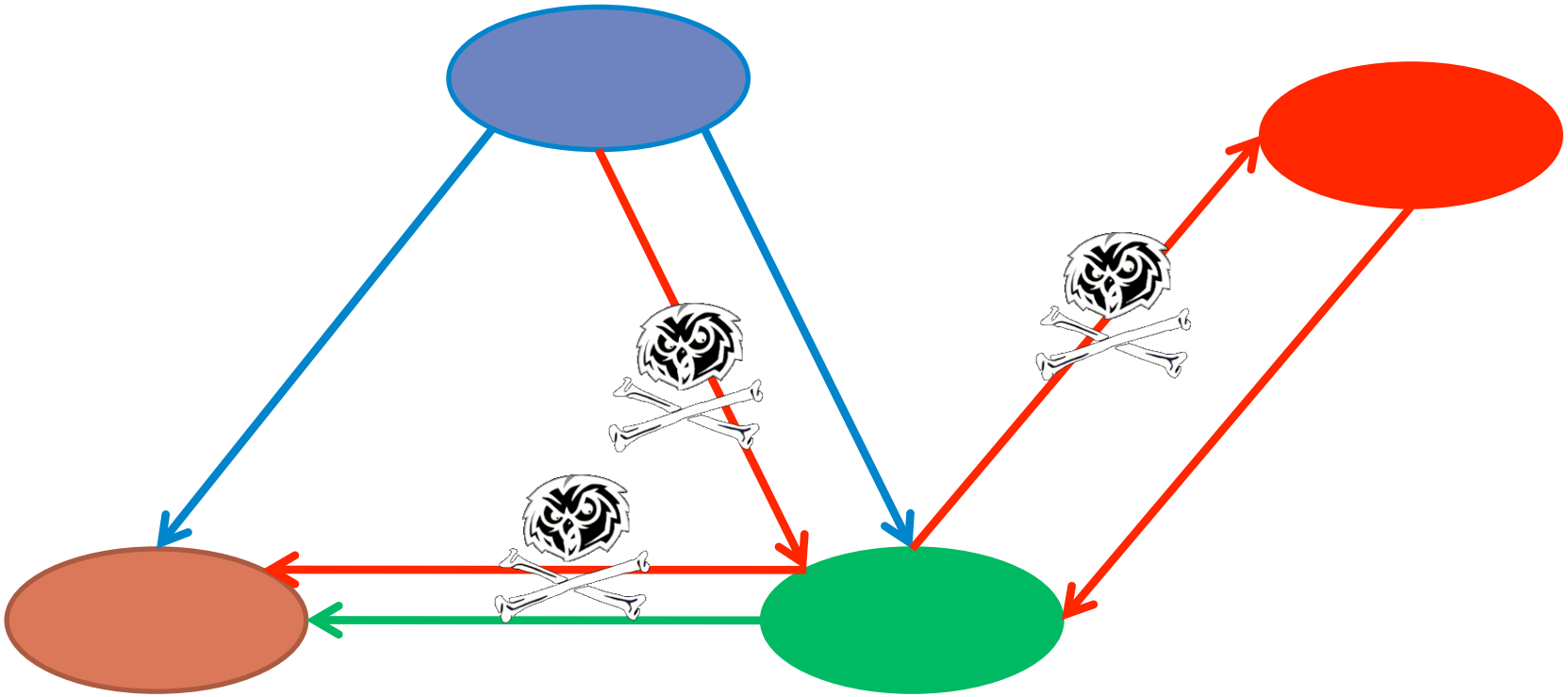
**164 million instances**

- **5 authoritative inferences**

- **26 additional non-authoritative inferences (excluding 100s possible through `rdfs:Resource`)**
  - **14 anonymous classes**
  - **12 named classes**

| Class | (Raw) Count |
|---|---|
| *Authoritative* | |
| foaf:Agent | 8,165,989 |
| wgs84:SpatialThing | 64,411 |
| contact:Person | 1,704 |
| dct:Agent | 35 |
| contact:SocialEntity | 1 |
| *Non-Authoritative* (additional) | |
| po:Person | 852 |
| wn:Person | 1 |
| aifb:Kategorie-3AAIFB | 0 |
| b2r2008:Controlled_vocabularies | 0 |
| foaf:Friend_of_a_friend | 0 |
| frbr:Person | 0 |
| frbr:ResponsibleEntity | 0 |
| pres:Person | 0 |
| po:Category | 0 |
| sc:Agent_Generic | 0 |
| sc:Person | 0 |
| wn:Agent-3 | 0 |

# **Authoritative** Reasoning



= schema vocab  ⟶ = defines translation to

# Noisy Data: ~~Redefining everything~~

**More proof (courtesy of http://www.eiao.net/rdf/1.0)**

```
rdf:type rdf:type owl:Property .
rdf:type rdfs:label "type"@en .
rdf:type rdfs:comment "Type of resource" .
rdf:type rdfs:domain eiao:testRun .
rdf:type rdfs:domain eiao:question .
rdf:type rdfs:domain eiao:subject .
rdf:type rdfs:domain eiao:scenario .
rdf:type rdfs:domain eiao:rangeLocation .
rdf:type rdfs:domain eiao:startPointer .
rdf:type rdfs:domain eiao:endPointer .
rdf:type rdfs:domain eiao:header .
rdf:type rdfs:domain eiao:runs .
```

Not Authoritative

# **Authoritative** Reasoning: read more …w/ essential plugs

Gong Cheng, Yuzhong Qu.

"*Integrating Lightweight Reasoning into Class-Based Query Refinement for Object Search*." ASWC 2008.

Aidan Hogan, Andreas Harth, Axel Polleres.

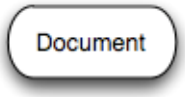
"*Scalable Authoritative OWL Reasoning for the Web*." IJSWIS 2009.

Aidan Hogan, Jeff Z. Pan, Axel Polleres and Stefan Decker.

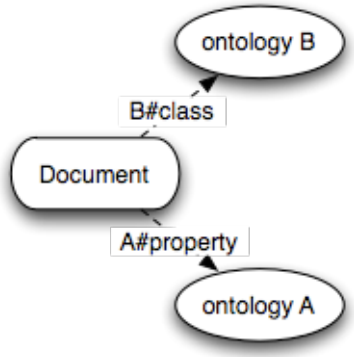
"*SAOR: Template Rule Optimisations for Distributed Reasoning over 1 Billion Linked Data Triples*." ISWC 2010.

My thesis: http://aidanhogan.com/docs/thesis/

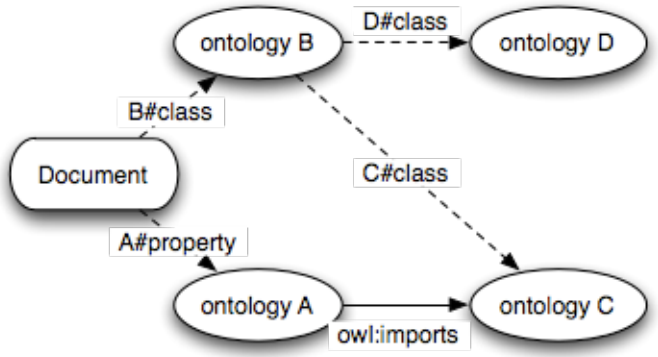
# **Quarantined** Reasoning [Delbru et al.; 2008]



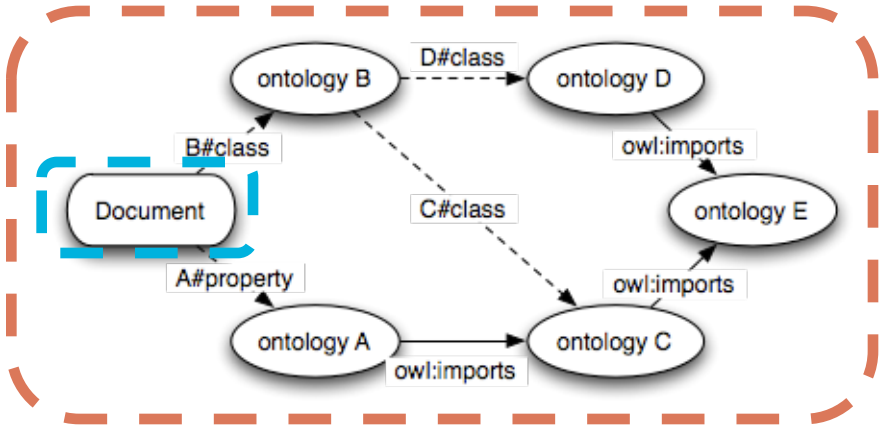
# **Quarantined** Reasoning [Delbru et al.; 2008]



# **Quarantined** Reasoning [Delbru et al.; 2008]



# **Quarantined** Reasoning [Delbru et al.; 2008]




**A-Box / Instance Data**
**(e.g, a FOAF file)**

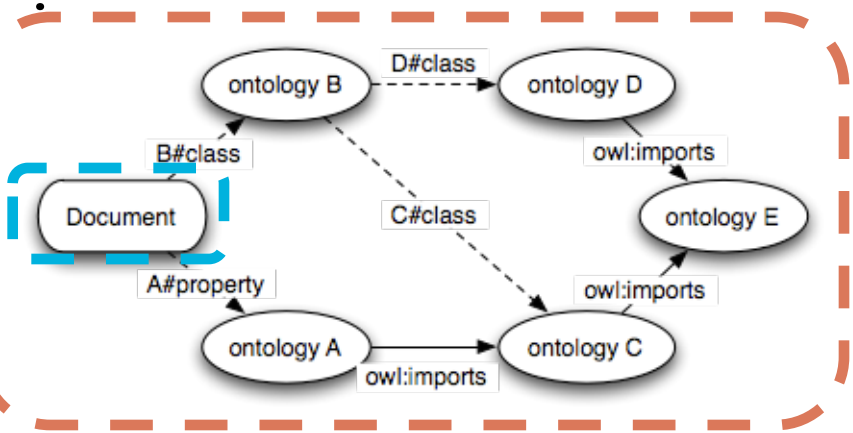
**T-Box / Ontology Data**
**(e.g., the FOAF ontology and its indirect imports)**

# Noisy Data: ~~Redefining everything~~

**More proof (courtesy of http://www.eiao.net/rdf/1.0)**

```
rdf:type rdf:type owl:Property .
rdf:type rdfs:label "type"@en .
rdf:type rdfs:comment "Type c
rdf:type rdfs:domain eiao:tes
rdf:type rdfs:domain eiao:sit
rdf:type rdfs:domain eiao:sce
rdf:type rdfs:domain eiao:rangeLocation .
rdf:type rdfs:domain eiao:startPointer .
rdf:type rdfs:domain eiao:endPointer .
rdf:type rdfs:domain eiao:header .
rdf:type rdfs:domain eiao:runs .
```

Not In Here



# **Quarantined** Reasoning: read more

R. Delbru, A. Polleres, G. Tummarello and S. Decker.

"*Context Dependent Reasoning for Semantic Documents in Sindice*. " 4th International Workshop on Scalable Semantic Web Knowledge Base Systems, 2008.

**Upcoming talk at RR 2011!**

# Resolving Inconsistency?

1. **Use links-analysis (PageRank) to rank documents and triples**
2. **Use annotated reasoning to rank inferences**
3. **Repair each consistency by removing the weakest triple**

**Read more:**

Piero A. Bonatti, Aidan Hogan, Axel Polleres and Luigi Sauro. "*Robust and Scalable Linked Data Reasoning Incorporating Provenance and Trust Annotations*". In the Journal of Web Semantics (in press).

…we're gonna' need a bigger boat…

# SCALABLE LINKED DATA REASONING

# Materialisation

**Forward-chaining Materialisation**

- Avoid runtime expense
  - Users taught impatience by Google
- Pre-compute for quick retrieval
- Web-scale systems should scale well
  - More data = more disk-space/machines

**One size does not fit all!**

**Don't materialise too much!**

**INPUT:**
- **Flat file of triples (quads)**

**OUTPUT:**
- **Flat file of (partial) inferred triples (quads)**

# …**cautious subset** of OWL 2 RL/RDF rules

- **OWL 2 RL/RDF rules are tractable! ☺**
- **OWL 2 RL/RDF rules are cubic! ☹**
- **Two triples give cubic inferences:**
  - `owl:sameAs owl:sameAs rdf:type .`
  - `rdf:type rdfs:domain owl:Thing .`
  - …homework to figure out why…
- **Also contains rules like `eq-ref`:**
  - `?s ?p ?o .`

    `⇒ ?s owl:sameAs ?s . ?p owl:sameAs ?p . ?o owl:sameAs ?o .`

# Scalable Reasoning: **In-mem T-Box**

- **Main optimisation: Store T-Box/schemata in memory**
- **T-Box: (loosely) data describing classes and properties.**
  - Aka. schemata/vocabularies/ontologies/terminologies.
  - E.g.,
    - `foaf:topic owl:inverseOf foaf:page .`
    - `sioc:UserAccount rdfs:subClassOf foaf:OnlineAccount .`
- **Most commonly accessed data for reasoning**
- **Quite small (~0.1% for our Linked Data corpus)**
  - High selectivity (if you prefer)
- **A-Box:** Lots `?s foaf:page ?o .` VS.
- **T-Box:** Few `foaf:page ?p ?o .` + `?s ?p foaf:page .`

# Scalable Reasoning: **Two Scans**

- **Scan 1:** Scan input data separate T-Box statements, load T-Box statements into memory
  - **Do T-Box level reasoning if required**
- **Scan 2:** Scan all on-disk data, join with in-memory T-Box.

# Scalable Reasoning: **No A-Box Joins**

- **Execution of three rules:**

**OWL 2 RL/RDF rule `prp-rng`**

`?p rdfs:range ?c .`

`?x ?p ?y .`

⇒ `?y a ?c .`

**OWL 2 RL/RDF rule `prp-spo1`**

$?p_1$ `rdfs:subPropertyOf` $?p_2$ `.`

`?x` $?p_1$ `?y .`

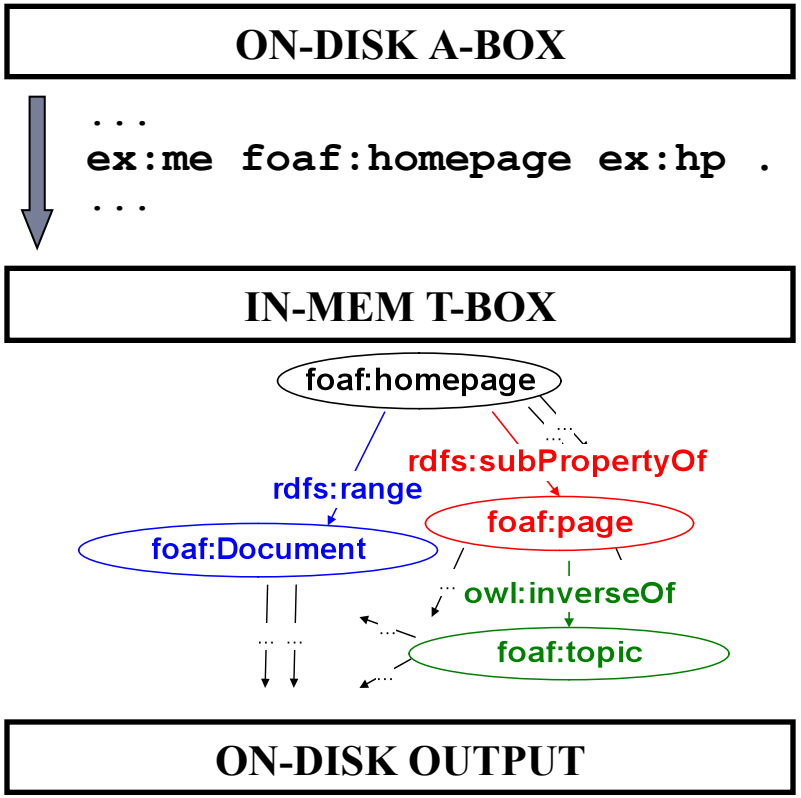
⇒ `?x` $?p_2$ `?y .`

**OWL 2 RL/RDF rule `prp-inv1`**

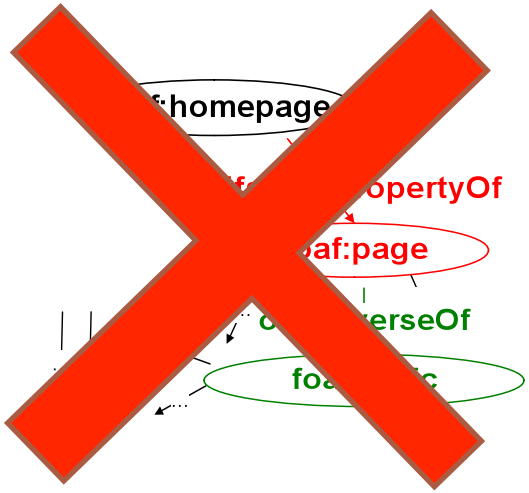
$?p_1$ `owl:inverseOf` $?p_2$ `.`

`?x` $?p_1$ `?y .`

⇒ `?y` $?p_2$ `?x .`



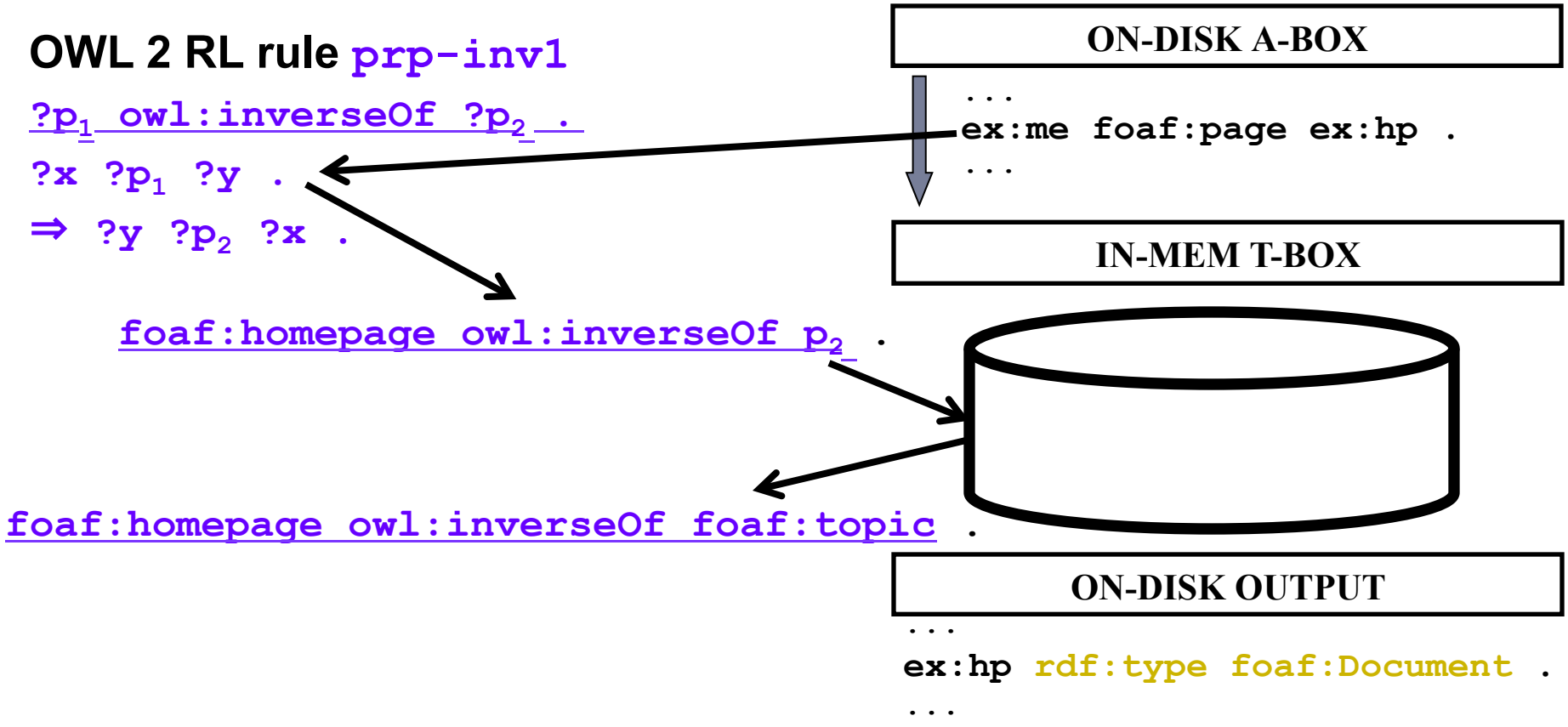
# Hard-coded T-Box graph




- Instead use generic optimisations…

# Baseline…

## Apply each supported OWL 2 RL/RDF rule to each triple during scan



# Baseline…

- 118 hours
- Infer 1.58 billion “raw” quadruples
  - 1.14 after filtering
  - 962 million novel/unique triples

# Partially evaluate schema in rules…

**OWL 2 RL rule prp-inv1**

$?p_1$ *owl:inverseOf* $?p_2$ .

?x $?p_1$ ?y . ⇒ ?y $?p_2$ ?x .

**OWL 2 RL rule prp-inv1**

foaf:primaryTopic *owl:inverseOf* foaf:isPrimaryTopicOf .

foaf:topic *owl:inverseOf* foaf:page .

?x foaf:topic ?y . ⇒ ?y foaf:page ?x .

---

**OWL 2 RL rule prp-spo1**

$?p_1$ *rdfs:subPropertyOf* $?p_2$ .

?x $?p_1$ ?y. ⇒ ?x $?p_2$ ?y .

**OWL 2 RL rule prp-spo1**

foaf:homepage *rdfs:subPropertyOf* foaf:isPrimaryTopicOf .

?x foaf:isPrimaryTopicOf ?y . ⇒ ?x foaf:page ?y .

foaf:isPrimaryTopicOf *rdfs:subPropertyOf* foaf:page .

?x foaf:primaryTopic ?y . ⇒ ?x foaf:topic ?y .

foaf:primaryTopic *rdfs:subPropertyOf* foaf:topic .

# Naïve: **Apply all rules to all triples**

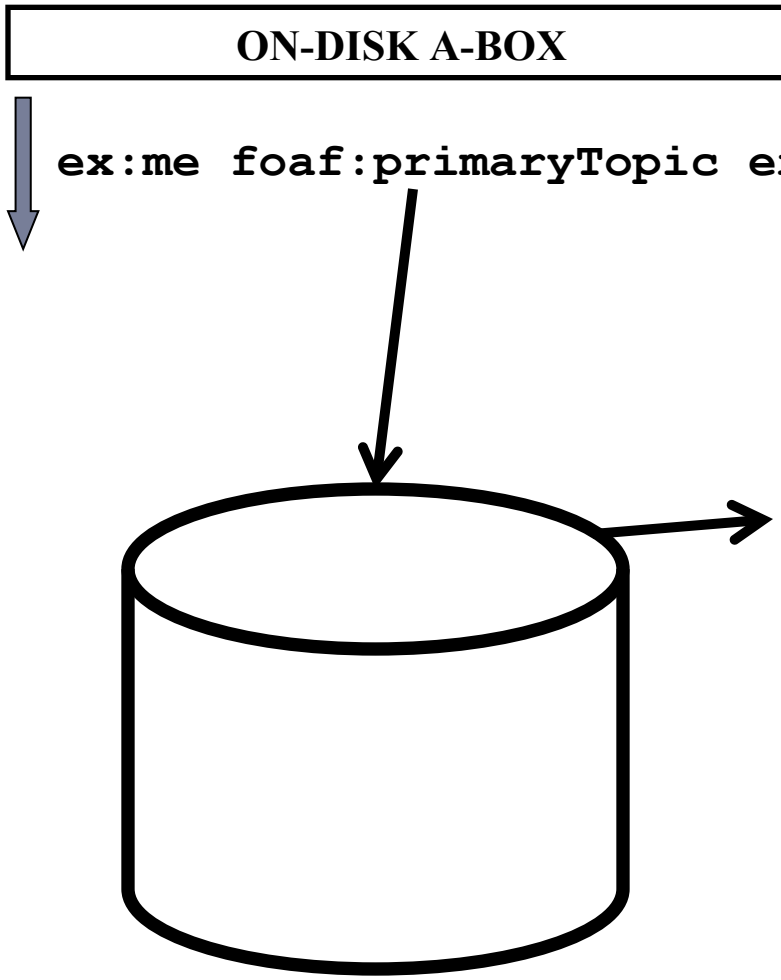
- **Apply all partially evaluated rules to all triples**
- **…erm**

- **We generated 301k such rules…**
  - **Will need to run over 2.4 billion triples**
  - **Requires 750 trillion (simple) rule applications**
    - **…erm**

**Estimated at 19 years**

**(114940% of baseline)**

# Optimisation: **In-memory Rule Index**




**OWL 2 RL rule** `prp-inv1`

```
?x foaf:primaryTopic ?y .
⇒ ?y foaf:isPrimaryTopicOf ?x .
```

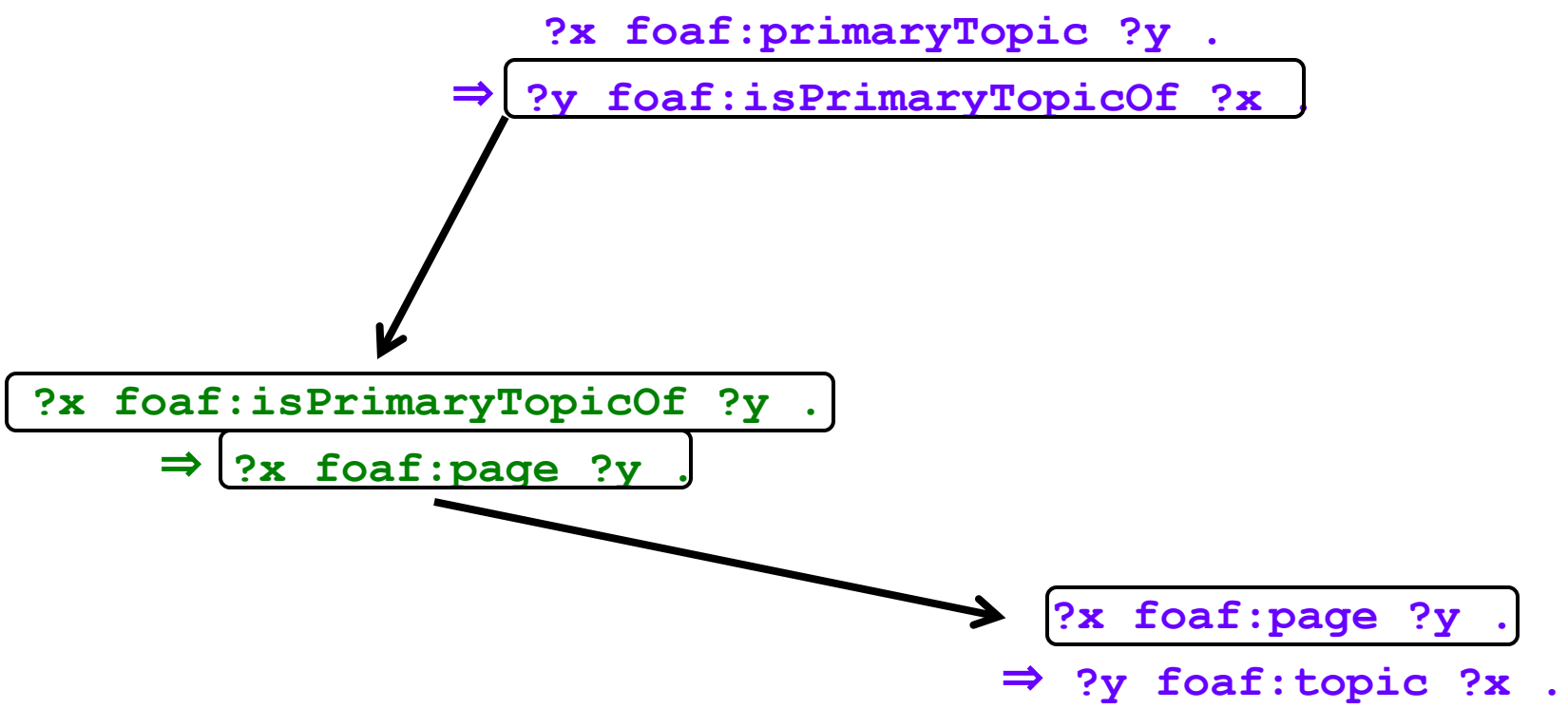
```
?x foaf:topic ?y .
⇒ ?y foaf:page ?x .
```

**OWL 2 RL rule** `prp-spo1`

```
?x foaf:homepage ?y.
⇒ ?x foaf:isPrimaryTopicOf ?y .
```

```
?x foaf:primaryTopic ?y .
⇒ ?x foaf:topic ?y .
```

# Optimisation: Linked Rule Index



# Optimisation: Linked Rule Index

**22.1 hours**

**(19% runtime of baseline)**

# Optimisation: **Merge** Rules

```
?y foaf:primaryTopic ?x .
⇒ ?x foaf:isPrimaryTopicOf ?y .
```

+

```
?x foaf:primaryTopic ?y .
⇒ ?x foaf:topic ?y .
```

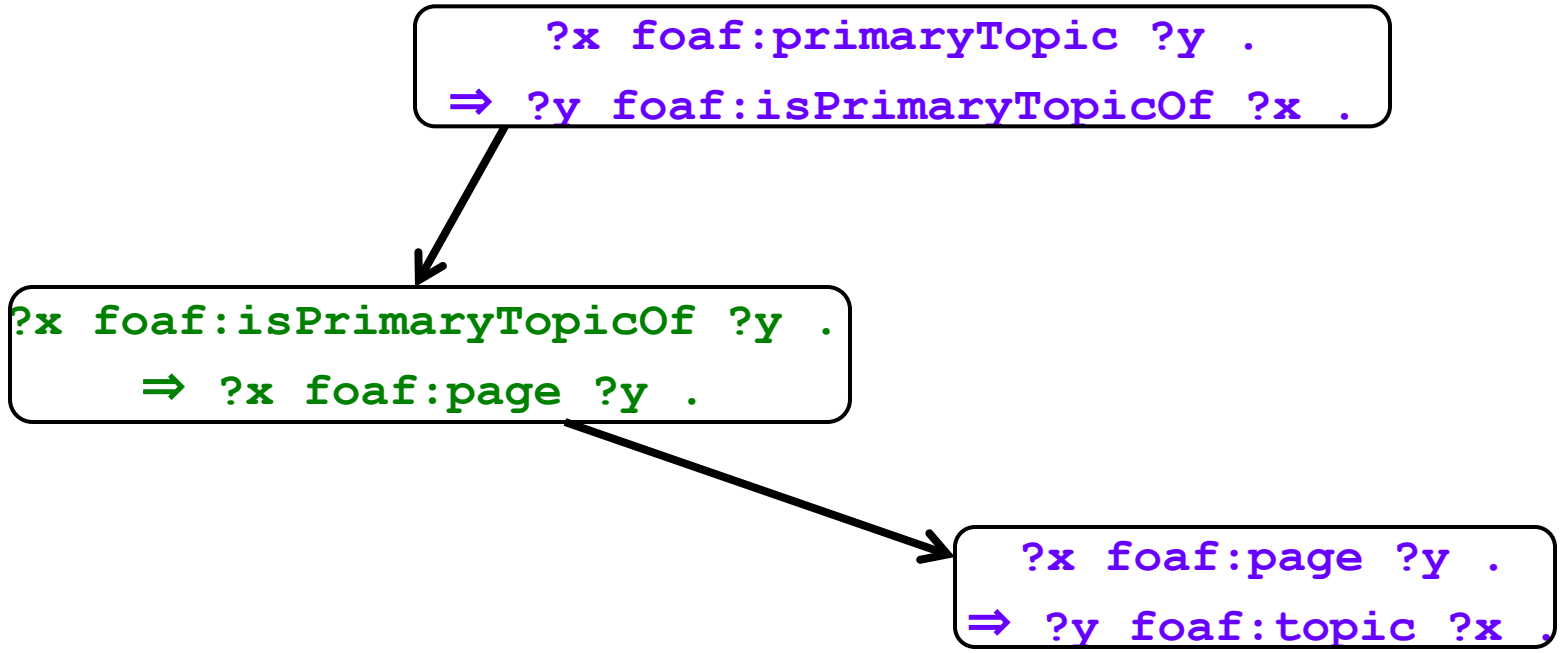
```
?y foaf:primaryTopic ?x .
⇒ ?x foaf:isPrimaryTopicOf ?y .
?y foaf:topic ?x .
```

# Optimisation: **Merged** Linked Rule Index

**17.7 hours**

**(16.5% runtime of baseline)**

# Failed Optimisation: "Saturate" rules




```
?x foaf:primaryTopic ?y .
⇒ ?y foaf:isPrimaryTopicOf ?x .
    ?y foaf:page ?x .
    ?x foaf:topic ?y .
```

```
?x foaf:isPrimaryTopic ?y .
⇒ ?x foaf:page ?y .
  ?y foaf:topic ?x .
```

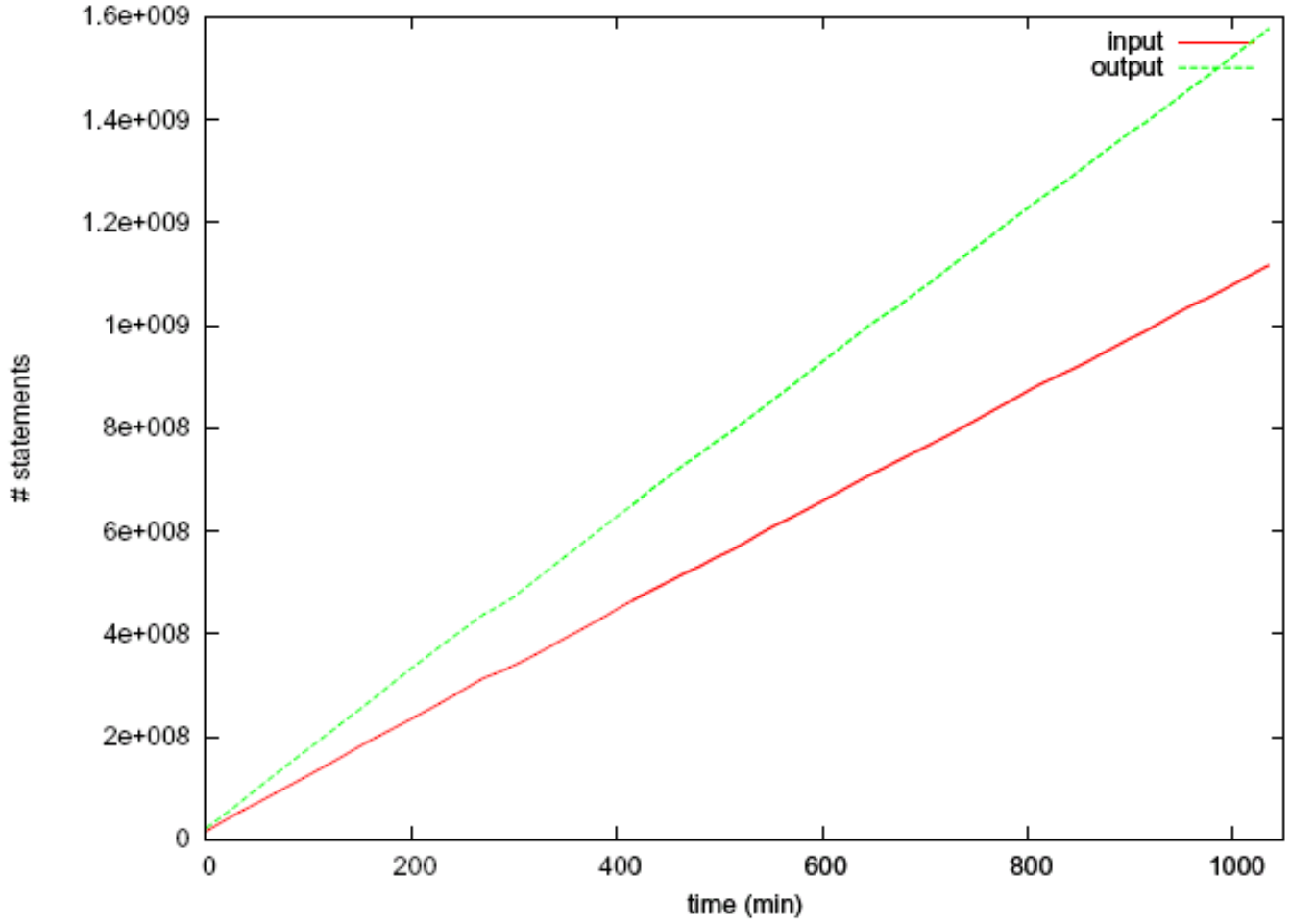
# Failed Optimisation:
## "Saturated" Merged Linked Rule Index

**19.5 hours**

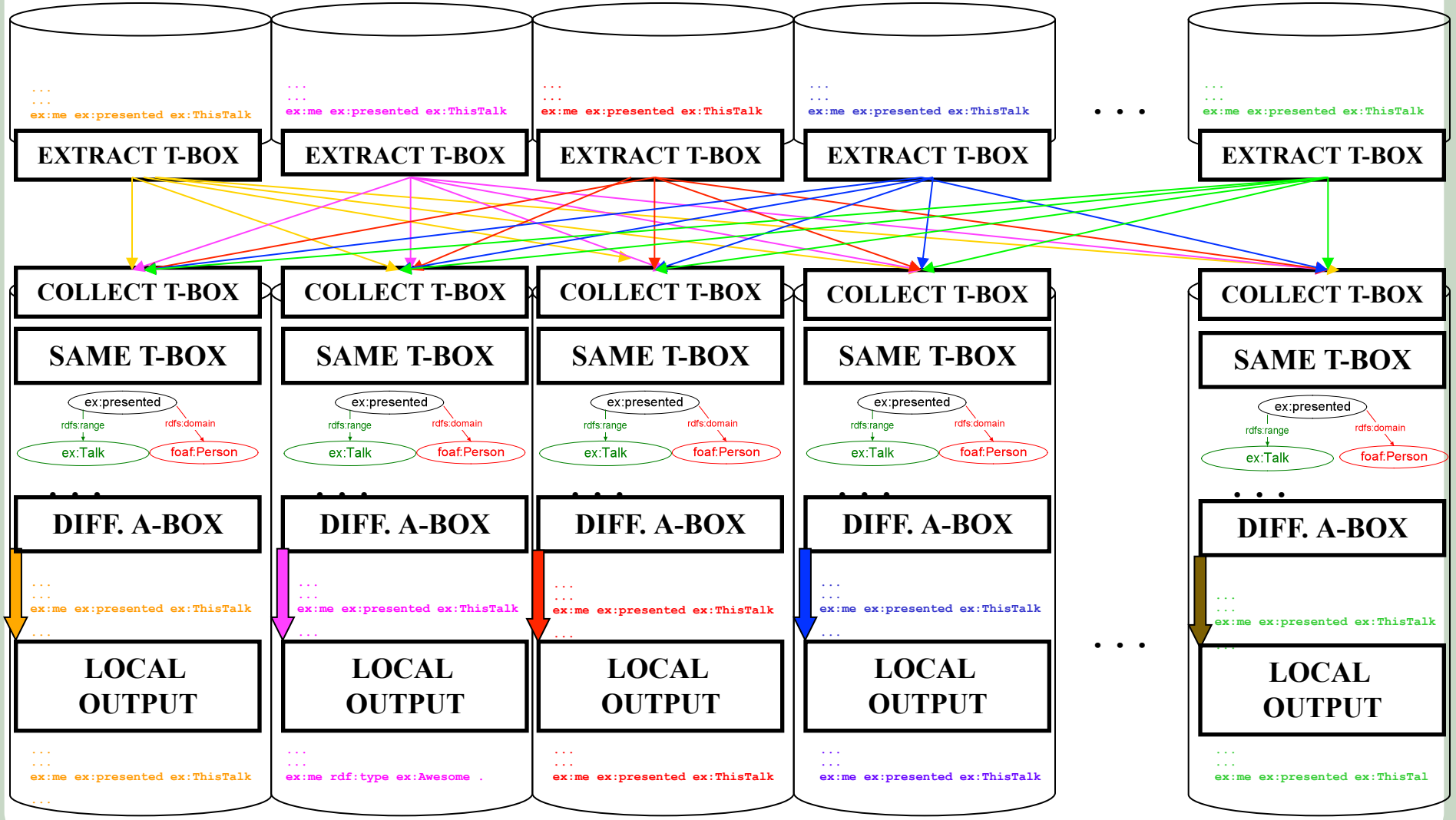
**(15% runtime of baseline)**

**110% runtime without saturation**

# Reasoning Performance (1 machine)



# Scalable **Distributed** Reasoning






NUI Galway OÉ Gaillimh · sfi Science Foundation Ireland · DERI · 1495 UNIVERSITY OF ABERDEEN

# Scalable **Distributed** Reasoning

- **Eight machines, 4GB main memory, 2.2 GHz**

| *Machines* | *Extract T-Box* | *Build T-Box* | *Reason A-Box* | *Total* |
|---|---|---|---|---|
| 1 | 492 | 8.9 | 1062 | 1565 |
| 2 | 240 | 10.2 | 465 | 719 |
| 4 | 131 | 10.4 | 239 | 383 |
| 8 | 67 | 9.8 | 121 | 201 |

**minutes**

- **Fastest:**

**8 machines: Total 3.35 hours**

# **Distributed** Reasoning: read more


Aidan Hogan, Jeff Z. Pan, Axel Polleres, Stefan Decker: SAOR: Template Rule Optimisations for Distributed Reasoning over 1 Billion Linked Data Triples. International Semantic Web Conference (1) 2010: 337-353

Jesse Weaver, James A. Hendler: Parallel Materialization of the Finite RDFS Closure for Hundreds of Millions of Triples. International Semantic Web Conference 2009: 682-697

Jacopo Urbani, Spyros Kotoulas, Eyal Oren, Frank van Harmelen: Scalable Distributed Reasoning Using MapReduce. International Semantic Web Conference 2009: 634-649

# **Linked Data vocabs**: top 15 RDFS/OWL features

| # | Axiom | Rank(Σ) | RDFS | Horst | O2R |
|---|---|---|---|---|---|
| 1. | rdfs:subClassOf | 0.295 | ✓ | ✓ | ✓ |
| 2. | rdfs:range | 0.294 | ✓ | ✓ | ✓ |
| 3. | rdfs:domain | 0.292 | ✓ | ✓ | ✓ |
| 4. | rdfs:subPropertyOf | 0.090 | ✓ | ✓ | ✓ |
| 5. | owl:FunctionalProperty | 0.063 | ✘ | ✓ | ✓ |
| 6. | owl:disjointWith | 0.049 | ✘ | ✘ | ✓ |
| 7. | owl:inverseOf | 0.047 | ✘ | ✓ | ✓ |
| 8. | owl:unionOf | 0.035 | ✘ | ✘ | ~ |
| 9. | owl:SymmetricProperty | 0.033 | ✘ | ✓ | ✓ |
| 10. | owl:TransitiveProperty | 0.030 | ✘ | ✓ | ✓ |
| 11. | owl:equivalentClass | 0.021 | ✘ | ✓ | ✓ |
| 12. | owl:InverseFunctionalProperty | 0.030 | ✘ | ✓ | ✓ |
| 13. | owl:equivalentProperty | 0.030 | ✘ | ✓ | ✓ |
| 14. | owl:someValuesFrom | 0.030 | ✘ | ~ | ~ |
| 15. | owl:hasValue | 0.028 | ✘ | ✓ | ✓ |

# **Linked Data vocabs**: top 15 RDFS/OWL features

- **...summary please?**

  **Adding up the ranks of all vocabularies our rules *fully* support gives 77% of the total rank of all vocabularies**

  **Adding up the ranks of all vocabularies our authoritative rules *fully* support gives 70% of the total rank of all vocabularies**

  **The highest ranked document our rules do not fully support was 5th overall: SKOS**

  **The highest ranked document with non-authoritative axioms was 7th overall: FOAF**

# Scalable Reasoning: **A-Box joins?**

- **However: some rules do require A-Box joins**
  - `?p a owl:TransitiveProperty .` `?x ?p ?y .` `?y ?p z .`

    ⇒ `?x ?p ?z .`
  - Difficult to engineer a scalable solution (which reaches a fixpoint) for Linked Data (?)
  - **Can lead to quadratic inferences, even for small T-Box**
- **A lot of useful reasoning still possible without A-Box joins…**

# **Distributed** Reasoning: read more

Jacopo Urbani, Spyros Kotoulas, Jason Maassen, Frank van Harmelen, Henri E. Bal: OWL Reasoning with WebPIE: Calculating the Closure of 100 Billion Triples. ESWC (1) 2010: 213-227

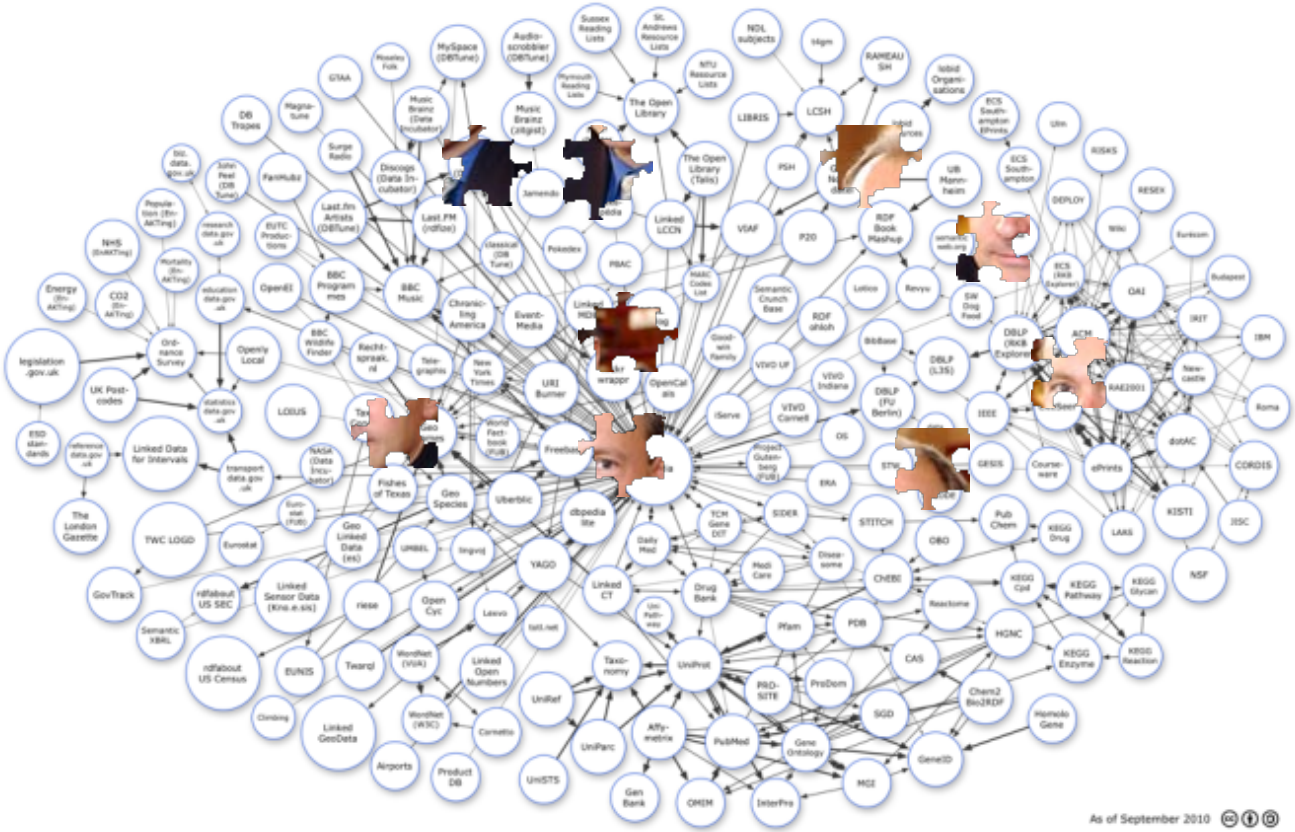
**A-Box Joins**

# “BONUS” MATERIAL

…what about `owl:sameAs`?…

# SCALABLE CONSOLIDATION

# Consolidation for Linked Data



# Consolidation: **Baseline**

- **Use provided `owl:sameAs` mappings in the data**

```
timbl:i owl:sameas identica:45563 .
dbpedia:Berners-Lee owl:sameas identica:45563 .
```

- **Store "equivalences" found**

```
timbl:i                 ->
identica:45563          ->
dbpedia:Berners-Lee     ->
```

```
timbl:i
identica:45563
dbpedia:Berners-Lee
```

# Consolidation: **Baseline**

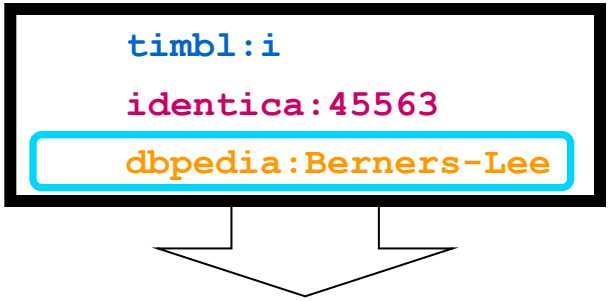
- **For each set of equivalent identifiers, choose a canonical term**

```
timbl:i
identica:45563
dbpedia:Berners-Lee
```

# Canonicalisation

- Afterwards, rewrite identifiers to their canonical version:

```
timbl:i rdf:type foaf:Person .
identica:48404 foaf:knows identica:45563 .
dbpedia:Berners-Lee dpo:birthDate "1955-06-08"^^xsd:date .
```




```
dbpedia:Berners-Lee rdf:type foaf:Person .
identica:48404 foaf:knows dbpedia:Berners-Lee .
dbpedia:Berners-Lee dpo:birthDate "1955-06-08"^^xsd:date .
```

# **Extended** Consolidation

- **Infer `owl:sameAs` through reasoning (OWL 2 RL/RDF)**
  1. **explicit `owl:sameAs` (again)**
  2. **`owl:InverseFunctionalProperty`**
  3. **`owl:FunctionalProperty`**
  4. **`owl:cardinality 1` / `owl:maxCardinality 1`**

```
foaf:homepage a owl:InverseFunctionalProperty .
timbl:i foaf:homepage w3c:timblhomepage .
adv:timbl foaf:homepage w3c:timblhomepage .
```

⇒

```
timbl:i owl:sameas adv:timbl .
```

**…then apply consolidation as before**

# Consolidation: **Results**

**For our Linked Data corpus:**

1. **~12 million explicit `owl:sameAs` triples (as before)**
2. **~8.7 million thru. `owl:InverseFunctionalProperty`**
3. **~106 thousand thru. `owl:FunctionalProperty`**
4. **none thru. `owl:cardinality`/`owl:maxCardinality`**

**In terms of equivalences found (baseline vs. extended):**

- **~2.8 million sets of equivalent identifiers**
  - **(1.31x baseline)**
- **~14.86 million identifiers involved**
  - **(2.58x baseline)**
- **~5.8 million URIs**
  - **!!(1.014x baseline)!!**

…and finally…

# CLEANING UP SOME INCONSISTENCIES

# Cannot compute…

$?c_1$ owl:disjointWith $?c_2$ .

?x rdf:type $?c_1$ .

?x rdf:type $?c_2$ .

⇒ false

foaf:Person owl:disjointWith foaf:Organization .

timbl:i rdf:type foaf:Person .

timbl:i rdf:type foaf:Organization .

⇒ false

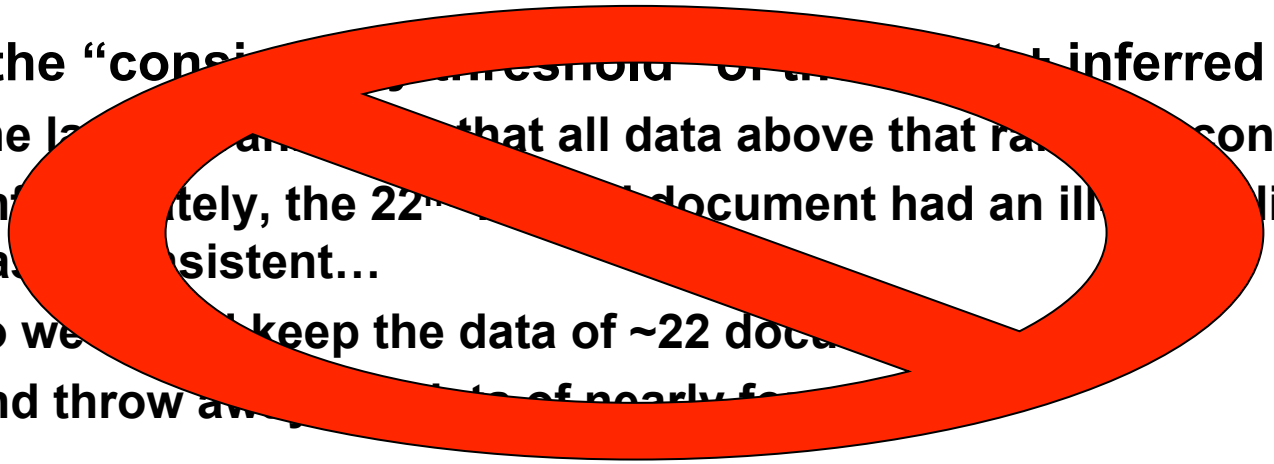
# Use ranking…

- **Apply PageRank over the documents in the Linked Data corpus**
  - **~measure of how well-linked/important documents are**
- **Assign scores to triples/inferences and use annotated reasoning**
  - **min**-based aggregation: inference assigned lowest score of triple or rule involved in proof
- **Use scores to decide which sources should be "trusted" in the event of inconsistency**

# Fixing inconsistencies

**Considered two approaches:**

1. **Find the "cons… threshold" of t… inferred data:**
   - **The la… that all data above that ra… consistent**
   - **Unf…tely, the 22nd… document had an ill… literal, and so wa… sistent…**
   - **So we… keep the data of ~22 doc…**
   - **And throw aw… …s of nearly fo…**



# Fixing inconsistencies

**Time for Plan B:**

2. **Perform a "granular" repair of the data**
   - **Remove the weakest triple causing each contradiction**

```
foaf:Person owl:disjointWith foaf:Organization 0.3 .
timbl:i rdf:type foaf:Person 0.007 .
timbl:i rdf:type foaf:Organization 0.002 .
```

# Inconsistencies found

- ~294k ill-typed datatypes
- ~7k members of disjoint classes

| Class 1 | Class 2 | Violations |
|---|---|---|
| ~~foaf:Agent~~ | ~~foaf:Document~~ | ~~3,842~~ |
| ~~foaf:Document~~ | ~~foaf:Person~~ | ~~2,918~~ |
| sioc:Container | sioc:Item | 128 |
| foaf:Person | foaf:Project | 100 |
| ecs:Group | ecs:Individual | 38 |
| skos:Concept | skos:Collection | 36 |
| foaf:Document | foaf:Project | 26 |
| foaf:Organization | foaf:Person | 7 |
| sioc:Community | sioc:Item | 3 |
| ecs:Fax | ecs:Telephone | 3 |

**Bonatti et al.** *"Robust and Scalable Linked Data Reasoning Incorporating Provenance and Trust Annotations".* JWS 2011 (in press).

# CONCLUSIONS

# Linked Data Reasoning Wrap-Up

**Heterogeneity poses a significant problem for consuming Linked Data**

1. ***Heterogenity in schema***
2. ***Heterogenity in naming***

**…but we can use the mappings provided by publishers to integrate heterogeneous Linked Data corpora (*with a little caution*)**

1. **Lightweight rule-based reasoning can go a long way**
2. **Deceit/Noise ≠ End Of World**
   - **Consider source of data!**
3. **Inconsistency ≠ End Of World**
   - **Useful for finding noise in fact!**
4. **Explicit `owl:sameAs` vs. extended consolidation:**
   - **Extended consolidation *mostly (but not entirely)* for consolidating blank-nodes from older FOAF exporters**