

Semantic Web and the Post Relational World

ITARC/ IASA
April 20, 2010

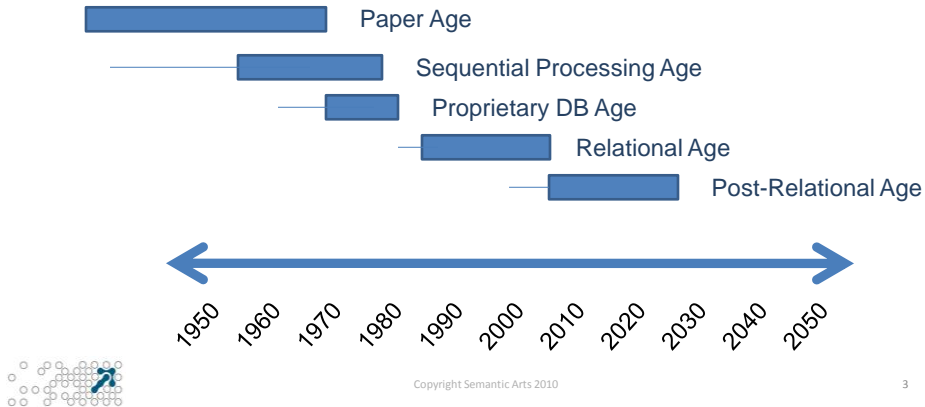
Dave McComb



The “Post-Relational” World

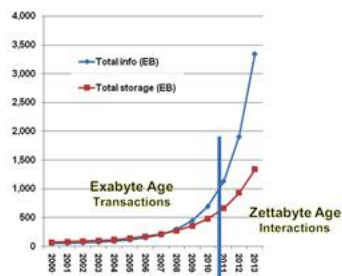


Time line



3

These days, we are drowning in data



www.c-data.nl

We have created more data since 2005 than we have in the previous 40,000 years.



Copyright Semantic Arts 2010

4

Limits of Relational

Storage

Google doesn't store its data in relational databases. There isn't a database that could handle it

Performance

There are many DBMS's now that outperform relational by 50x (Hadoop, Mapreduce, Columnar DBs etc)

Complexity

Relational databases rely on metadata to create and communicate distinctions.



Copyright Semantic Arts 2010

5

Sir Tim Berners-Lee



Copyright Semantic Arts 2010

6

Web 3.0/ Semantic Web

- What is it, and how does it turn data into information and knowledge?



Copyright Semantic Arts 2010

7

Three key distinctions in Web 3.0 (aka the Semantic Web)

- A uniform way to refer to specific instances.
- One way to declare all facts.
- A way to describe classes/categories/types/sets in a way that allows computers to categorize some data for us.



Copyright Semantic Arts 2010

8

Instances

- Anything individually identifiable
 - People
 - Organizations
 - Cars
 - Documents
 - Contracts
 - Transactions
 - Etc.



Copyright Semantic Arts 2010

9

Historic Treatment

- Create a table
- Give it an “id” or “key”

Securities	
ID	Company
02209S103	Phillip Morris
761713106	Reynolds

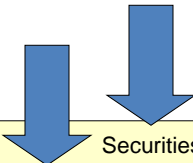


Copyright Semantic Arts 2010

10

Historic Treatment

- To get data out, you need to know the table and the column (the metadata).



Securities	
ID	Company
02209S103	Phillip Morris
761713106	Reynolds

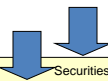


Copyright Semantic Arts 2010

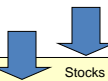
11

Historic Treatment

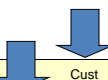
- To get data out of dozens of systems, you'd need to know dozens of bits of metadata.




Securities	
ID	Company
02209S103	Phillip Morris
761713106	Reynolds



Stocks	
Key	Company
02209S103	Phillip Morris
761713106	Reynolds



Cust	
CID	Company
02209S103	Phillip Morris
761713106	Reynolds



Companies	
Comp	Company
02209S103	Phillip Morris
761713106	Reynolds



Copyright Semantic Arts 2010

12

This doesn't scale



Copyright Semantic Arts 2010

13

Web 3.0 approach

- Skip the metadata (for identification).
- Everything is a resource.
- Everything gets a “URI” (think URL).

cusip:02209S103



<http://www.cusip.com>



Copyright Semantic Arts 2010

14

I know what you're thinking...

Does this mean everyone has to use the same id?

Won't this be like every "universal id" system we've ever seen?

Nope.

This doesn't require that everyone use the same id, only that if you use the id you refer to the same thing.



Copyright Semantic Arts 2010

15

What if two systems assign different identifiers to the same thing?



Copyright Semantic Arts 2010

16

<sameAs>

interlinking the Web of Data

The Web of Data has many equivalent URIs. This service helps you to find co-references between different data sets. Enter a known URI, or use Sindice to search first.

<sameAs>

Sindice

Search results from [Sindice](#), with co-references applied...

"City of Fort Collins, Colorado"

<sameAs> {

- 1 http://dbpedia.org/resource/Fort_Collins,_Colorado
- 2 <http://rdf.freebase.com/ns/guid.9202a8c04000641f800000000018714>
- 3 <http://sws.geonames.org/5577147/>
- 4 http://umbel.org/umbel/ne/wikipedia/Fort_Collins,_Colorado
- 5 http://www.rdfabout.com/rdf/usgov/geo/us/co/counties/larimer_county/fort_collins

rdf+xml · n3 · json · text · show fewer items

"Fort Collins Public Library"

<sameAs> {

- 1 http://dbpedia.org/resource/Fort_Collins_Public_Library
- 2 http://mpii.de/yago/resource/Fort_Collins_Public_Library
- 3 <http://rdf.freebase.com/ns/guid.9202a8c04000641f800000000065beec>
- 4 http://umbel.org/umbel/ne/wikipedia/Fort_Collins_Public_Library

rdf+xml · n3 · json · text

"Fort Collins-Loveland Municipal Airport"

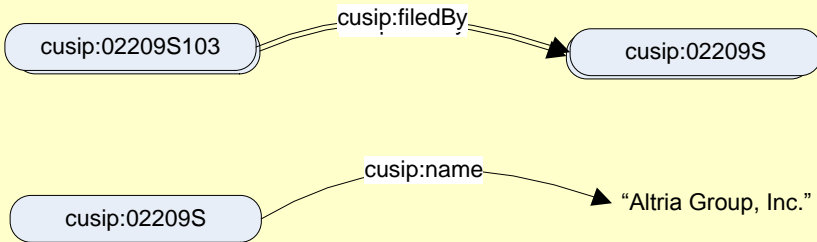
- 1 http://dbpedia.org/resource/Fort_Collins-Loveland_Municipal_Airport



Copyright Semantic Arts 2010

17

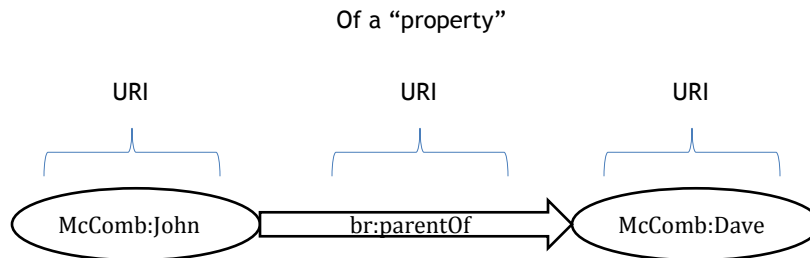
One (or two) ways to assert facts



Copyright Semantic Arts 2010

18

Anatomy of a triple



Copyright Semantic Arts 2010

19

The Logical Model of the Semantic Web

Subject	Predicate	Object
URI	URI	URI
URI	URI	Literal

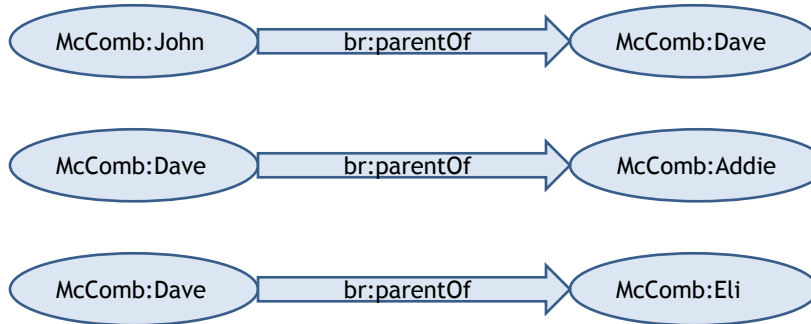


Copyright Semantic Arts 2010

20

Triples

Subject	Predicate	Object
McComb:John	br:parentOf	McComb:Dave
McComb:Dave	br:parentOf	McComb:Addie
McComb:Dave	br:parentOf	McComb:Eli



Copyright Semantic Arts 2010

21

Here's where the magic is....

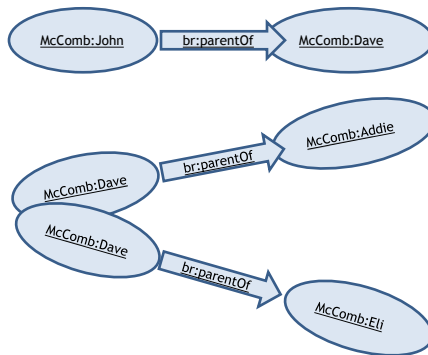


Copyright Semantic Arts 2010

22

Triples to Graphs

Linked Data



Copyright Semantic Arts 2010

23

The “join” is done at the instance level

- No metadata was harmed in the making of this join.
- The metadata wasn’t even interrogated.
- In fact, it just isn’t necessary.



Copyright Semantic Arts 2010

24

Schema

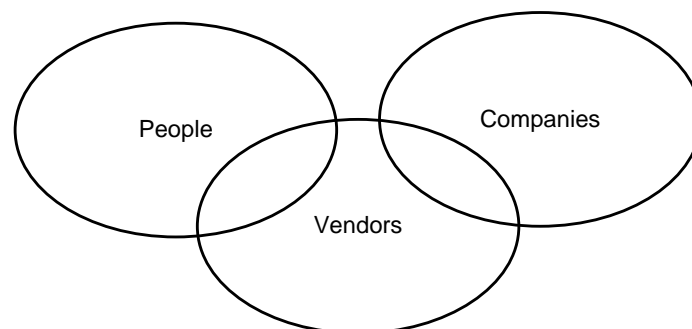
- In traditional systems, “schema” defines physical structure as well as hinting at meaning, and must be defined before data can be stored.
- In “Semantic-land” the schema is “logical” (not physical) and “late” (can be bound after the instances have been created).



Copyright Semantic Arts 2010

25

Classes, Categories, Sets, Types



Copyright Semantic Arts 2010

26

Current Approaches

- Humans assign things to categories.
- Things get one primary category and that category's parents.
- Once assigned, items stay in their categories.



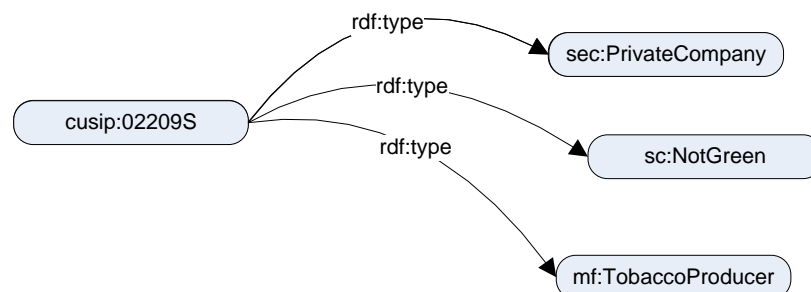
Copyright Semantic Arts 2010

27

Web 3.0 approach

“Type” or “Class” is not structural. It's just another assertion.

Any instance may be many types simultaneously.

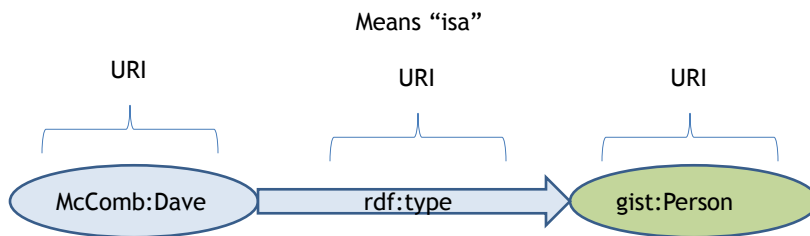


Copyright Semantic Arts 2010

28

Using Existing Classes

Subject	Predicate	Object
McComb:Dave	rdf:type	gist:Person

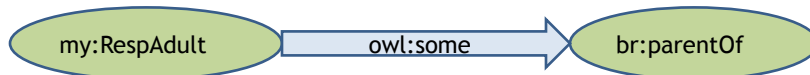


Copyright Semantic Arts 2010

29

Making Up Classes (schema later and local)

Subject	Predicate	Object
my:RespAdult	owl:some	br:parentOf



Means: "you are a member of the class of Responsible Adults, if you are the parent of at least one [Person]"

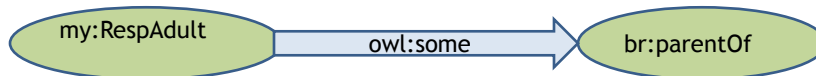


Copyright Semantic Arts 2010

30

Inference

Subject	Predicate	Object
McComb:John	br:parentOf	McComb:Dave
McComb:Dave	br:parentOf	McComb:Addie
McComb:Dave	br:parentOf	McComb:Eli
McComb:John	rdf:type	my:RespAdult
McComb:Dave	rdf:type	my:RespAdult

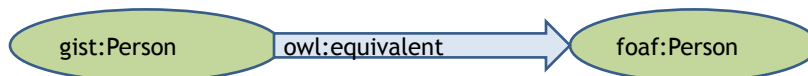


Copyright Semantic Arts 2010

31

Integration

Subject	Predicate	Object
gist:Person	owl:equivalent	foaf:Person



Means: "any one who is a 'gist:Person' is also a 'foaf:Person' and vice versa."



Copyright Semantic Arts 2010

32

IRS definition of a passenger automobile

- A passenger automobile is a 4-wheeled vehicle manufactured primarily for use on public roads that is rated at 6,000 pounds unloaded gross vehicle weight or less. Certain vehicles, such as ambulances, hearses, and taxicabs, are not considered passenger automobiles and are not subject to the line 36 limits...

Form 4797



Copyright Semantic Arts 2010

33

The screenshot shows the TopBraid Eclipse SDK interface for editing the `PassengerVehicle` class in an OWL ontology. The `Classes` tree on the left shows the hierarchy: `owl:Thing` (parent), `Vehicle` (child), `NotPassengerVehicle` (child), `PassengerVehicle` (child), `Truck` (child), `Van` (child), `Weight` (child), and `Wheel` (child). The `Class Form` editor for `PassengerVehicle` shows the following configuration:

- Name:** PassengerVehicle
- Annotations:** (None listed)
- Class Axioms:**
 - `rdfs:subClassOf`: Vehicle
 - `owl:disjointWith`: NotPassengerVehicle, Ambulances, Taxicab, Hearse
 - `owl:equivalentClass`:
 - `hasWheel` **max** 4
 - `ratedWeight` **has** lessThan6000
 - `ratedWeight` **some** Weight

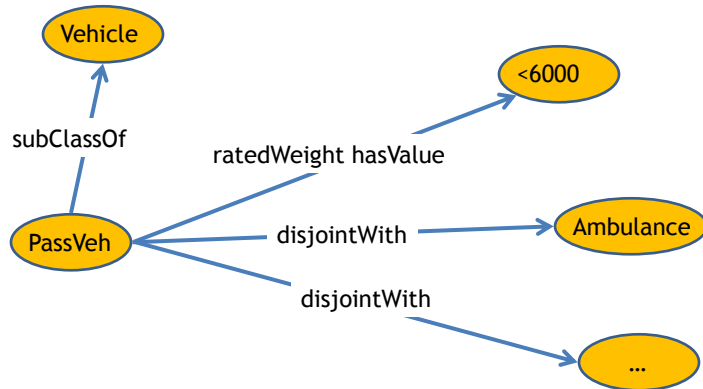
The **Properties** panel on the right lists the following properties: `hasWheel`, `ratedWeight`, `unloadedGrossVehicleWeight`, `owl:versionInfo`, `rdfs:comment`, `rdfs:label`, and `rdfs:seeAlso`.



Copyright Semantic Arts 2010

34

Even the definition of classes is done in triples



Copyright Semantic Arts 2010

35

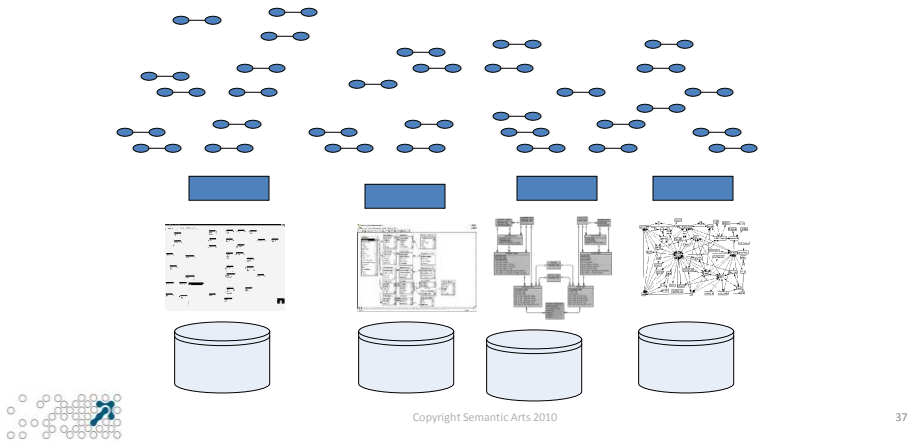
So where are these “triples” coming from ?



Copyright Semantic Arts 2010

36

Triples from Databases



c:123 $\xrightarrow{m:P_ID}$ p:P27

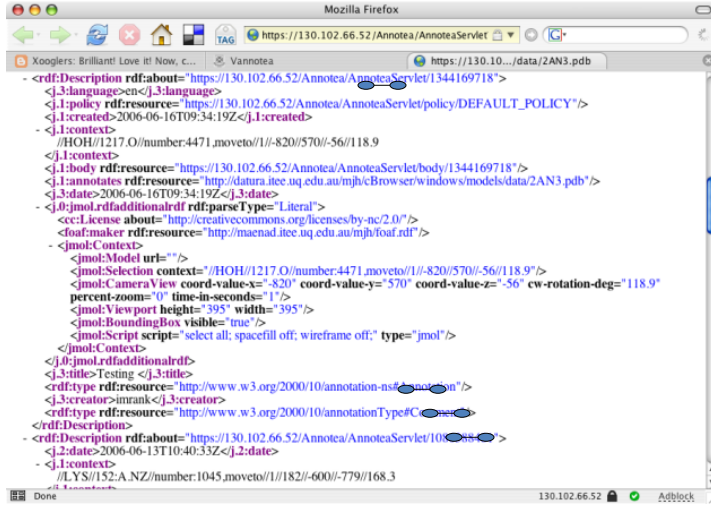
p:P27 $\xrightarrow{m:OnHand}$ "55"

C ID	O_ID	P_ID	Qty	Price
123	AA001	P27	4	1.00
456	AA002	P27	6	1.00

Prod	Desc	OnHand
P27	Bolt	55
P28	Nut	66



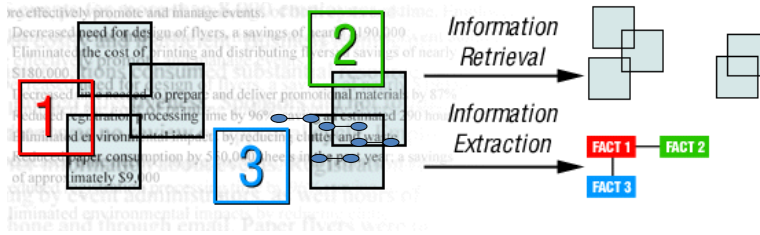
Triples from XML/HTML



Copyright Semantic Arts 2010

39

Triples from Text



Aerotext from Lockheed Martin



Copyright Semantic Arts 2010

40

Triples and Triple Stores

- Once we've harvested a large number of "triples" we need a place to store and efficiently access them.
- This is the role of a "triple store," essentially a database for these assertions.



Copyright Semantic Arts 2010

41

Querying

- The equivalent of SQL is SPARQL

```

PREFIX foaf:
<http://xmlns.com/foaf/0.1/>
SELECT ?name ?mbox
WHERE
  { ?x foaf:name ?name .
    ?x foaf:mbox ?mbox }

```

- A service that accepts SPARQL queries is called a "SPARQL End Point"



Copyright Semantic Arts 2010

42

ROA

- Semantic Web Technology is a natural fit with Resource Oriented Architecture
 - Based on RESTful design principles
 - Few verbs, many nouns
 - Nouns are the Resources, expressed as URIs



Copyright Semantic Arts 2010

45

“Web of Things”

- Semantic Technology is also a good fit for the newly emerging “web of things”
 - Sensors
 - Motes
 - RFID



Copyright Semantic Arts 2010

46

Transparency

- What governments, and companies, are discovering.
- Our chance of becoming wise (acquiring wisdom) will be increased if we can give more people access to more information and knowledge.



Copyright Semantic Arts 2010

47

WEDNESDAY, OCTOBER 14, 2009 Text AT A

DATA.GOV Share | Facebook | Twitter | RSS

HOME | CATALOGS | STATE/LOCAL | ABOUT | FAQ | CONTACT US | SUGGEST OTHER DATASETS

**DISCOVER.
PARTICIPATE.
ENGAGE.**

Search the following Data.gov catalogs:

"RAW" DATA CATALOG

TOOL CATALOG

GEODATA CATALOG

**FEATURED TOOL:
NATIONAL AERONAUTICS AND SPACE
ADMINISTRATION (NASA)
MY NASA DATA**

Mentoring and inquiry using NASA data on Atmospheric and Earth Science for teachers and amateurs. The MY NASA DATA Live Access Server (LAS) is now available to create your own microsets for your class or your interests. The LAS contains over 148 parameters in atmospheric and earth science from five NASA scientific projects.

[VIEW THIS TOOL](#) ▶

Welcome to Data.gov

The purpose of Data.gov is to increase public access to high value,

How to use Data.gov

Data.gov includes searchable [data catalogs](#) providing access to

1 2 3 4 5

<http://data.gov>



Copyright Semantic Arts 2010

48

This We Know: What's This We Know?

Fort Collins, CO

- There are **8 Factories** (within 11 mi.) [tweet this](#)
- 21,294 pounds of 6 Pollutants** were released (within 11 mi.) [tweet this](#)
- 8 Officials** reported on **8 Factories** (within 11 mi.) [tweet this](#)
- 418 Violent Crimes** occurred or **3.2 per 1000 people** (in this town) [tweet this](#)
- Demographics:** **12,993** people were **Hispanic**, **1,441** were **African American**, **2,741** were **Asian**, **126,146** were **White**, **70,728** were **Male**, and **70,004** were **Female** (in this town)
- There are **32,970 Home Owners** and **21,473 Renters** (in this town) [tweet this](#)
- 62%** of people **Relocated** in the past 15 years (in this town) [tweet this](#)
- 5 Bills** have been introduced about this location by **3 Members of Congress** since 1993 [tweet this](#)

Tip: Click on any of the highlighted items above to explore the underlying data.
Sources: 2005 Toxics Release Inventory • 2007 Crime in the United States • US Census 2000 • GovTrack



Population: 140,712
 Households: 56,489
 Land Area: 110.4
 Water Area: 8.0

NEARBY...

- Boulder, CO 40 mi
- Denver, CO 58 mi
- Lakewood, CO 60 mi
- Colorado Springs, CO 121 mi
- Pueblo, CO 162 mi



This We Know:

Fort Collins, CO

418 Violent Crimes occurred or **3.2 per 1000 people** (in this town)

VIOLENT CRIME	BURGLARY	ROBBERY	AGGRAVATED ASSAULT	VEHICLE THEFT	LARCENY THEFT	FORCIBLE RAPE	PROPERTY CRIME	MURDER
418	695	33	314	294	3487	70	4476	1

Source: 2007 Crime in the United States

[Have some Feedback? Help us improve this site.](#) [Speak SPARQL? Show the query that generated this page >](#)



SPARQL ▶ **Speak SPARQL?**
Show the query that generated this page >

```

PREFIX o: <http://www.data.gov/ontology#>
PREFIX ui: <http://www.thisweknow.org/ui#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX geo: <http://www.w3.org/2003/01/geo/wgs84_pos#>
PREFIX samp: <http://www.rdfabout.com/rdf/schema/usensus/details/samp/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>

SELECT ?violent_crime ?town
       ?burglary ?town
       ?robbery ?town
       ?aggravated_assault ?town
       ?vehicle_theft ?town
       ?larceny_theft ?town
       ?forcible_rape ?town
       ?property_crime ?town
       ?murder ?town
       ?population ?town
FROM <data> WHERE {
  ?crime_stat o:town ?town;
              o:violentCrime ?violent_crime;
              o:burglary ?burglary;
              o:robbery ?robbery;
              o:aggravatedAssault ?aggravated_assault;
              o:motorVehicleTheft ?vehicle_theft;
              o:larcenyTheft ?larceny_theft;
              o:forcibleRape ?forcible_rape;
              o:propertyCrime ?property_crime;
              o:murder ?murder;
              o:population ?population .
FILTER(?town=<http://www.rdfabout.com/rdf/usgov/geo/us/co/counties/larimer_county/fort_collins>)
}

```



Copyright Semantic Arts 2010

51

Summary

- We are transitioning into the Post-Relational Age
- Our need to transition is that we can't deal with the complexity or rigidity of schema-first design
- Semantic Technology offers a powerful new approach: modular, roll-your-own schema accessing far more data than you will ever have in your corporation
- Compatible with ROA & "Web of Things"



Copyright Semantic Arts 2010

52

Questions?



Copyright Semantic Arts 2010

53