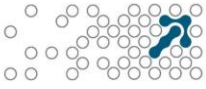
Semantic Web and the Post Relational World

ITARC/ IASA

April 20, 2010

Dave McComb

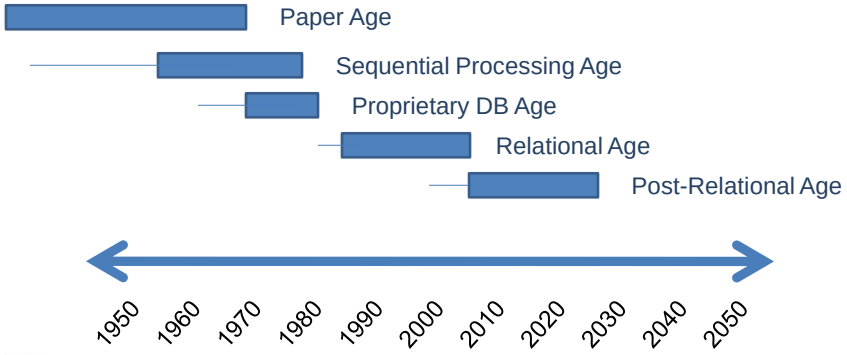


Semantic Arts

The "Post-Relational" World

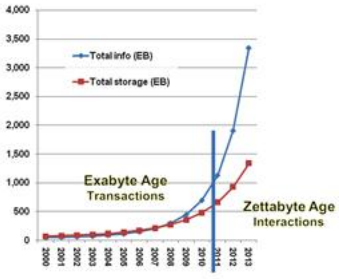
Copyright Semantic Arts 2010

Time line



Copyright Semantic Arts 2010

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

Limits of Relational

Storage Google doesn't store it's 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

Sir Tim Berners-Lee

Copyright Semantic Arts 2010

Web 3.0/ Semantic Web

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

Copyright Semantic Arts 2010

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

Instances

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

Copyright Semantic Arts 2010

Historic Treatment

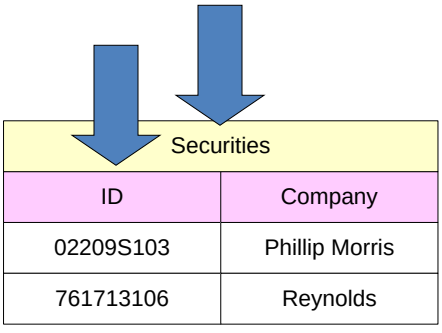
- Create a table
- Give it an “id” or “key”

Securities	
ID	Company
02209S103	Phillip Morris
761713106	Reynolds

Copyright Semantic Arts 2010

Historic Treatment

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

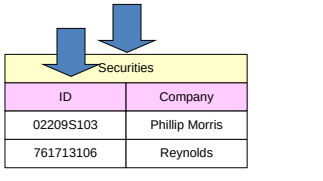


Securities	
ID	Company
02209S103	Phillip Morris
761713106	Reynolds

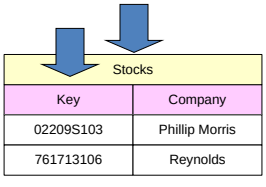
Copyright Semantic Arts 2010

Historic Treatment

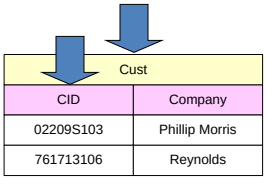
- To get data out of dozens of systems, you'd need to know dozens of bits of metadata.



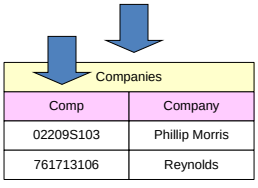
Securities	
ID	Company
02209S103	Phillip Morris
761713106	Reynolds



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

This doesn't scale

Copyright Semantic Arts 2010

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

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?

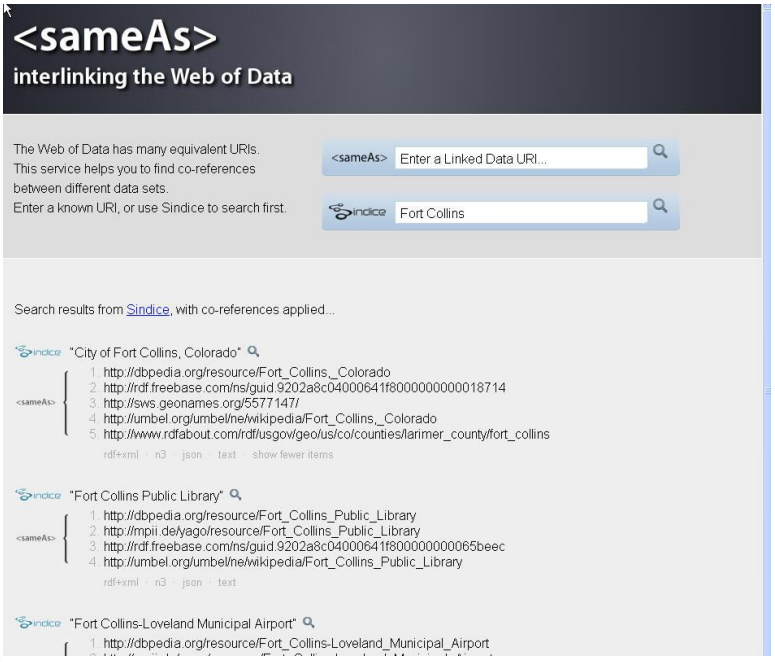
Nope.

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

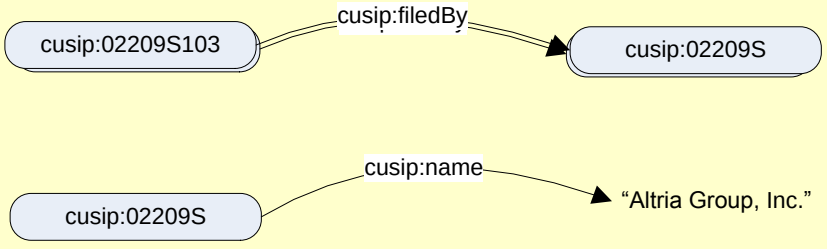
What if two systems assign different identifiers to the same thing?

Copyright Semantic Arts 2010



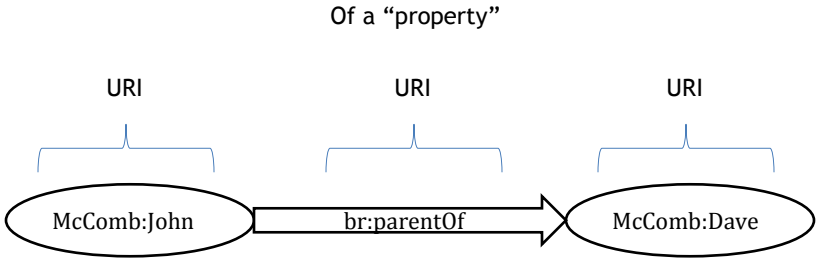
Copyright Semantic Arts 2010

One (two) ways to assert facts



Copyright Semantic Arts 2010

Anatomy of a triple



Copyright Semantic Arts 2010

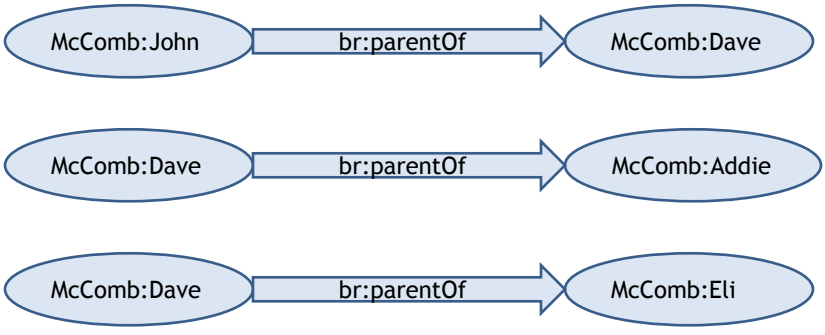
The Logical Model of the Semantic Web

Subject	Predicate	Object
URI	URI	URI
URI	URI	Literal

Copyright Semantic Arts 2010

Triples

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



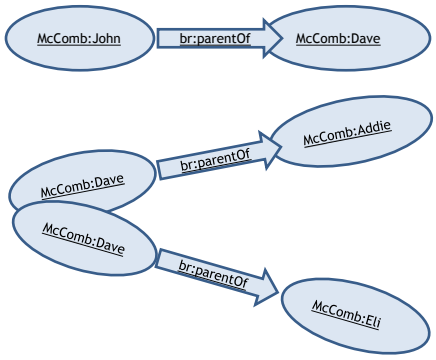
Copyright Semantic Arts 2010

Here's where the magic is....

Copyright Semantic Arts 2010

Triples to Graphs

Linked Data



Copyright Semantic Arts 2010

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.

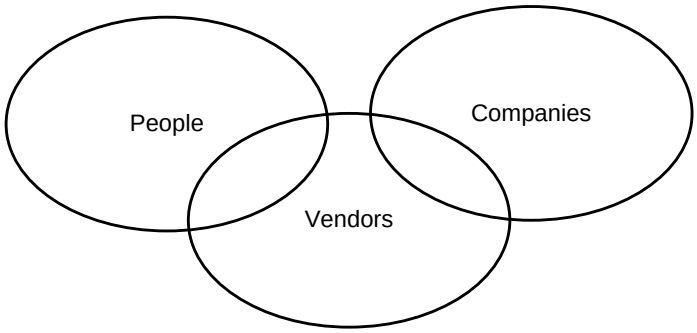
Copyright Semantic Arts 2010

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

Classes, Categories, Sets, Types



Copyright Semantic Arts 2010

Current Approaches

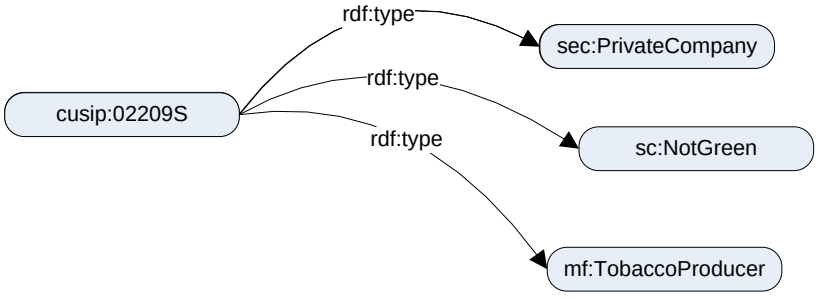
- Humans assign things to categories.
- Things get one primary category and that category's parents.
- Once assigned, items stay in their categories.

Copyright Semantic Arts 2010

Web 3.0 approach

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

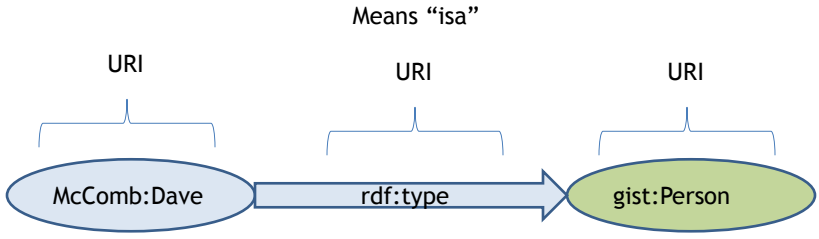
Any instance may be many types simultaneously.



Copyright Semantic Arts 2010

Using Existing Classes

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



Copyright Semantic Arts 2010

Making Up Classes (schema later and local)

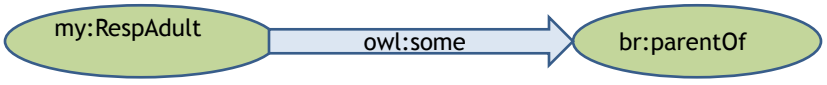
Subject	Predicate	Object
my:RespAdult	owl:some	br:parentOf

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

Inference

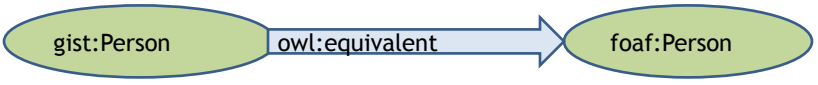
Subject	Predicate	Object
McComb:John	br:parentOf	McComb:Dave
McComb:Dave	br:parentOf	McComb:Addie
McComb:Dave	br:parentOf	McComb:Eli
McComb:John	rdf:type	my:RespAdult
McComb:Dave	rdf:type	my:RespAdult



Copyright Semantic Arts 2010

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.

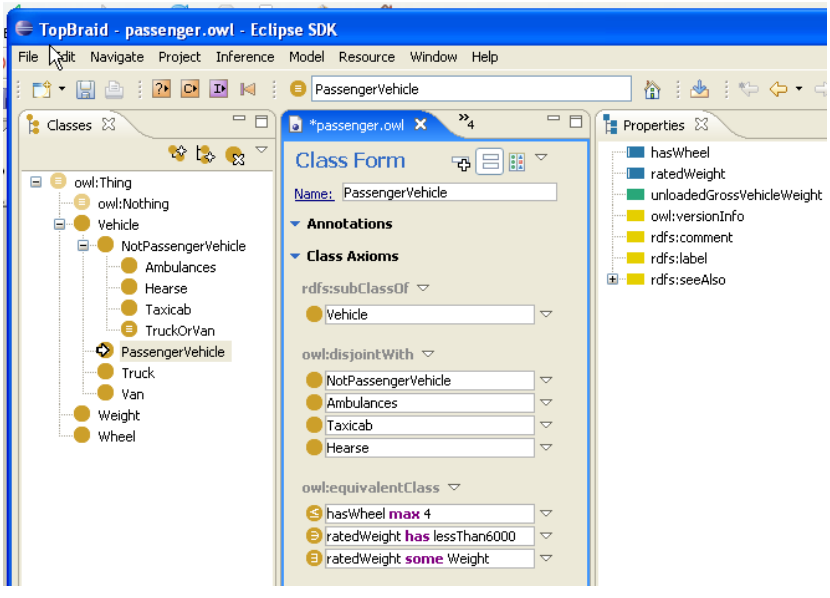
Copyright Semantic Arts 2010

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…

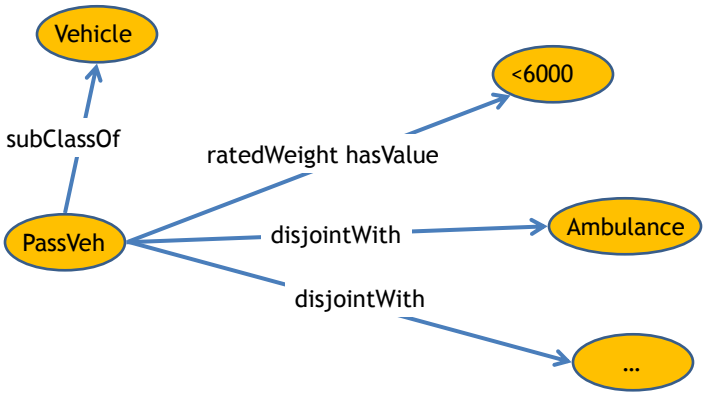
Form 4797

Copyright Semantic Arts 2010



Copyright Semantic Arts 2010

Even the definition of classes is done in triples

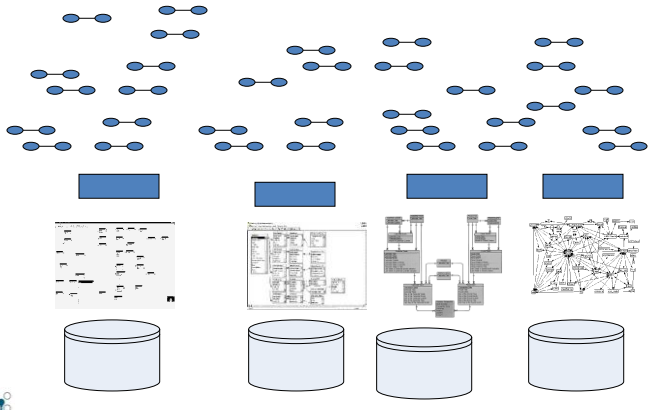


Copyright Semantic Arts 2010

So where are these "triples" coming from ?

Copyright Semantic Arts 2010

Triples from Databases

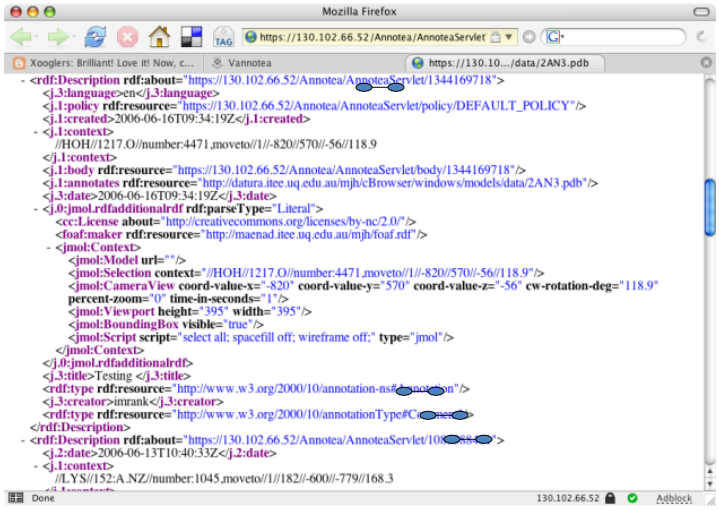


Copyright Semantic Arts 2010

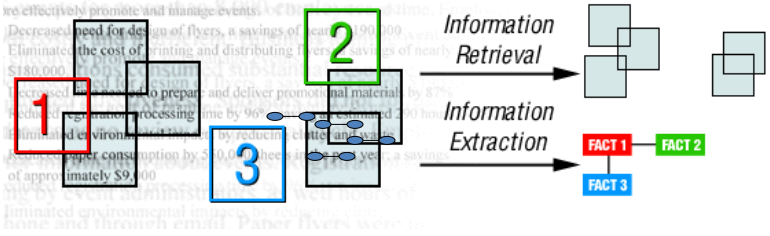
Prod	Desc	OnHand
P27	Bolt	55
P28	Nut	66

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

Triples from XML/HTML



Triples from Text



Aerotext from Lockheed Martin

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

Querying

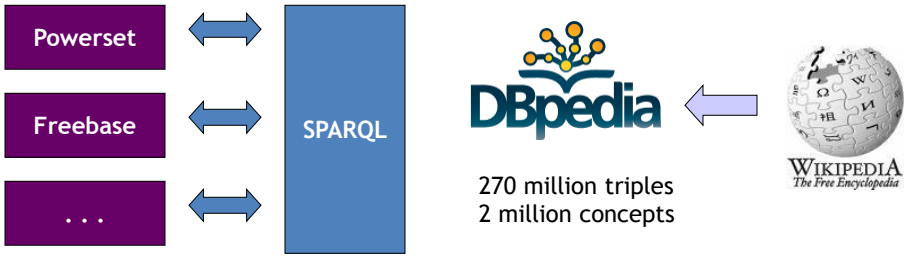
- The equivalent of SQL is SPARQL

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

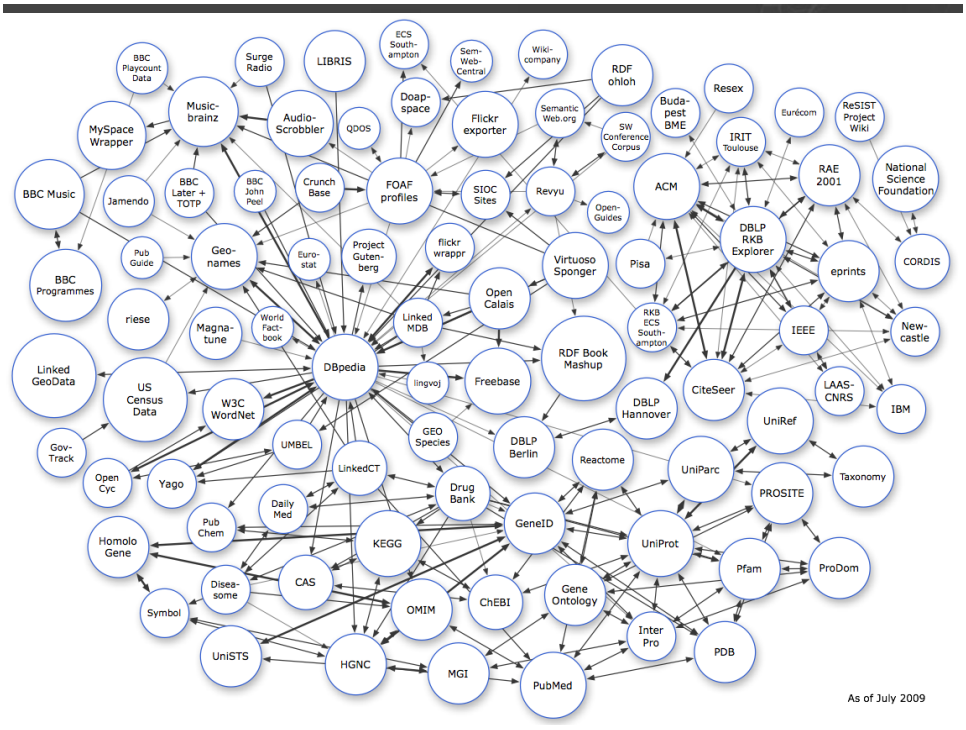
- A service that accepts SPARQL queries is called a "SPARQL End Point"

Copyright Semantic Arts 2010

SPARQL End Points



Copyright Semantic Arts 2010



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

"Web of Things"

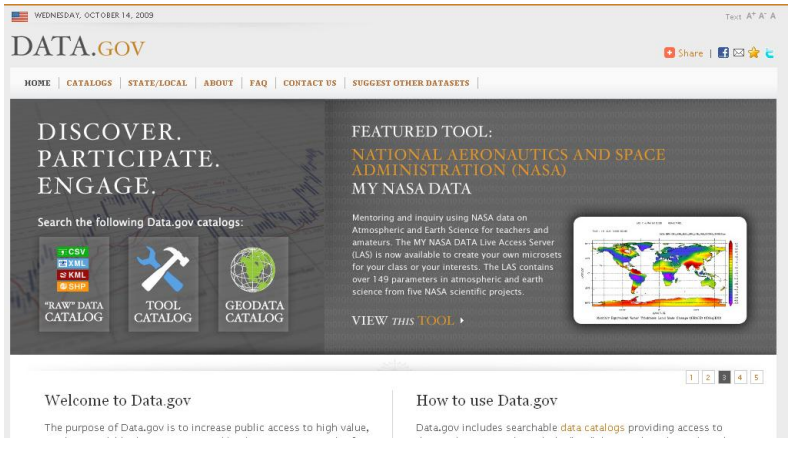
- Semantic Technology is also a good fit for the newly emerging "web of things"
  - Sensors
  - Motes
  - RFID

Copyright Semantic Arts 2010

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.

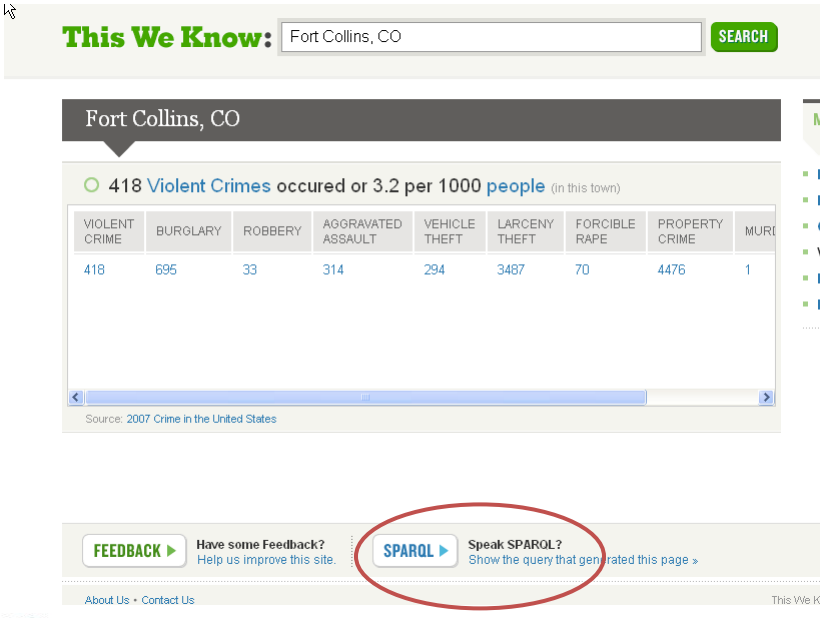
Copyright Semantic Arts 2010



http://data.gov

Copyright Semantic Arts 2010

Copyright Semantic Arts 2010



Copyright Semantic Arts 2010

Copyright Semantic Arts 2010

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

Questions?

Copyright Semantic Arts 2010