An Argument For Semantics

Why developers should give a hoot about OWL

Brian Panulla

The quest for a smarter Web

- What is a Semantic Web, and why would I want one? Isn't Web 2.0 good enough?
- The "O" Word
- Using SW technologies today

There's too much confusion…

The Semantic Web

Web 3.0

Linked Data

New W3C Languages

O RLY?

- RDF
- RDF Schema
- OWL

- Each builds on one another, but all are fundamentally RDF

- (Just as XHTML, MXML, or SVG are all XML)

PWC Spring 2009 Technology Forecast

Implied Meaning

- "Jane Smith is a *designer* in New York"
- "Bob Walker is a *designer* in San Francisco"
- "Alice Reed is a *designer* for online learning."

Implied Meaning

- Meaning inferred from context

    – "Jane Smith is a *designer* in New York"
    – "Bob Walker is a *designer* in San Francisco"
    – "Alice Reed is a *designer* for online learning."

- What kind of designer? Are each the same?

Semantics

- *"meaning"* of symbols
  - Words – usage, connotation
  - Images - symbolism

- Become really useful when shared
  - Between individuals
  - Within a community or culture

Making the best of dumb data

Moving to Smart Data

- We can add semantics (meaning) to our data to relieve the burden on our applications

- Smarter data provides a better foundation for smarter software and systems.

Is this more "Cathedral" thinking?

- Top-down Ivory tower approach has led to our current network of *walled garden*s of data:
  - Data Warehouses
  - Data Mining
  - Data Stewards

- Could some of our data be more *open?*

http://en.wikipedia.org/wiki/File:Stained_glass_Tower_of_Ivory.jpg

Sharing Enterprise data

E-mail link to us instead of what's below

Regular Method	
Select Course Abbreviation	Abbrev ▾
Enter Course Number	Text
Semester	▾
○ Resident ○ World Campus ○ Contin Ed	
Location	▾
Section	Text
Advanced Method	

- Why can't we pull non-sensitive enterprise data from an open, central source?
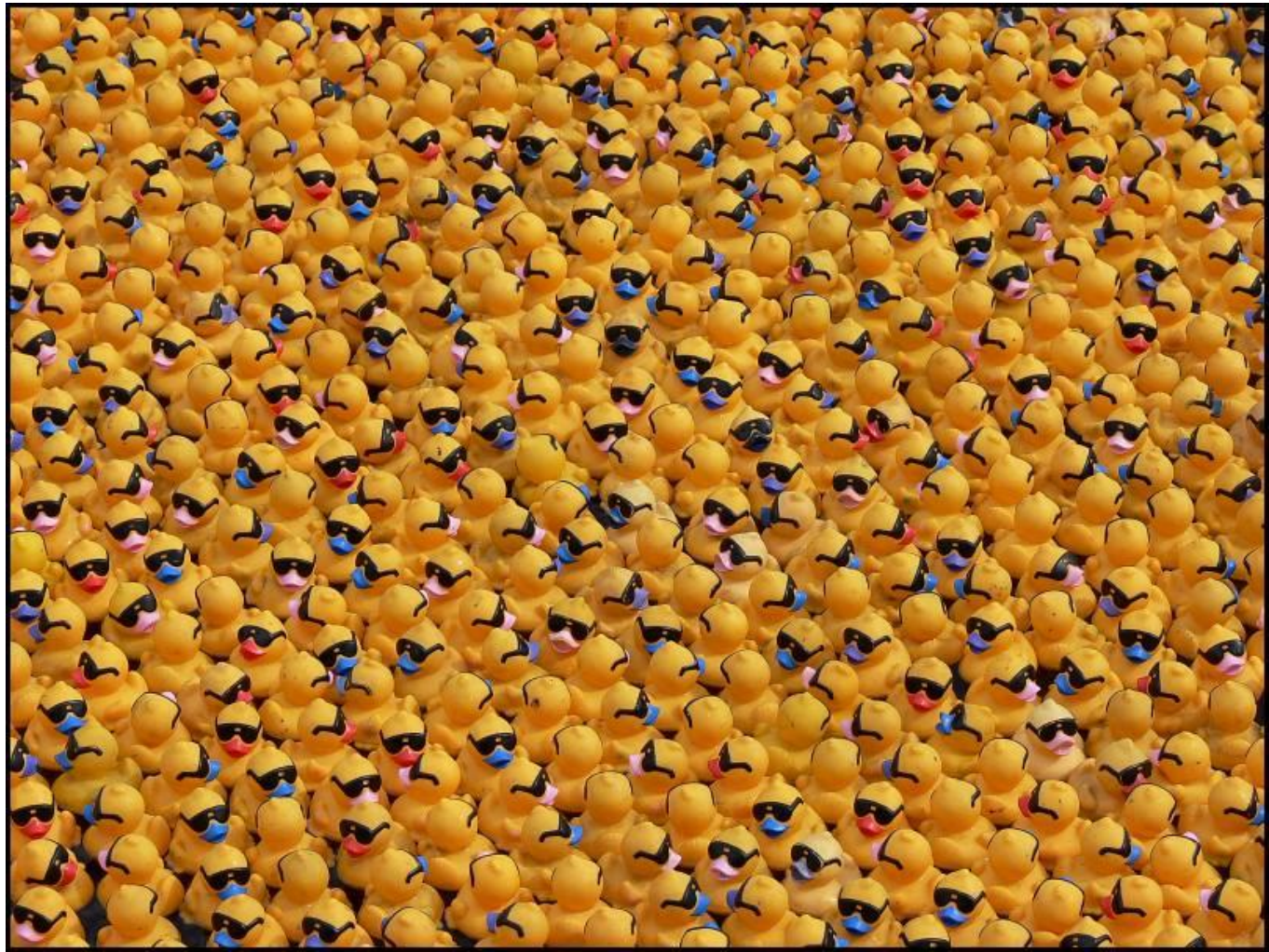
Redundant Data

How many Web applications have local copies of:

- States
- Countries
- Campuses
- Majors
- Courses

Why are we maintaining them?

Separation of Concerns

- Smarter data is driving new levels of *separation of concerns*

  - Content
  - Presentation
  - Behavior
  - *Rules*

Whither HTML and XML: Does the old Web die?

The *SW is infrastructure*

- A parallel information architecture *design pattern* for *smarter applications*

- Web content, pages and sites **do not need to change** to be made ready for *"Teh Webs 3.0" ...*
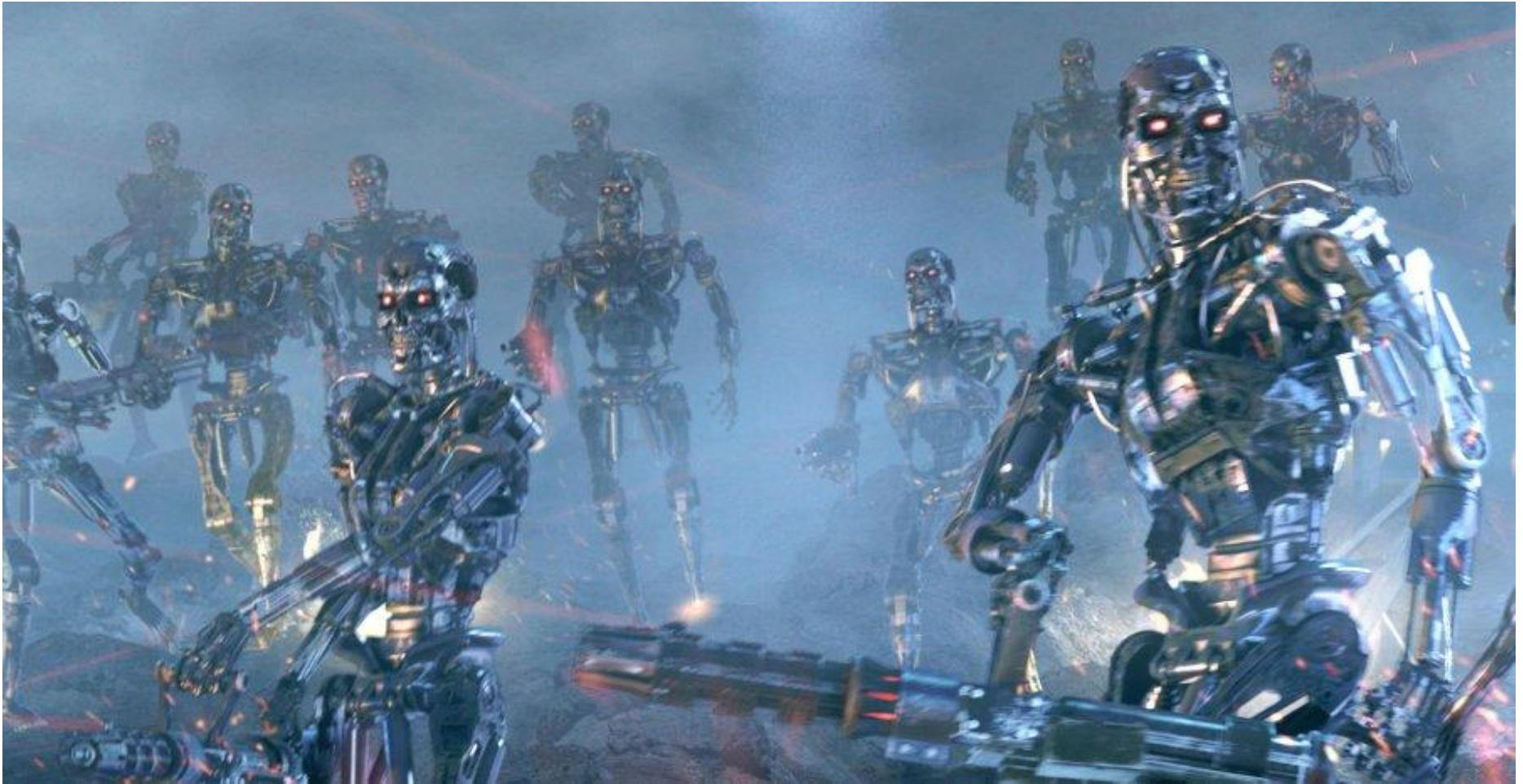
Software Inferencing

- More meaningful markup allows software to make more robust decisions

- Sounds good, but:
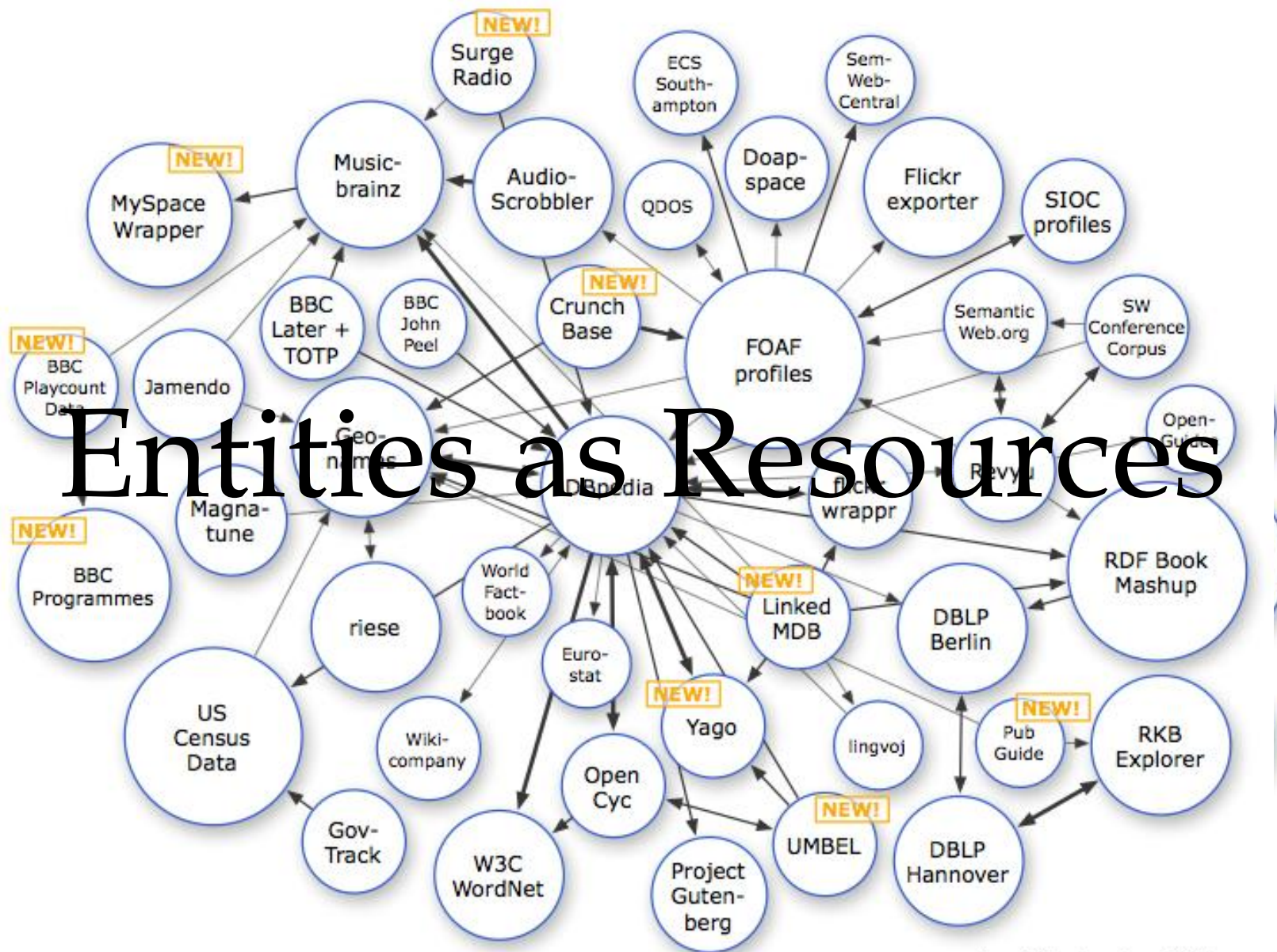   - How do we get there?
    - And is it safe?

Rise of the Machines?

http://www.movieprop.com/tvandmovie/terminator/t3endoskeletons1.jpg

Roadmap to Smart Data

- Entities as Resources
- Specifying Relationships
- Drawing Inferences

Entities as Resources

As of September 2008

The Non-Semantic Web

The Non-Semantic Web

Google™   [IST   ] (Search)   Advanced Search
                                                                    Preferences

Web   Show options...

Indian Standard Time - Wikipedia, the free encyclopedia 🔼❌
Indian Standard Time (IST) is the time observed throughout India and Sri Lanka, with a time
offset of UTC+5:30. India does not observe daylight saving time, ...
en.wikipedia.org/wiki/Indian_Standard_Time - 85k - Cached - Similar pages - 💬

Penn State College of **Information Sciences and Technology** 🔼❌
Penn State's visionary response to the rapidly growing need in almost every field for
leadership in information sciences and related technologies.
ist.psu.edu/ - 12k - Cached - Similar pages - 💬

IST -- The Company with Passion 🔼❌
Mail, courier, shipping, receiving, and other services, for companies nationwide. Based in
Atlanta, Georgia.
www.**ist**management.com/ - 9k - Cached - Similar pages - 💬

See results for: *indian standard time*

It gets worse…

- Locally "IST" can refer to:
  - The administrative unit (where I work)
  - The building (where I work, in a different way)
  - The college (degree-granting organization)
  - 5 degree programs (A.S., B.S., B.A., M.S., Ph.D.)
  - The family of courses (independent of major)
  - The field of study/research

How do we identify entities?

- Differentiating between conceptual entities creates the need for an identifier

  – Indefinite article: *A* college of IST

  – Definite article: *The* College of IST **at Penn State**

How do we identify entities?

- Convention allows us to simplify integration of data across systems

  – USPS State Code for states
  – ISO Country Code for countries
  – 2-letter symbol for elements

- Convention is *implicit semantics*

How do we identify entities?

- In the absence of a good candidate key, each organization usually make an *ad hoc* artificial one.

  - Conferences - **#heweb09**
  - Speakers - **@bpanulla**
  - Tracks - **TPR**
  - Sessions – **TPR9**

HELLO
my name is

24601

Identification

- A global Web needs a global ID scheme

- How do we differentiate between entities, even if they have the same name?

- Turns out we already have a handy tool…

URIs

- Entities are identified by a Uniform Resource Identifier (URI):

  http://www.psu.edu/owl/ist.owl#CollegeOfIST

  http://highedweb.org /owl/hew2009.owl#TPR9

URIs

- URIs can also be found in your XHTML DTD and html tag's *xmlns* attribute:

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitic
<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="e
<head>
```

URIs

- A *URI* looks like a Uniform Resource Locator (URL)

- There is no guarantee that a URI may be accessible as with a browser.

  … but it usually is. Hence….

URIs

I *link* therefore I *am.*

http://en.wikipedia.org/wiki/File:Frans_Hals_-
_Portret_van_Ren%C3%A9_Descartes.jpg

URIs and identity

- Normally, entities identified by two different URIs are *distinct*

- We can override this and have different URIs refer to the same entity

Describing Relationships

As of September 2008

RDF: Resource Description Framework

- RDF is a fundamental knowledge representation
  - Declares *resources*
  - Specifies *properties*

- RDF can be used to specify *is-a, is-a-member-of,* and *has-a* relationships

New W3C Languages

- RDF defines basic *type ("class") framework, further extended by OWL*

- RDF (and OWL) are stored natively *as triples:*
    - *subject ("Penn State")*
    - *predicate ("is a")*
    - *object ("University")*
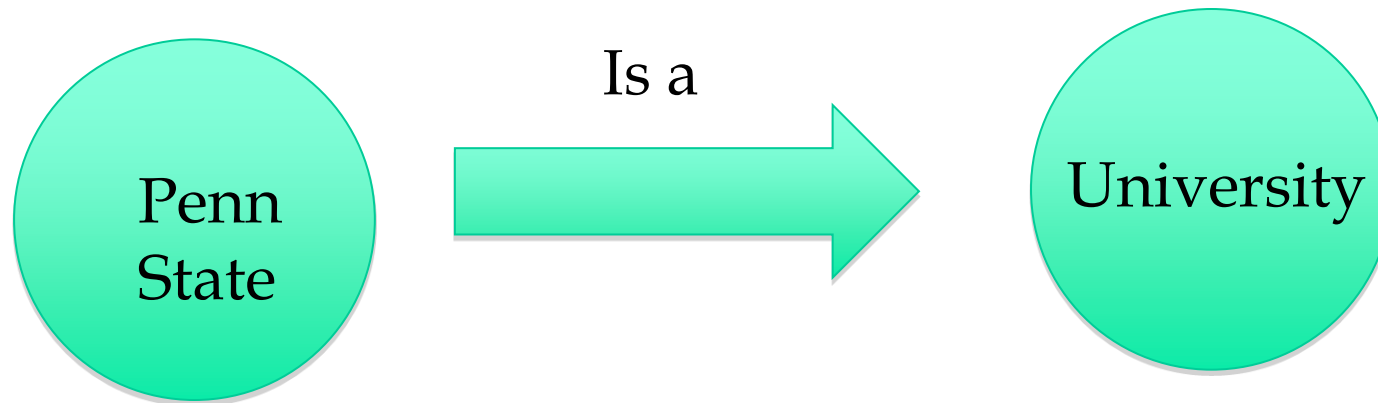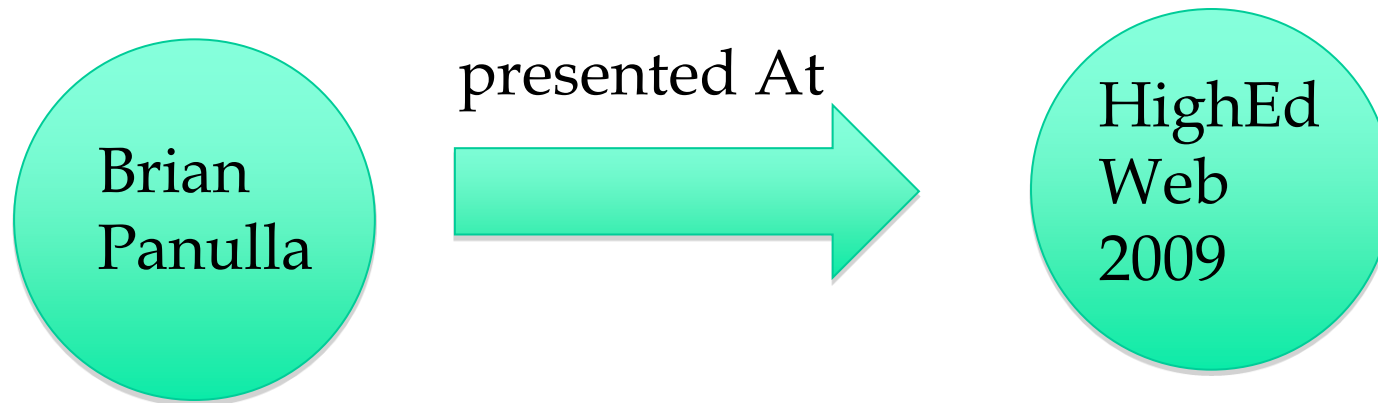
A HighEdWeb Model

Types:

- Conference
- Presentation
  - Poster
  - Session
  - Keynote
- Presenter
- Track
- Room

Properties:

- presentedAt
- presentedBy
- presentedIn
- partOfTrack

New W3C Languages

- Another example:
  - *subject ("Brian Panulla")*
  - *predicate ("presentedAt")*
  - *object ("HighEdWeb 2009")*

presented At

Brian Panulla

HighEd Web 2009

RDF/XML

```
<rdf:RDF
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:rdfs="http://www.w3.org/2002/07/owl#"
    xmlns:owl="http://www.w3.org/2002/07/owl#"
    xmlns:hew="http://highedweb.org/owl/hew.owl#"
    xml:base="http://highedweb.org/owl/hew2009.owl">

<hew:Session rdf:ID="TPR9" rdfs:label="An Argument For Semantics">
    <hew:heldIn rdf:resource="#Mitchell" />
    <hew:givenBy rdf:resource="#Brian Panulla" />
    <hew:partOfTrack rdf:resource="#TPR" />
</hew:Session>

</rdf:RDF>
```

RDF/N3

:BrianPanulla a hew:Speaker;
    rdfs:label "Brian Panulla".

:Mitchell a hew:Room.

:TPR a hew:Track;
    rdfs:label "Technical: Propeller Hats Required".

:TPR9
    rdfs:label="An Argument For Semantics";
    hew:heldIn :Mitchell;
    hew:givenBy :BrianPanulla;
    hew:partOfTrack :TPR.

Structure

- Types and properties used in RDF
  documents can be defined in
  - RDF Schemas
  - Web *ontologies*

RDF Schemas

- *Think of classes* in RDFS as sets rather than OOP classes

- RDFS adds limited Set Theory properties
  - subClassOf
  - subPropertyOf
  - Domain
  - Range

DOGS — have fur all over

HUMANS — can run for President / have 2 legs

mammals

animals may eat bugs

have 4 legs

found in LEARN NC offices

reptiles have scales

LIZARDS

Ontologies

- *Ontology* is the study of *being or reality*.

- A Formal Ontology is a representation of a true ontology in some sort of communicable format

Aristotle ->
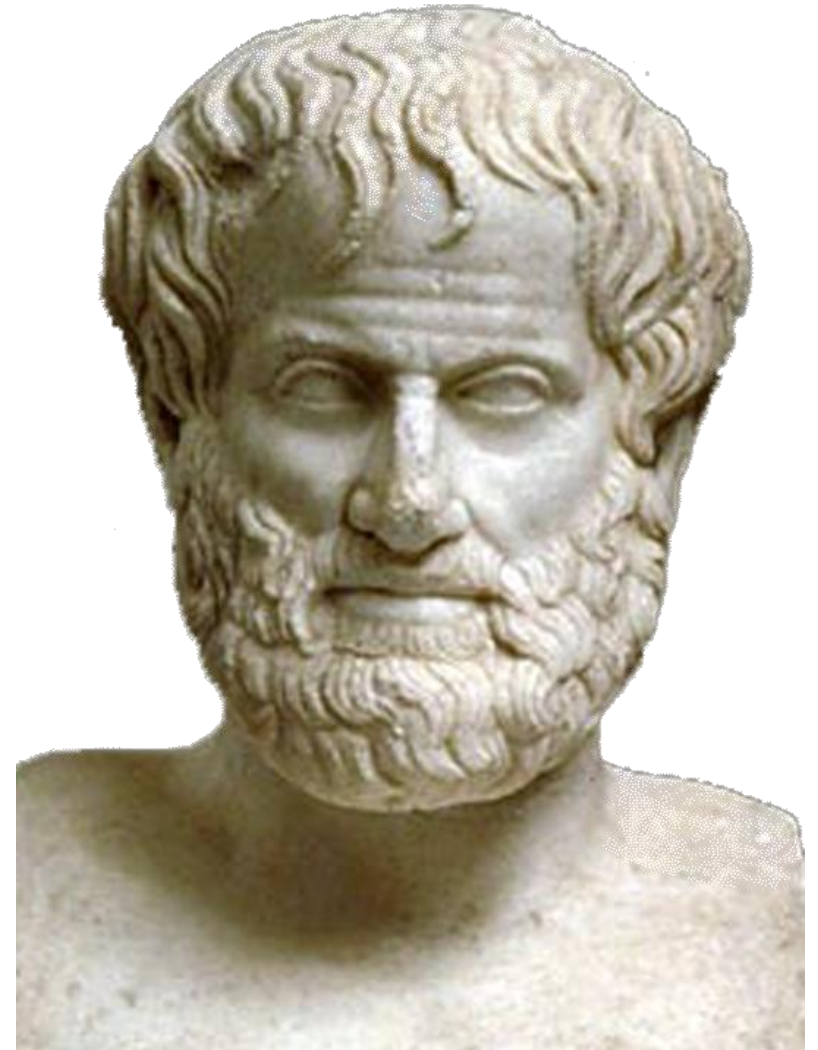
Web Ontologies

- Where *schemas* describe structure, *ontologies* describe meaning or intended use.

- OWL adds more expressiveness and many aspects of formal logic, superseding RDFS.

OWL Features

- Classes
  - SubClassOf
  - EquivalentClasses
  - DisjointClasses
  - Cardinality constraints (max/min)

- Properties
  - SubPropertyOf
  - EquivalentProperties
  - InverseProperties
  - FunctionalProperty
  - SymmetricProperty
  - TransitiveProperty

- Individuals
  - SameIndividual

OWL Classes

- Adds Equivalency and Disjointness for classes

  - Important to combine multiple OWL sources!
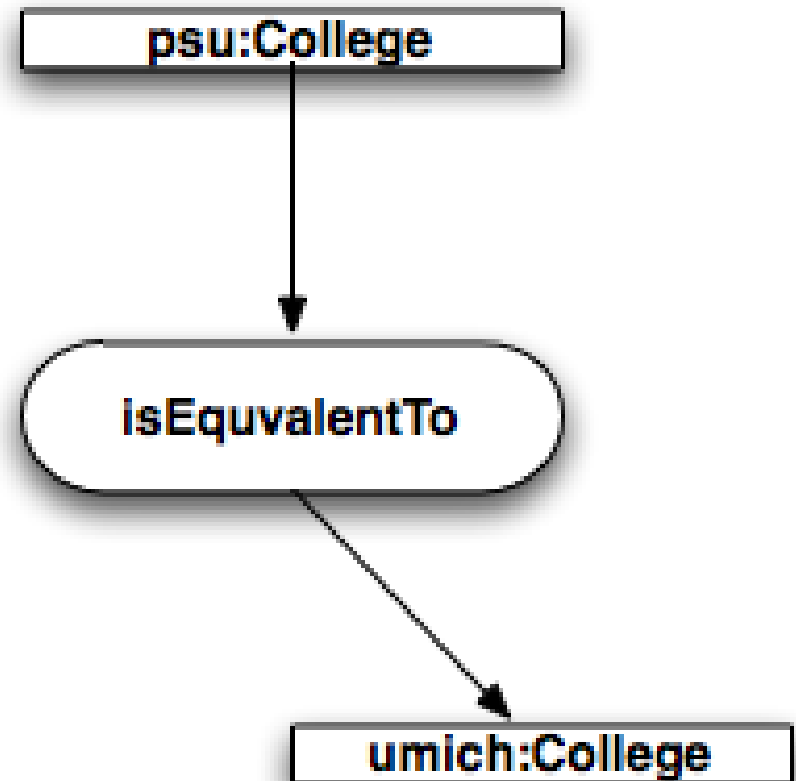
Multiple Inheritance
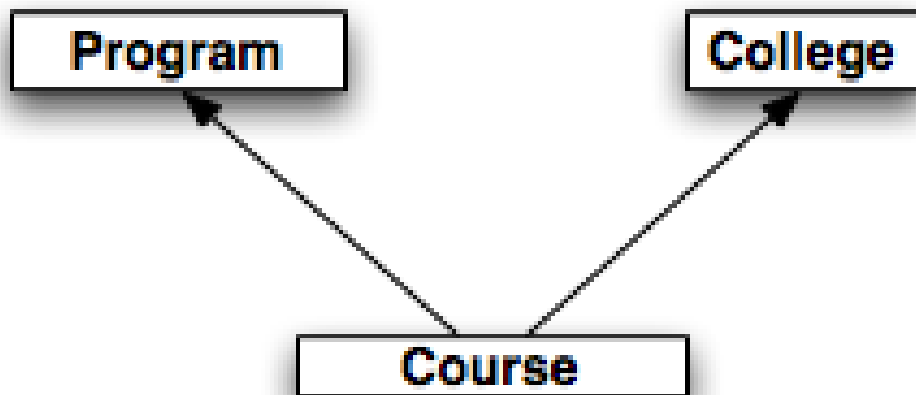
- Classes may have *more than one parent*

- Class structure is not a strict heirarchy

Drawing Inferences

As of September 2008

Inferencing

- *Inferences* are statements that are derived from other facts known to be true

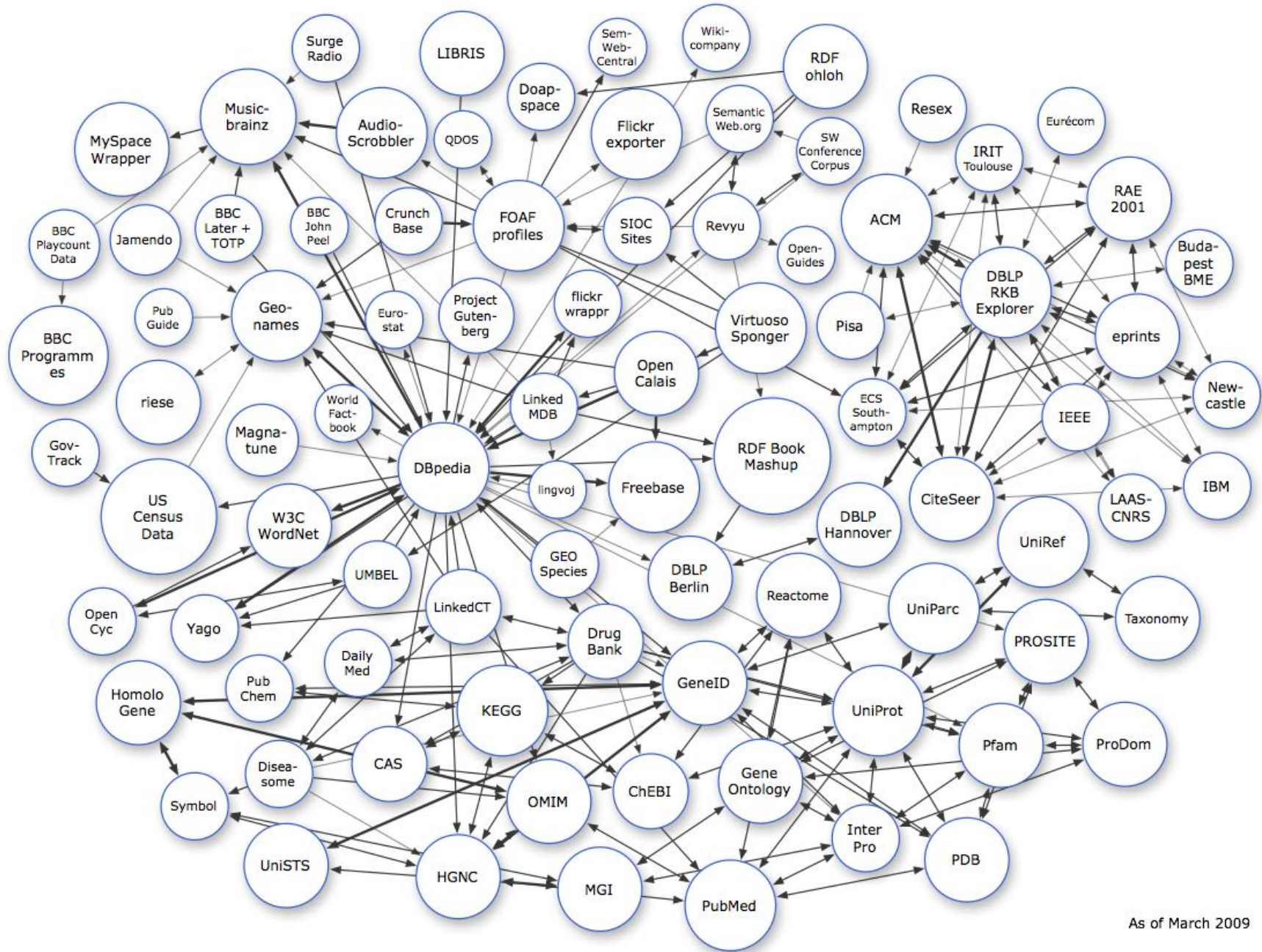- By declaring relationships as *equivalent, inverse, or transitive* we can make inferences on our data.

Transitive Inferences

- If
  - Portland is in Oregon
- and
  - Oregon is in the United States
- then
  - Portland is in the United States

open
connected

Inverse Inferences

- If
  - Christian *works with* Brian
- Then
  - Brian *works with Christian*

Linked Data

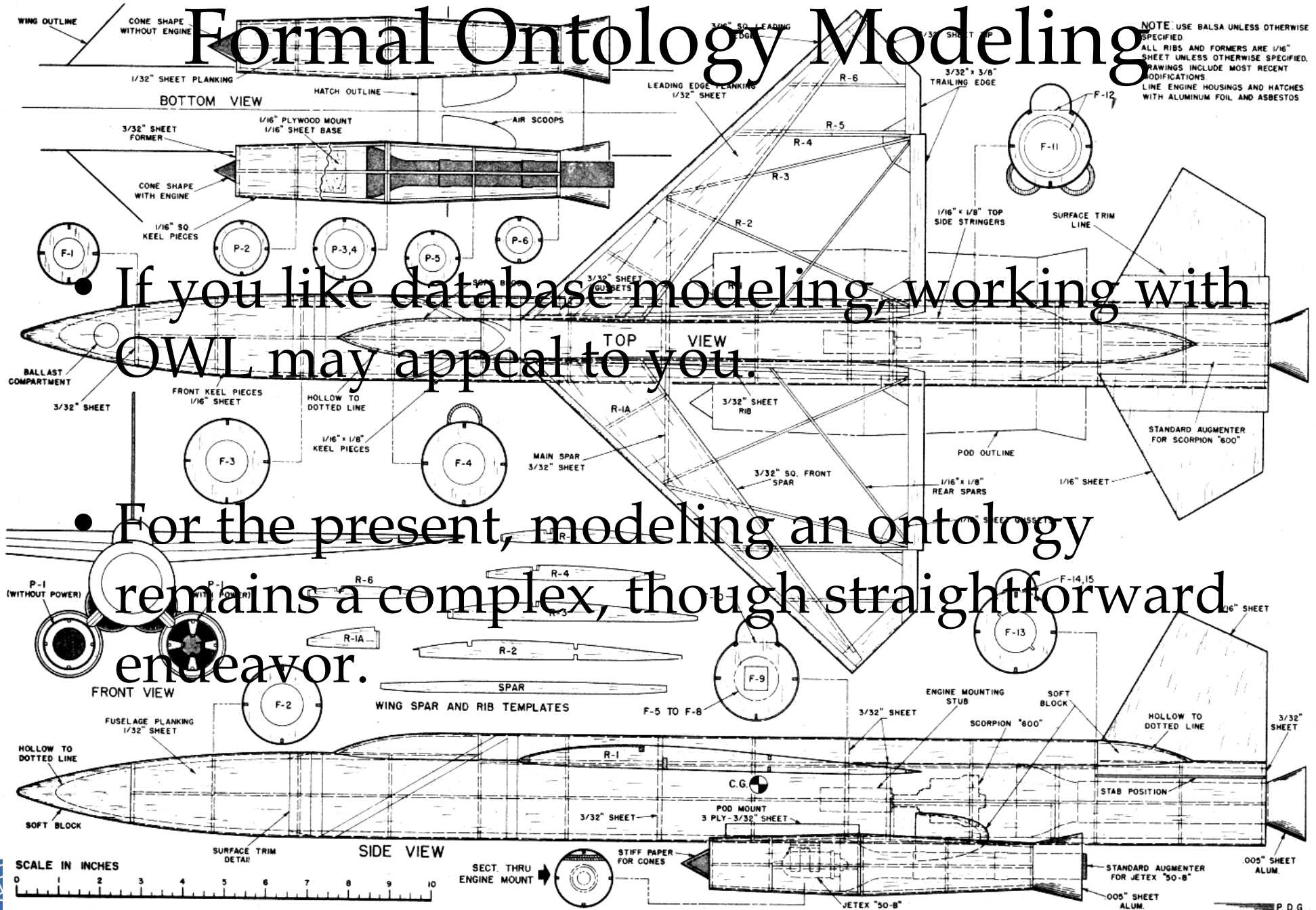As of March 2009

Using Ontologies

- Add a semantic data to your existing Web app by
  - Reading RDF/XML
  - Using an API (Java, Python, JavaScript)
  - Using a Semantic Store

- Query ontology files with SPARQL

Formal Ontology Modeling

- If you like database modeling, working with OWL may appeal to you.

- For the present, modeling an ontology remains a complex, though straightforward endeavor.

Ontologies: Do's

Building smarter data and smarter systems starts with *an ontology. To get started:*

1. *Look for an existing ontology for your domain/field*

2. Try building a small ontology to test your ideas

Ontologies: Don'ts

Things to avoid:

1. Don't try to model the universe.

2. Don't use the word 'ontology' in mixed company*.

* Philosophy students and non-Philosophy students

Ontologies: Do's

- Use a Upper/Foundational Ontology
  - BFO
  - Cyc
  - DOLCE
- Easier integrations later (?)
- Makes some philosophers happy
- Makes other philosophers unhapp

Ontologies: Don'ts

For more useful tips, see "Never Mind the Semantic Web (or, 13 Reasons Not to Let a Computer Scientist Choose a Name (or a Problem))"

http://www.furia.com/page.cgi?type=log&id=301#id301

open connected

Current Projects

PSU University Knowledge Engineering
https://weblion.psu.edu/trac/uke

Rubric Builder
https://weblion.psu.edu/trac/rubricgen

Smart Data

Good examples:

- http://dbpedia.org/
- http://www.geonames.org/
- http://linkeddata.org

Sample ontologies:

- http://semanticweb.org/wiki/Ontology

Resources
- PWC Technology Forecast
    http://www.pwc.com/us/en/technology-
        forecast/spring2009/index.jhtml

Contact Me

Brian Panulla

Twitter: @bpanulla

LinkedIn:

http://www.linkedin.com/in/brianpanulla

Backup

OWL Classes

- OWL classes replace RDFS classes

```
<owl:Class rdf:ID="AdministrativeUnit">
    <rdfs:subClassOf rdf:resource="http://www.lehigh.edu/
        ~zhp2/2004/0401/univ-bench.owl#Organization"/>
    <rdfs:label>Administrative unit</rdfs:label>
</owl:Class>
```
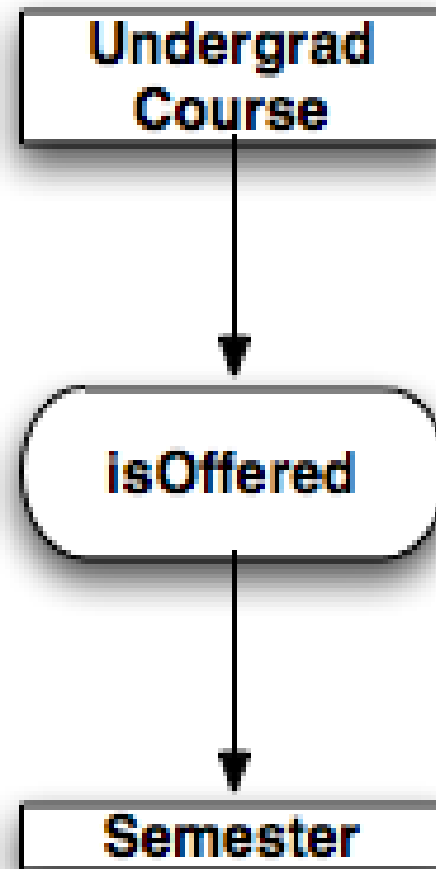
OWL

- OWL classes can have two types of properties: Datatype and Object

- Datatype properties: range of property is a simple atomic value

Course

label
description
courseNumber

OWL Properties

- Object properties:
  range of property is
  another OWL class

- Domain and Range can
  be specified for both
  types of properties

Undergrad
Course

isOffered

Semester

RDF Schemas

```xml
<rdfs:Class rdf:ID="AcademicProgram"/>

<rdfs:Class rdf:ID="DegreeProgram">
  <rdfs:subClassOf rdf:resource="#AcademicProgram"/>
</rdfs:Class>

<rdfs:Class rdf:ID="UndergraduateDegreeProgram">
  <rdfs:subClassOf rdf:resource="#DegreeProgram"/>
</rdfs:Class>

<rdfs:Class rdf:ID="Course"/>
```
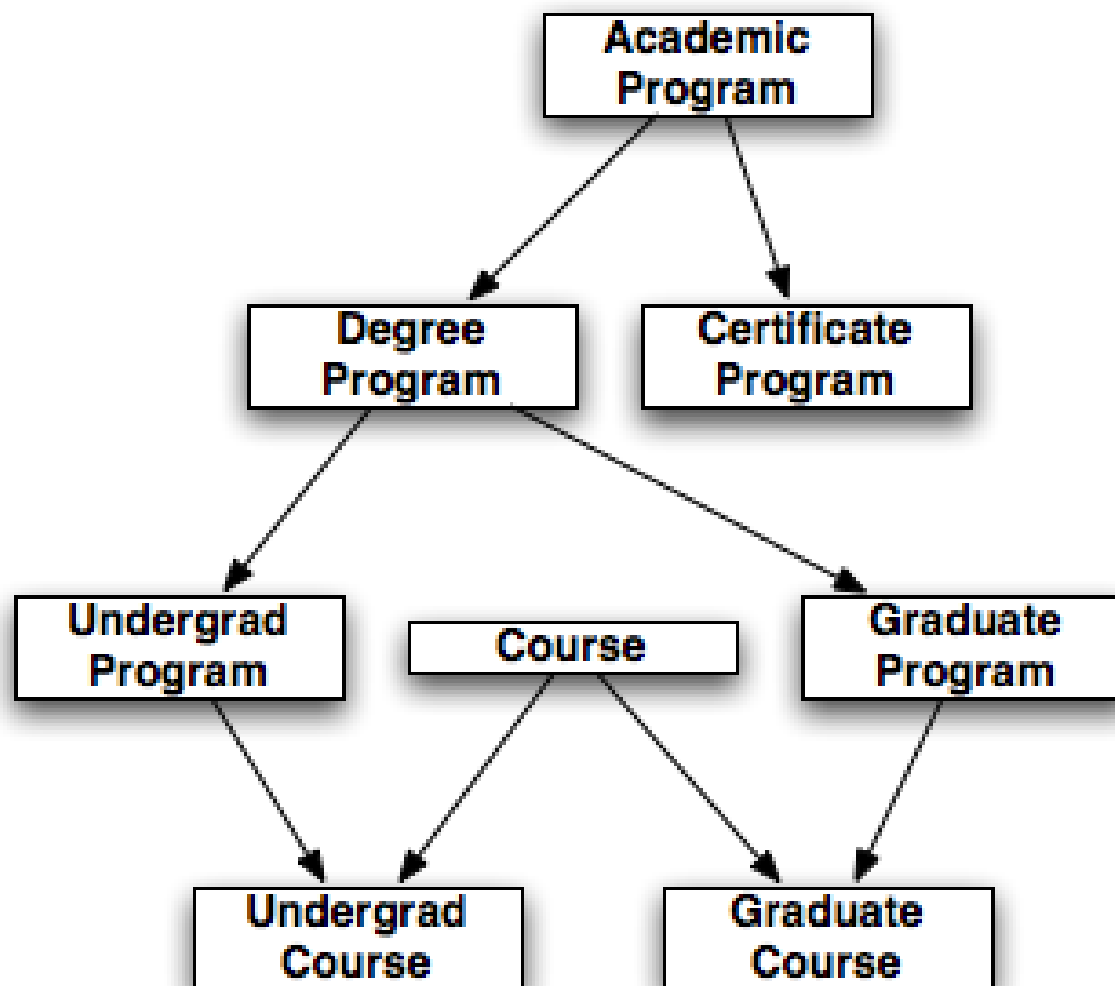
RDF Schema

Validating RDF

- RDF and RDFS documents can be validated by the W3C Validator:

  http://www.w3.org/RDF/Validator/

Will XML replace HTML? by John E. Simpson
December 13, 2000

This month, we tackle two related, neurosis-inducing questions common to Web developers just dipping their toes into XML.

Q: Will XML ever replace HTML?

A: Two answers, one philosophic and one pragmatic.

The philosophic answer is that XML isn't really meant, except incidentally, as a replacement for HTML. While the XML 1.0 Recommendation was under development, it was sometimes referred to as "SGML for the Web," and some residue of that perception remains. XML is superbly *adaptable* to the Web, true. But given the number of of XML-based markup languages possible (and the thousands already extant, for that matter), no imaginable Web browser could possibly figure out how to render all the corresponding documents. For example, what's a conventional browser to make of an `<employee emdID="emp73519">` tag? or an `<invoice_num>`? let alone an `<aperçu>`?

That leads us to the more pragmatic answer: XML is *already* replacing HTML...sort of.
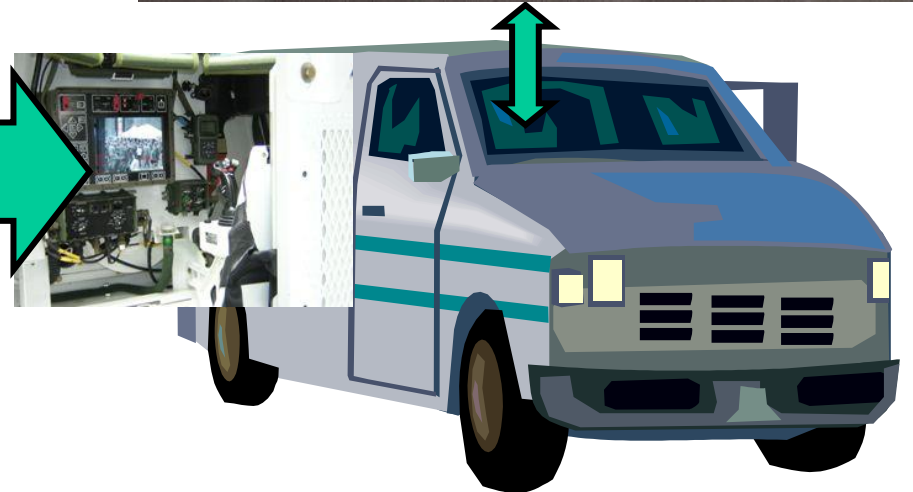
PSU Extreme Events Laboratory

PSU Extreme Events Laboratory

IST Command and Analysis Center

Ad hoc Community of Observers

Mobile van/command center

Differs from Syntax

- Rules of how symbols (words, letters, pictures) can be arranged.

I love Sushi

Semantics

- Can be unchanged despite symbols used

  I 💚 Sushi

Linked Data

- Distributed, linked datasets
- Moving rules and knowledge out of program code, into "documents"