Data Independence and The Semantic Web Roadmap

Michael C. Daconta
Chief Scientist, Advanced Programs Group
McDonald Bradley, Inc.
September 8th, 2003

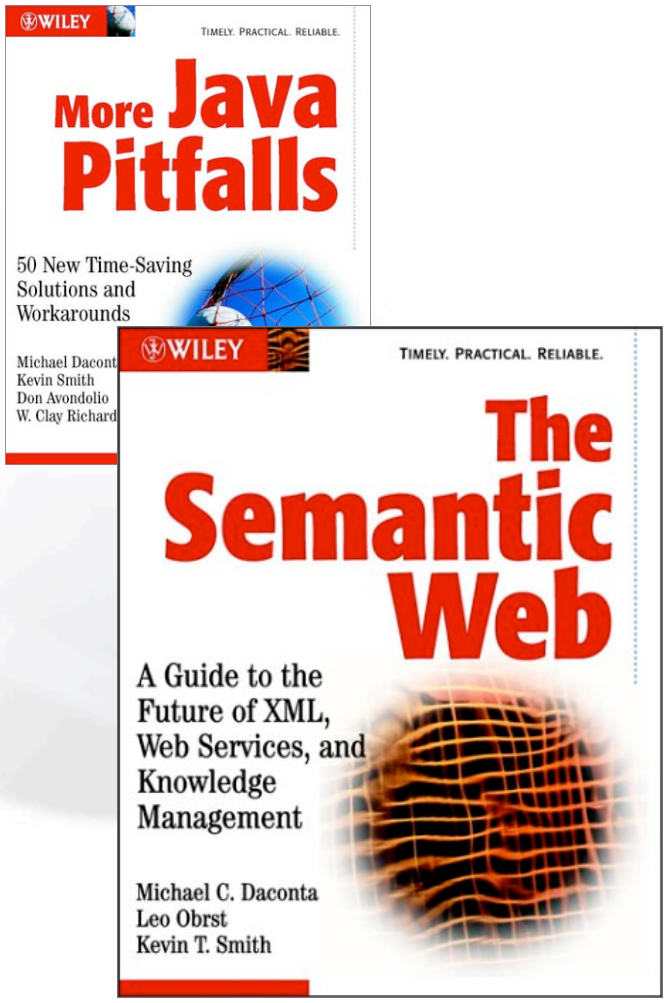
2250 Corporate Park Drive • Suite 500 • Herndon, VA 20171
703.326.1000 • http://www.mcdonaldbradley.com

Agenda

- Introduction
- The Semantic Web Book
- Declaration of Data Independence
- The Semantic Web Roadmap
- Conclusion

Introduction

- **Michael C. Daconta**
  - Chief Scientist, Advanced Programs Group, McDonald Bradley
  - Chief DIA Architect, VKB & Collateral Space/NCES
  - Author/co-author of 10 technical books
  - Inventor of Fannie Mae XML Electronic Mortgage Standards



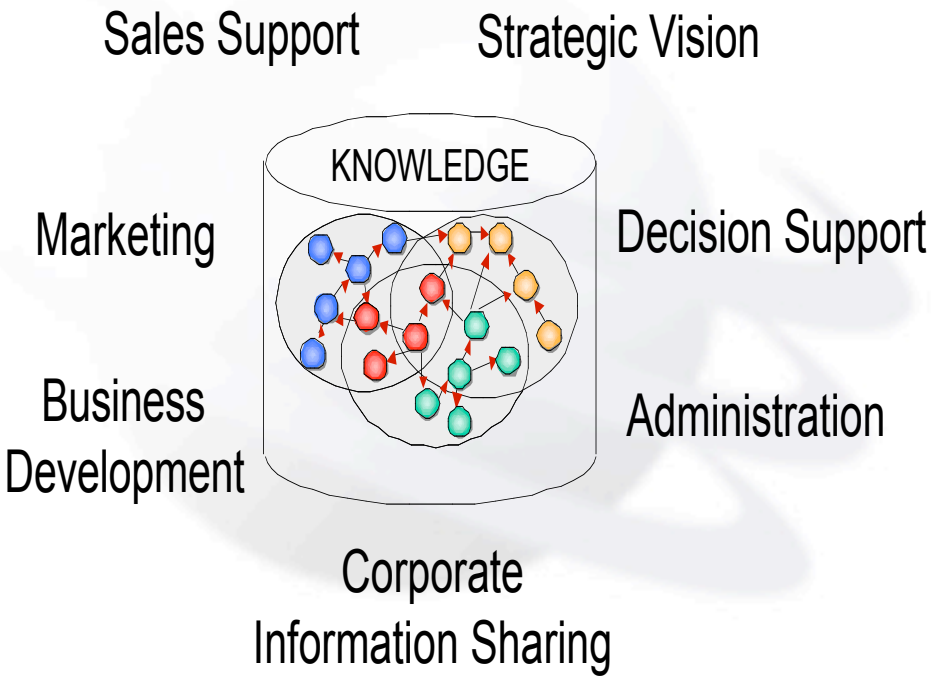
- **My Great Co-authors**:
  - Dr. Leo Obrst, Lead of Information Semantics Team at Mitre
    - Chapters 7, 8, part of 9
  - Kevin T. Smith, Chief Security Architect, McDonald Bradley
    - Chapters 2, 4, 6, most of 9

The Semantic Web Book

- **For CIO's**
  - What is the Semantic Web?
  - The Business Case for the Semantic Web.
  - Crafting your Company's Roadmap to the Semantic Web
- **Understanding where we are now…**
  - Understanding XML and its impact on the Enterprise
  - Understanding Web Services
  - Understanding the Rest of the Alphabet Soup
- **Understanding where we are going …**
  - Understanding the Resource Description Framework (RDF)
  - Understanding Taxonomies
  - Understanding Ontologies

Why do I need the Semantic Web?

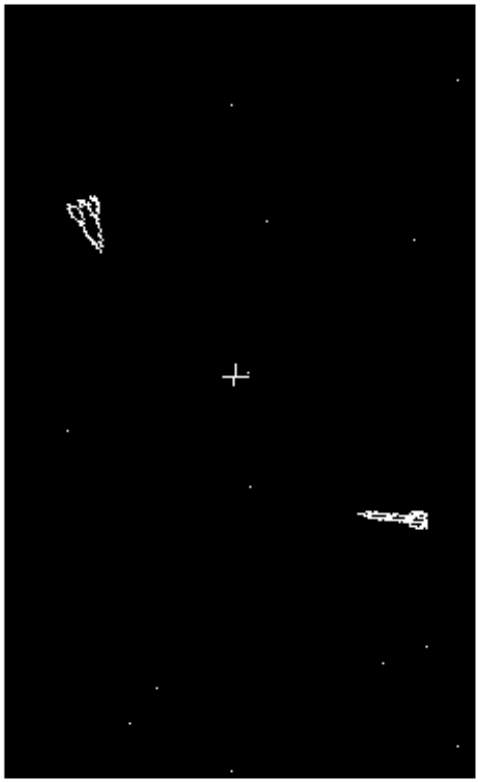
- **Solve Systemic Problems**
  - Information Overload
  - Stovepipe Systems
  - Poor content aggregation
  - Poor personalization
- **New Capabilities**
  - Decision Support
  - Expertise Capture
  - Agile Enterprise



Declaration of Data Independence

- 10 principles … some motherhood and apple pie.
- First step on the road to the semantic web.
- First 40 years we've perfected graphics fidelity …

In the next 40 years we will perfect data fidelity … let's begin!

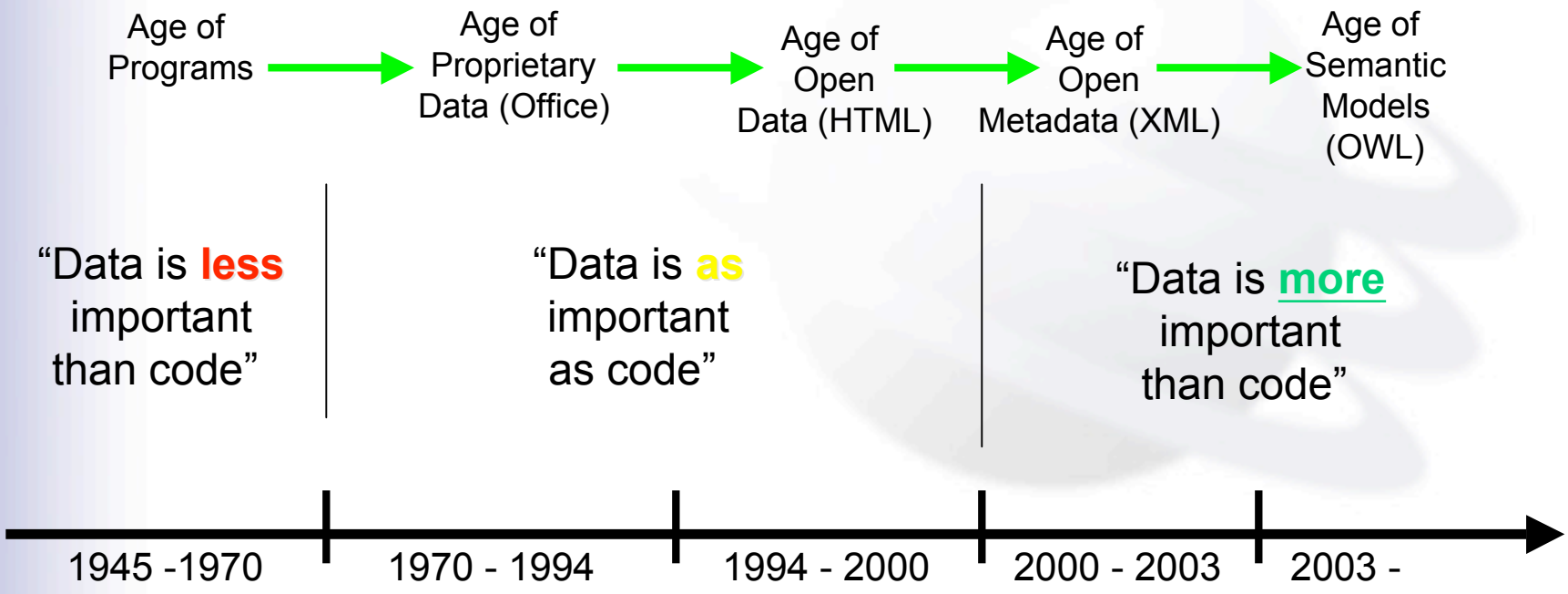


1961- SpaceWar

2001 - Halo

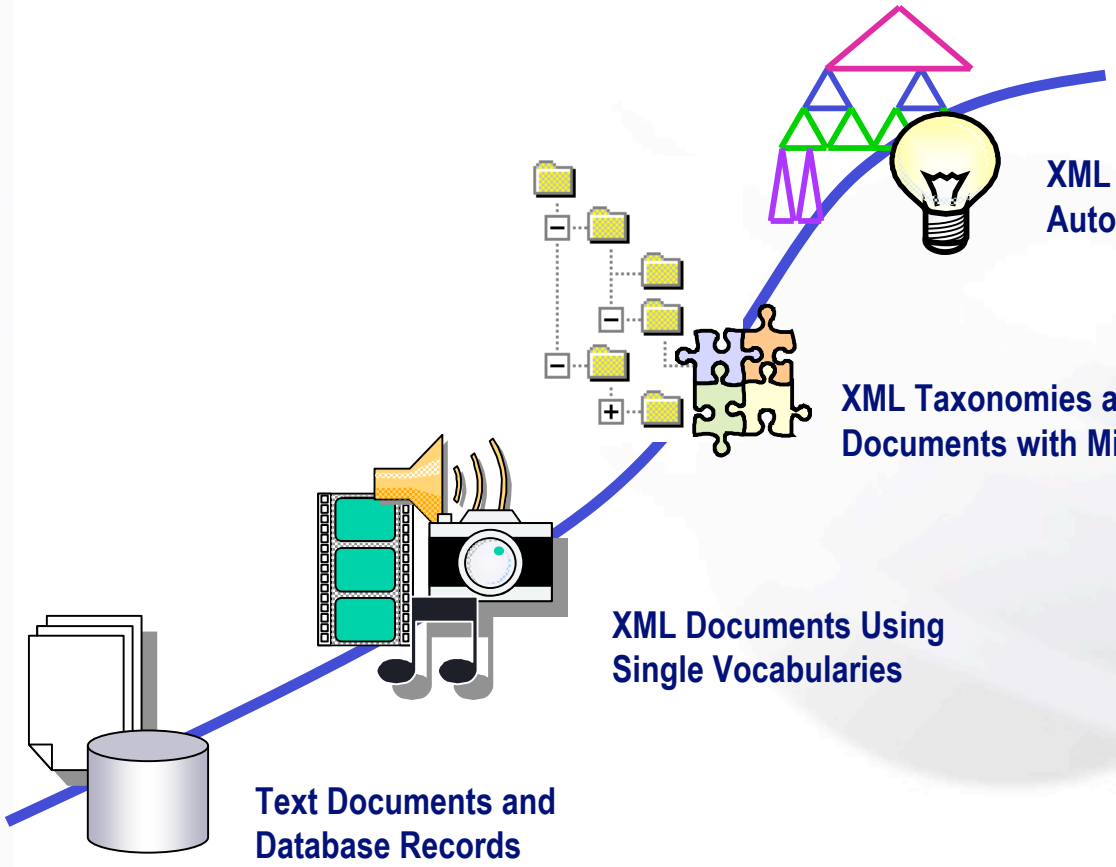
Data Independence – Principle 1

Data is more important than applications



The Data Evolution Timeline

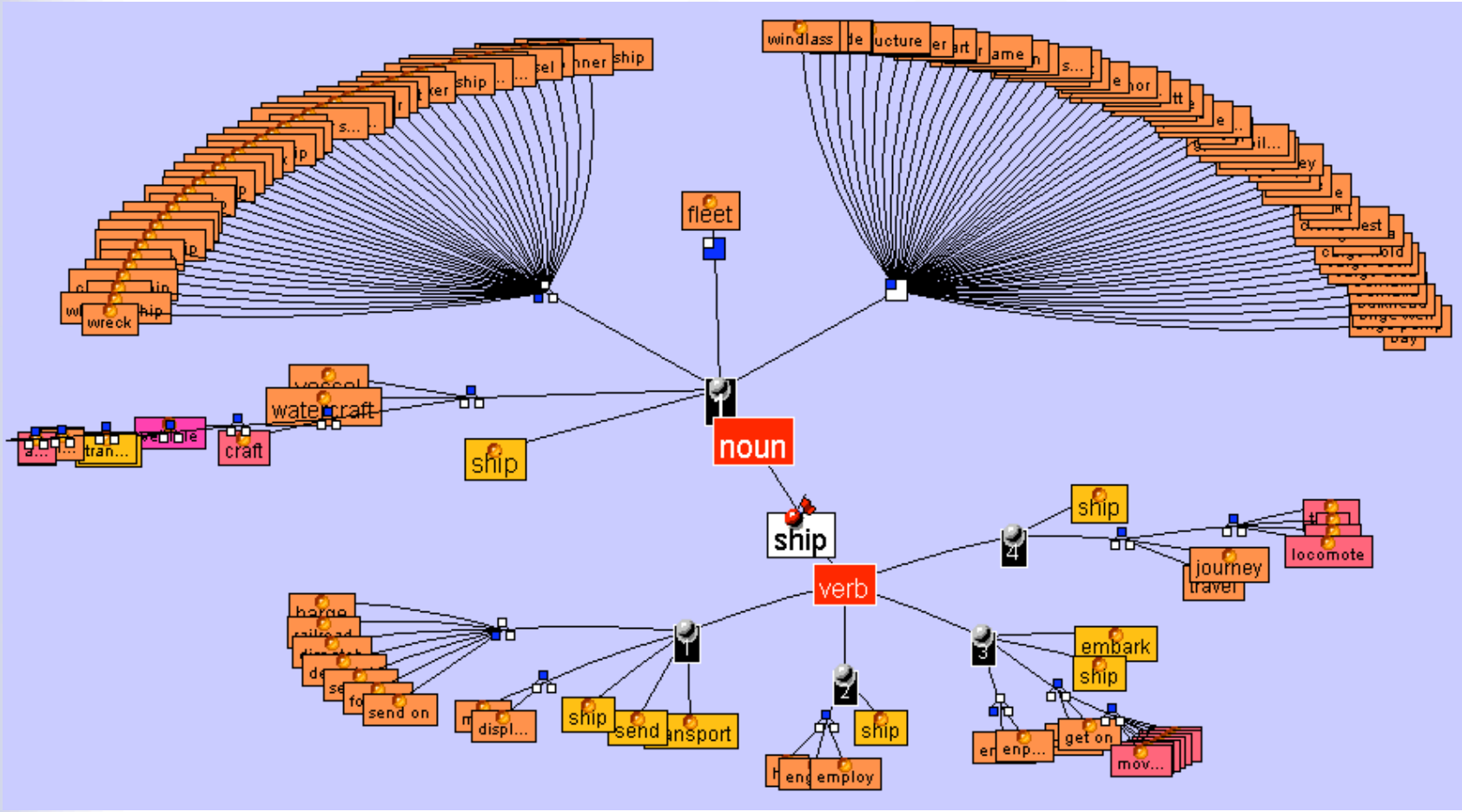
Smart Data Continuum



The trend is to put the "smarts" in the data, not in the applications.

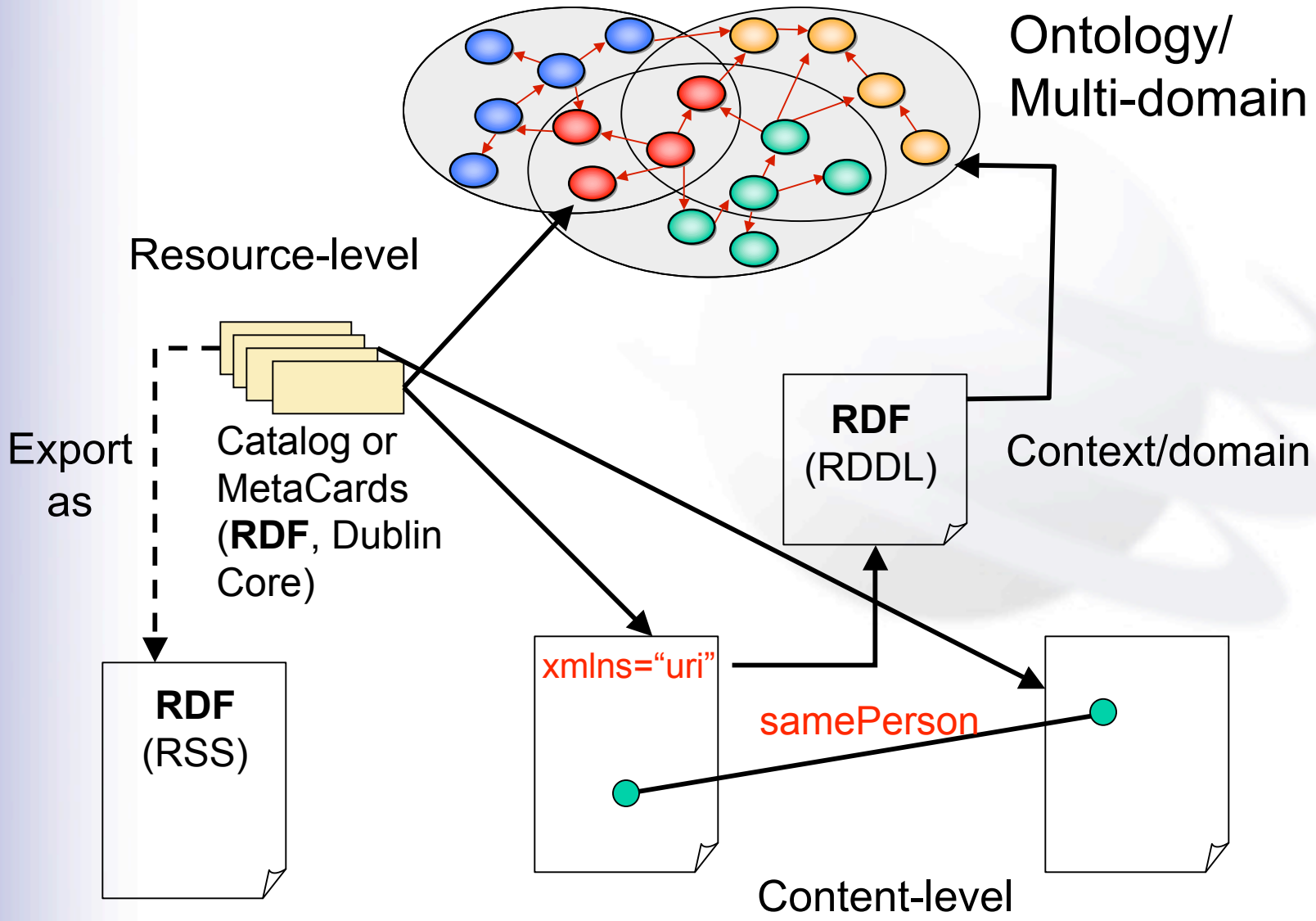
Data Independence – Principle 2

Data value increases with the number of connections it shares.



Apply the “network effect” to your data!

Link Types


Ontology/
Multi-domain
Resource-level
Export
as
Catalog or
MetaCards
(**RDF**, Dublin
Core)
RDF
(RDDL)
Context/domain
RDF
(RSS)
xmlns="uri"
samePerson
Content-level

Link Types (2)

- Resource-level: RDF and Dublin Core (card catalog)
- Content-level: (html, xlink, xpath)
- Context/domain links: namespaces, RDDL
- Ontology level (cross-domain): OWL object type/datatype properties
- Explicit (RDF) versus implicit (i.e. XML containment)

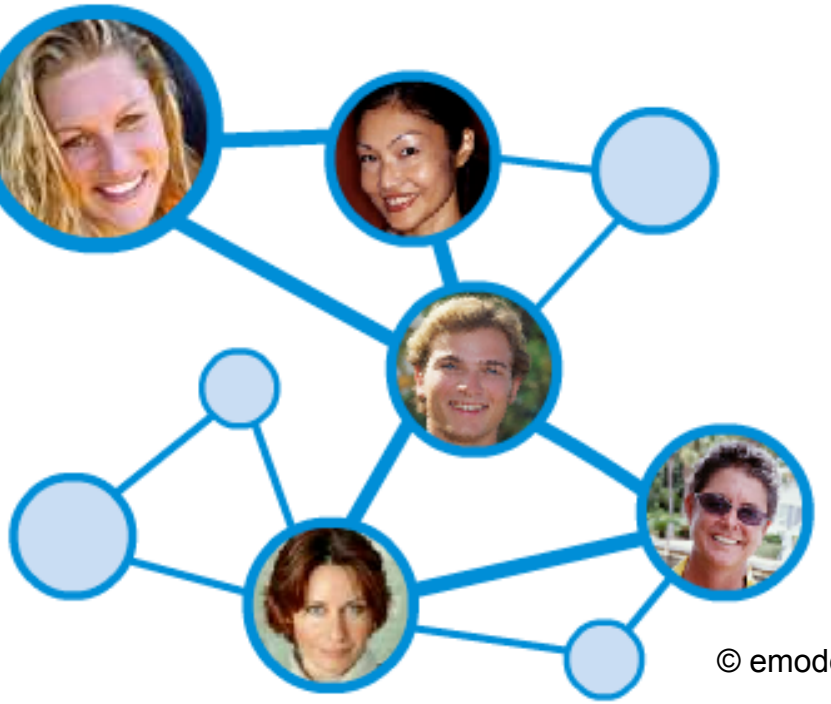
Explicit Linking example

- FOAF RDF vocabulary

```
<foaf:Person>
  <foaf:name>Mike Daconta </foaf:name>
  <foaf:knows>
    <foaf:Person>
      <foaf:name> Kevin Smith </foaf:name>
    </foaf:Person>
  </foaf:knows>
</foaf:Person>
```

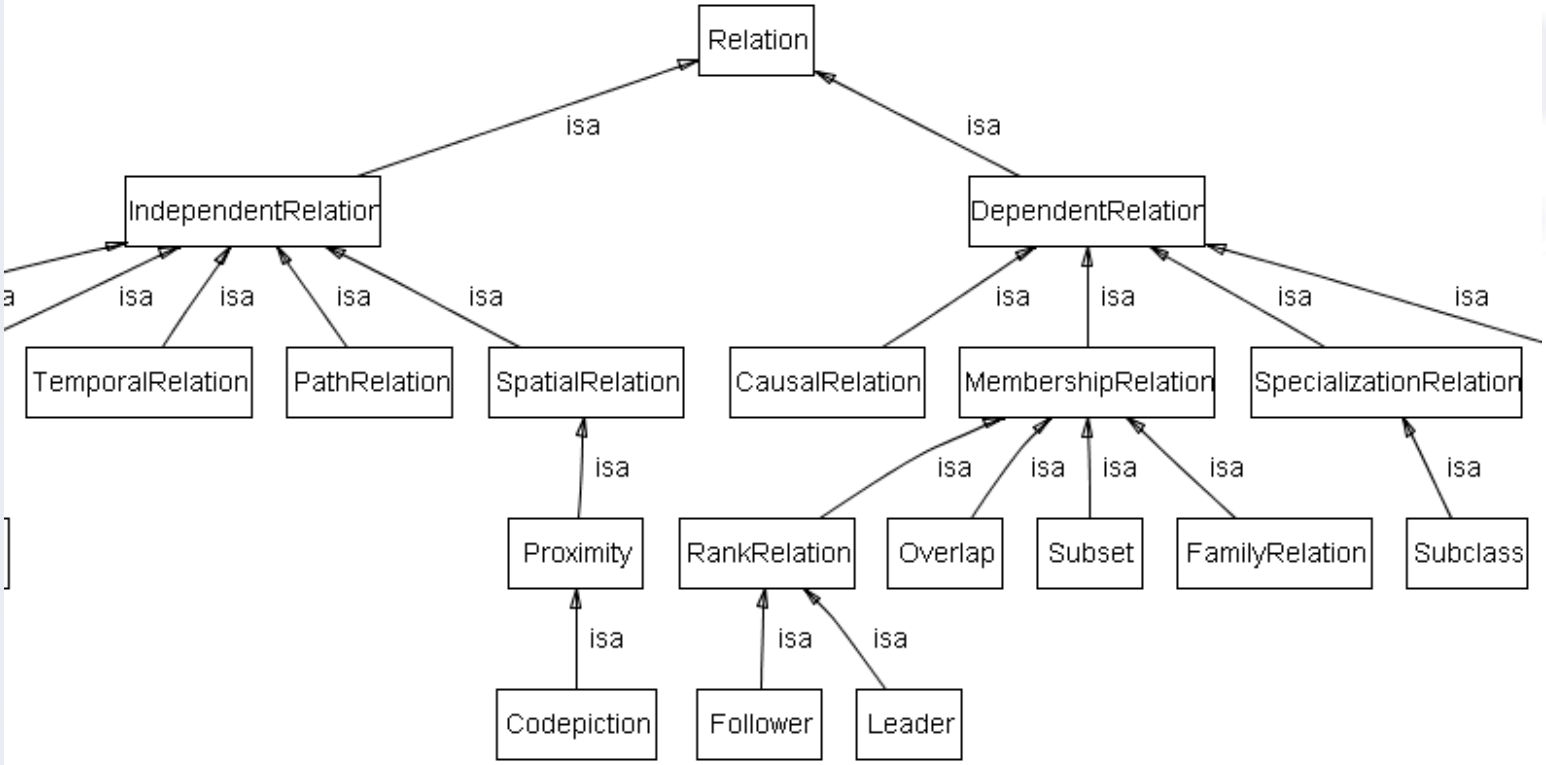
- Copy Cats
  - Friendster
  - Friends network

Use Explicit Links!



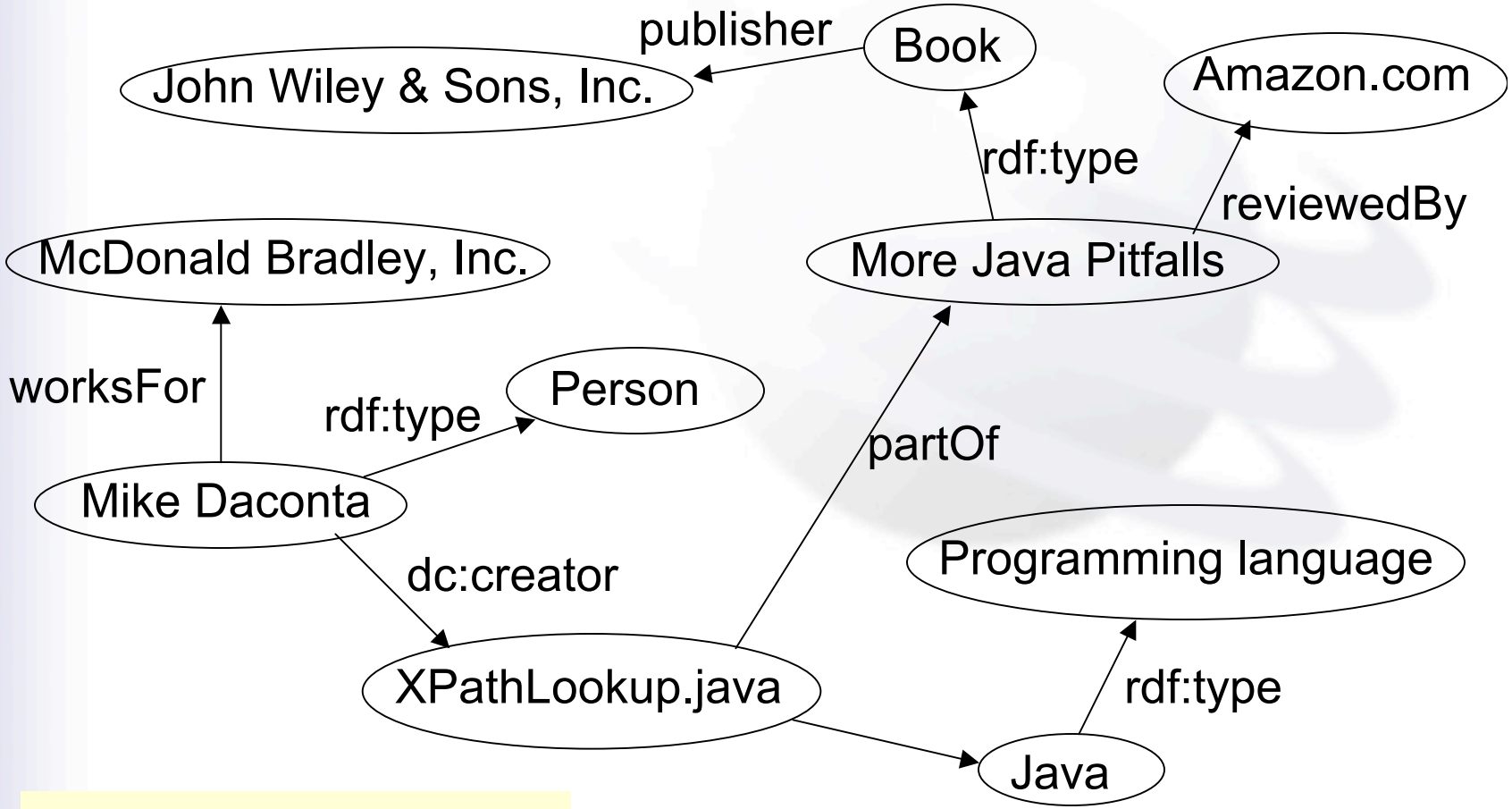
Relation Mining is the next Killer App!

- Clear mandate: **Connect the dots!**
- Need more effort modeling relations instead of things!
  - i.e. Programming languages do it wrong
  - OWL does it right. Relation versus Characteristic.
  - CYC has a large set of complex relations.
- Predictive intelligence requires detailed relation modeling.



Data Independence – Principle 3

Data about data can expand to as many layers are there are meanings



Let metadata proliferate!

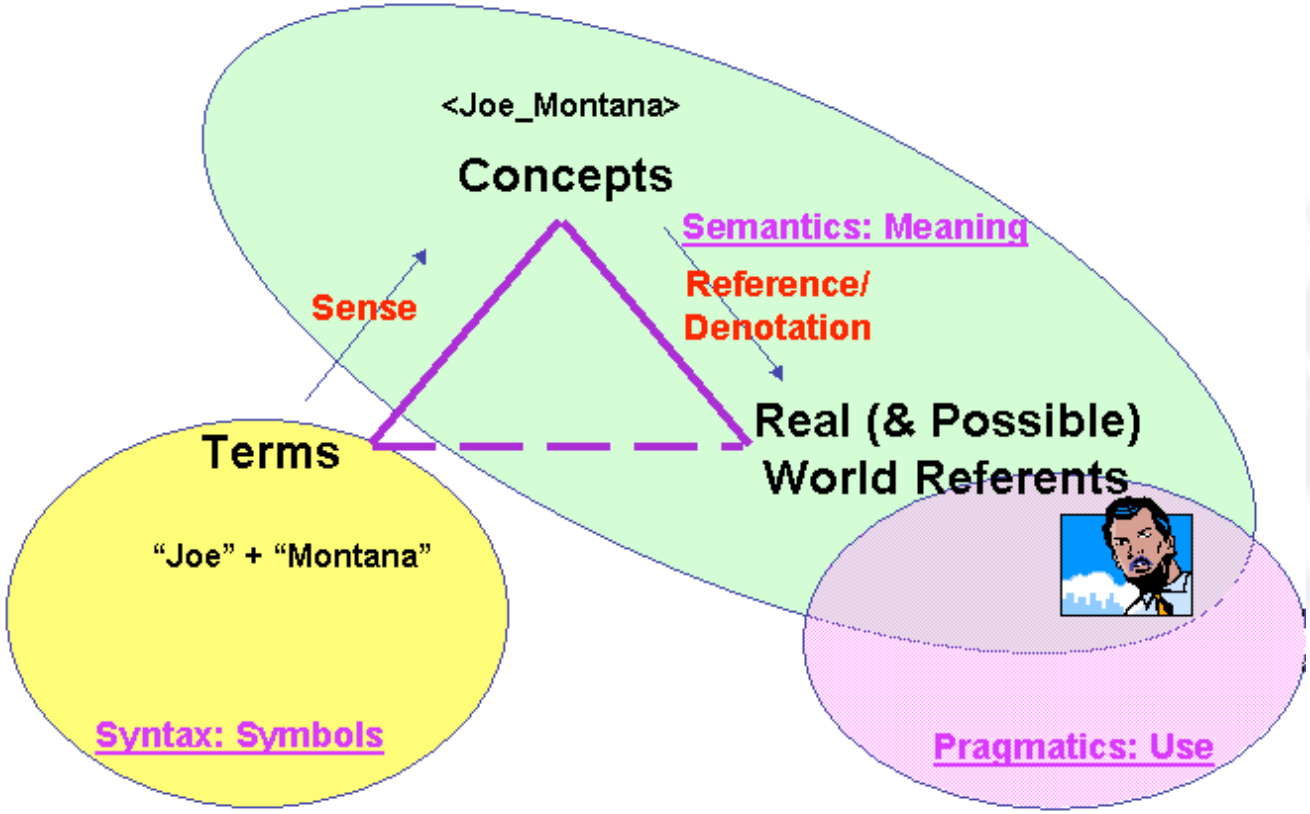
Metadata layers

- Do you think we can have too much metadata? Fugghedaboutit !!! (that means no)
- User context modeling is weak.
  - Choice stream
  - OpenTable
- GPS, cell-phones and voice recognition will ratchet up the requirements for **"real-time relevance!"**
- 64-bit computing will allow us to search massive data graphs in real-time!

Metadata layers allow multiple contexts!

Data Independence – Principle 4

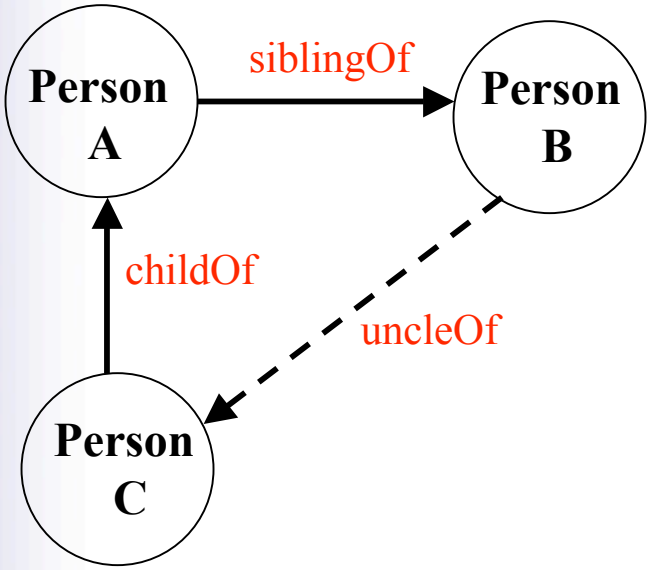
Data modeling harmony is the alignment of syntax, semantics and pragmatics



Triangle of Signification

Data Independence – Principle 5

Data and logic are the yin and yang of information processing



Rules

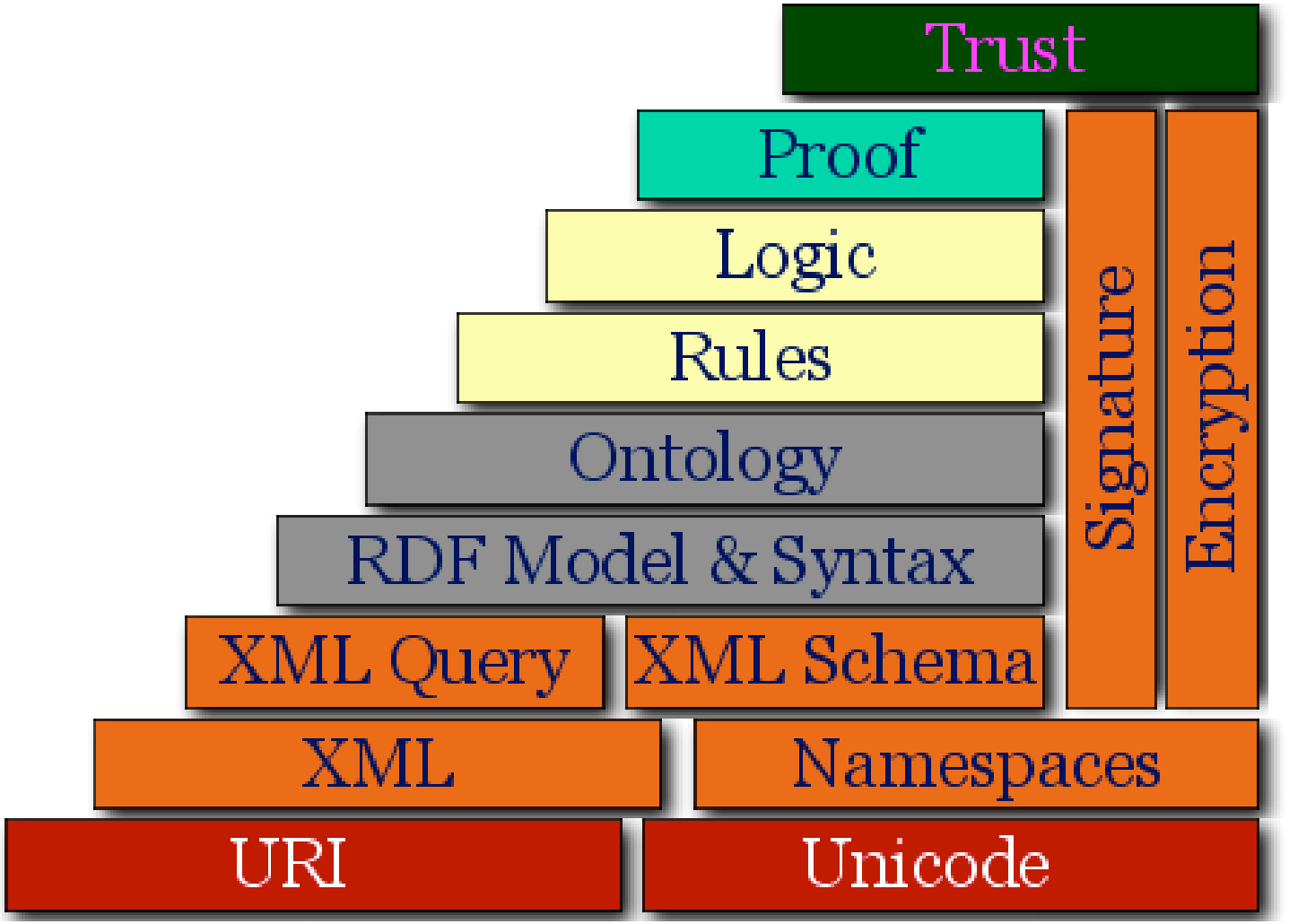
```
if (C.gender == "male" AND
     C == childOf(A))
    then C = sonOf(A);

if (B.gender == "male" AND
     B == siblingOf(A))
     then B == brotherOf(A);

if (C == sonOf(A) AND
     B == brotherOf(A))
     then B = uncleOf(C);
```

- Two given relations and one inferred relation (uncleOf)

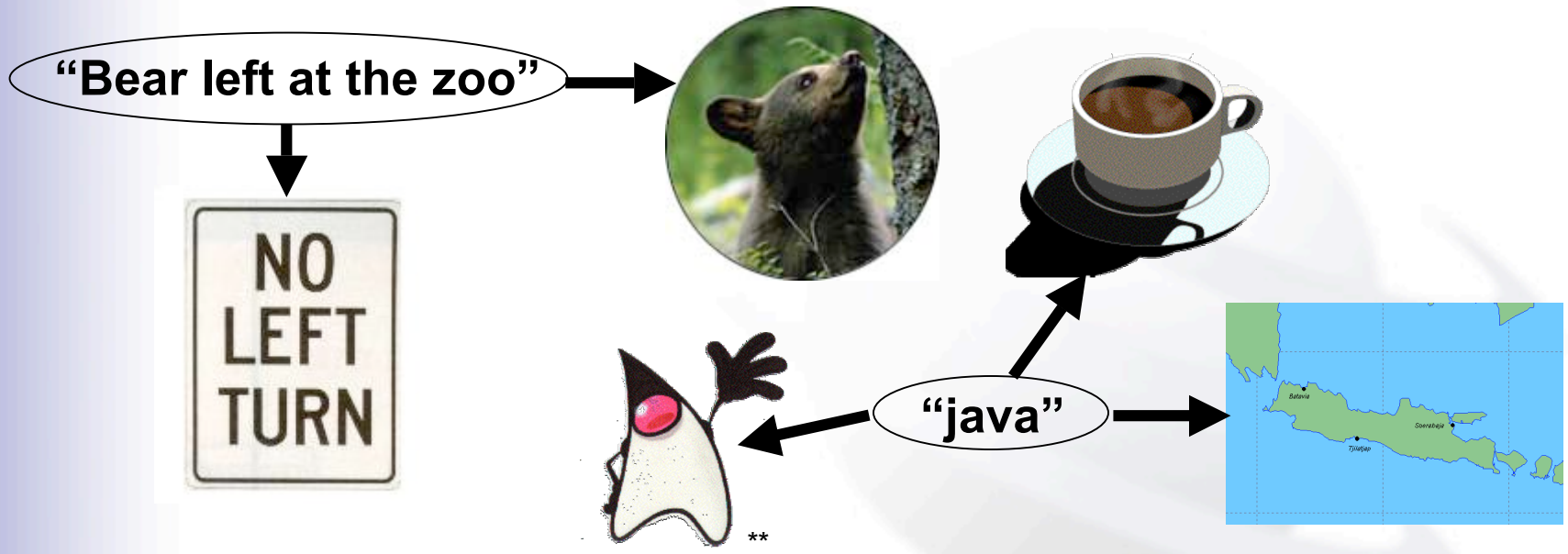
Semantic Web Stack



Ontology first, then rules!

Data Independence – Principle 6

Data modeling makes the implicit explicit and the transparent apparent



RDF gems:

- URIs for concepts
- Explicit relations
- Simple Foundation (S,P,O)

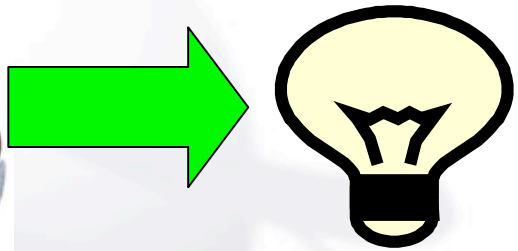
OWL gems:

- Transitive
- InverseFunctional
- Symmetric
- Many more!!

Data Independence – Principle 7

Data standardization is not amenable to competition

© Times Newspapers Ltd.

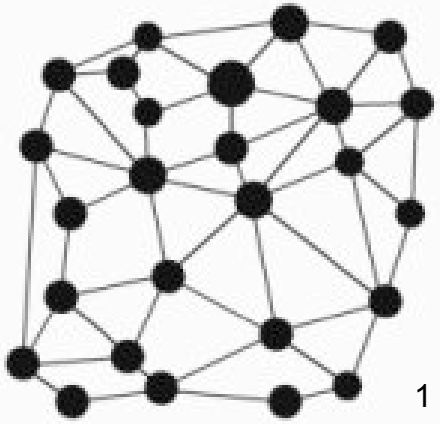


Government can lead here, if bureaucracy doesn't bog down the process

But you must <u>be aggressive</u>! Remember, Perfect is the enemy of the Good.

Data Independence – Principle 8

Data modeling must be decentralized



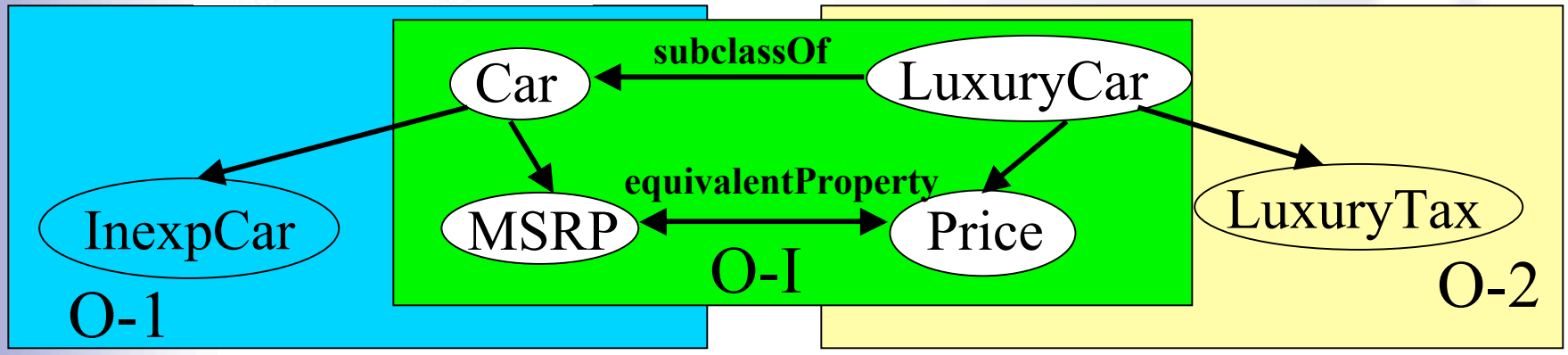
Avoid the expertise acquisition bottleneck!

Distributed…

Collaborative…

Ontology Mapping with OWL…

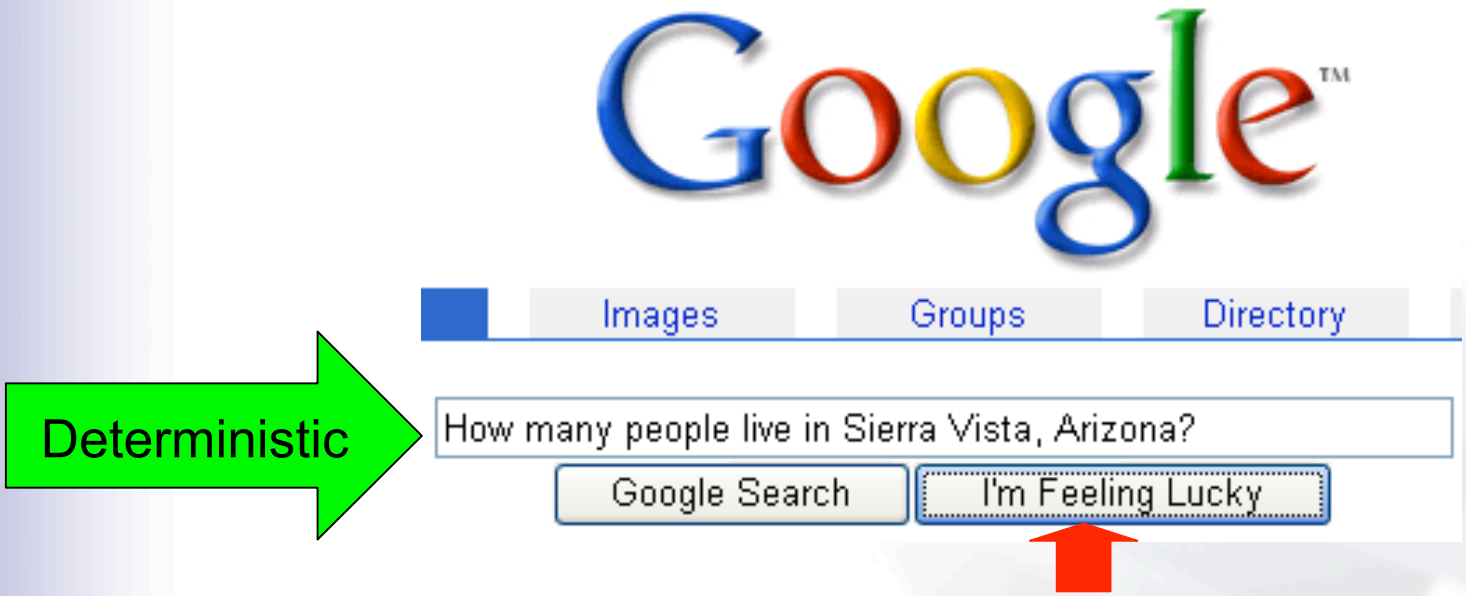
equivalentClass, equivalentProperty, sameAs, differentFrom, AllDifferent



The semantic web is the only practical way to achieve a global, general-purpose expert system.

Data Independence – Principle 9

Data relations must not be based on probability or luck.



"Is this the best we can do?"
Answer: Google doesn't think so.

"We need more content and ways to interact with search. For instance, you can't ask a question and have it answer you."
- Marisa Mayer, Google's Director of Consumer Products

For discovery, prefer deterministic over probabilistic approaches!

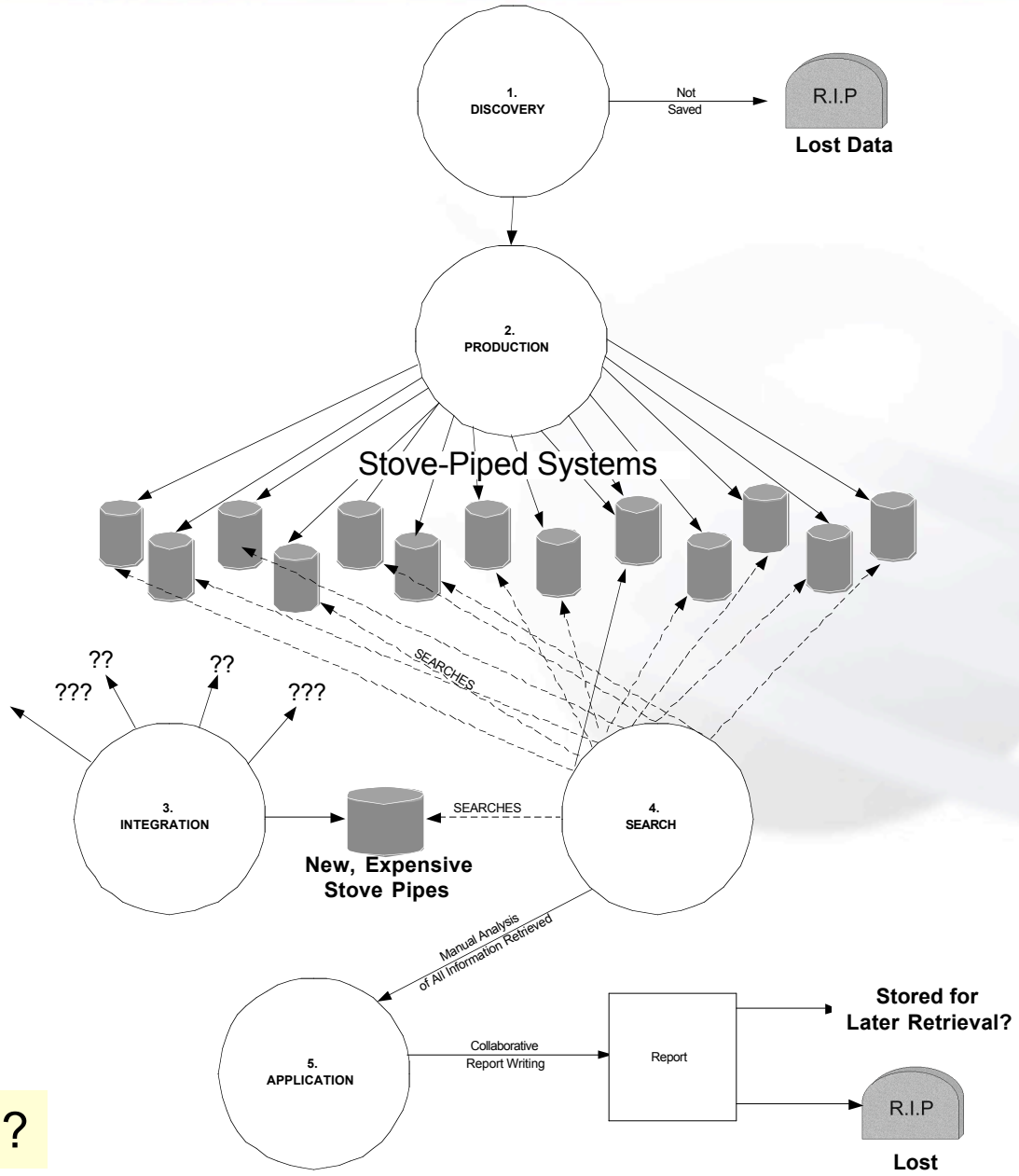
Data Independence – Principle 10

Data is truly independent when the next generation need not reinvent it

"We're a society that's used to losing information to new generations. Now more than ever, search is really important."

- Larry Page

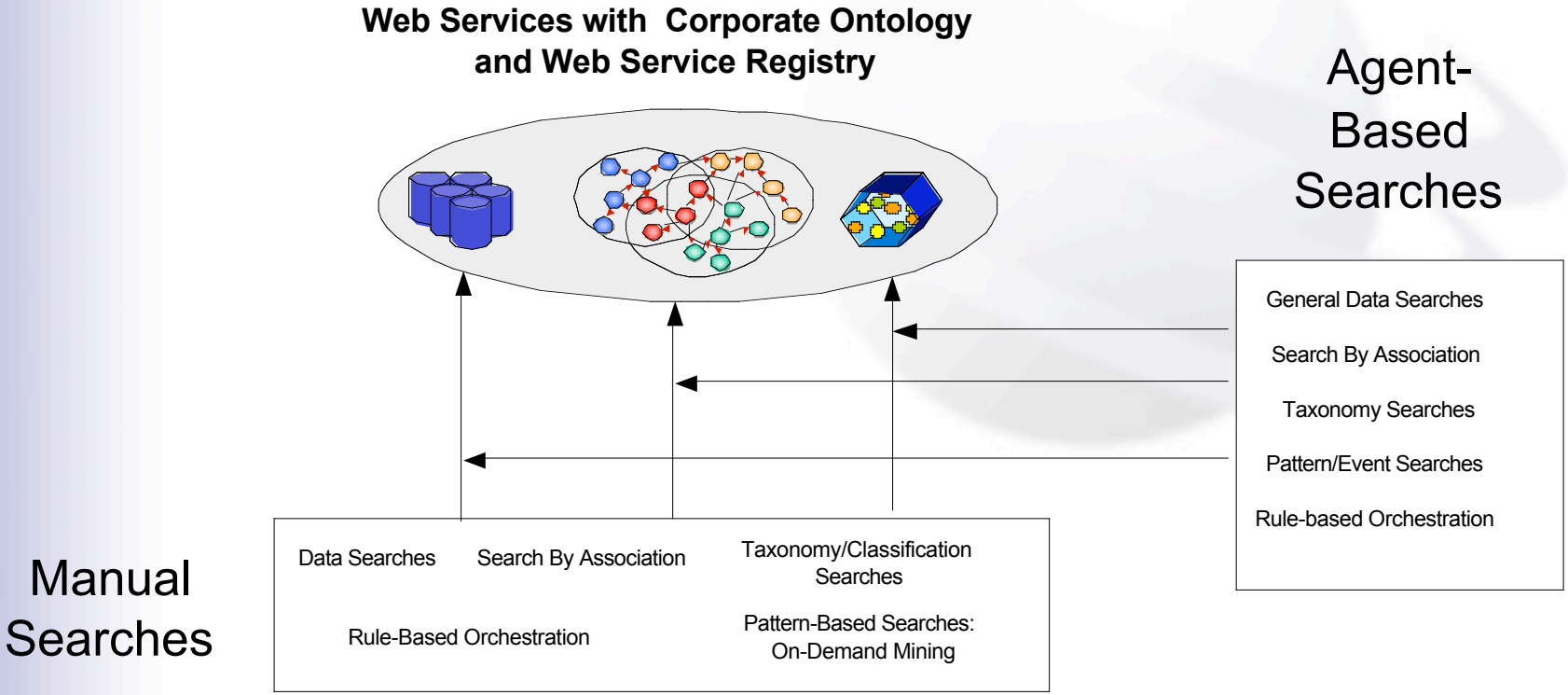
The Semantic Web Roadmap



Is This You?

Semantic Web Roadmap (2)

- Where you want to be…
- Climbing the smart data continuum…
  - XML Schema (fine-grained), domains, RDF catalog cards, **relations**, taxonomy, thesaurus, OWL ontology, Reasoning…



A journey of a thousand miles begins with a single step. -- Lao-tzu

Conclusion

- As We May Think[1]…
  - "Our ineptitude in getting at the record is largely caused by the artificiality of systems of indexing. …
    The human mind does not work that way. It operates by association. …
    Selection by association, rather than indexing, may yet be mechanized."
    - Vannevar Bush, 1945

- The Semantic Web is "Crossing the Chasm" now.
  - We'll see the tipping point within three years.
  - Businesses will see it in portals.
  - Consumers will see it in the integration of email/calendar/contacts with personal knowledge bases (music, video, vacation, etc.)

All Aboard!

© Ocean Cruise Guides

1 © Vannevar Bush