AusWeb06 W3C SIG

NICTA

Semantic Interoperability:
Vision, Issues and Strategies
July 5, 2006

Anne Cregan
Knowledge Representation & Reasoning Program
NICTA / UNSW PhD Student

What is NICTA? Objectives

NICTA = National Information and Communication Technology Centre of Excellence Australia

MISSION:

To be an enduring world-class research institute in Information and Communications Technology that generates national wealth for Australia.

- **Strategy:**

  Based on the "four pillars" of
  - **Research,**
  - **Commercialisation,**
  - **Education**
  - **Collaboration.**

- **Focus:**
  - **Reversing the "brain-drain":**
    attracting, developing, and networking exceptional talent for Australia's future prosperity.
  - **Use–inspired basic research:**
    NICTA's research efforts focus on the technology challenges facing industry, community, and the national interest.
  - **The central drivers of NICTA's research are its two *Priority Challenges*.**

What is NICTA? Structure

NICTA = National Information and Communication Technology Centre of Excellence Australia

NICTA's members and partners share the common vision of building Australia's Information and Communication Technology (ICT) research capability with the long-term aim of generating economic wealth for Australia.

- **NICTA currently has**
  - **Seven Sites:**
    - **Sydney(2),**
    - **Canberra(2),**
    - **Melbourne,**
    - **Brisbane**
    - **Adelaide**
  - **15 Research Programs**
  - **Plus liaison with Government and Industry**
  - **over 500 employees and NICTA-endorsed students**

Initiated in 2002 by :

- **the Federal Department of Communications, Information Technology & the Arts**
- **the Australian Research Council.**

NICTA Founding members :

- **ACT Government,**
- **NSW Government,**
- **the University of New South Wales,**
- **the Australian National University.**

Current Partners :

- **The University of Sydney**
- **Victorian Government**
- **University of Melbourne**
- **Queensland Government**
- **Griffith University**
- **Queensland University of Technology**
- **The University of Queensland**

Introduction to NICTA Research

and the environment, the central drivers of NICTA's research efforts are its Priority Challenges.

NICTA currently has two priority challenges:

1. Trusted Wireless Networks
2. From Data to Knowledge

Introduction to NICTA Research

and the environment, the central drivers of NICTA's research efforts are its Priority Challenges.

NICTA currently has two priority challenges:

NICTA Priority Challenges

NICTA Priority Challenges

Find out more about NICTA and its research at:

www.nicta.com.au

The Semantic Web Vision

Tim-Berners-Lee's original vision for the web :

"A plan for achieving a set of connected applications for data on the Web in such a way as to form a consistent logical web of data."

Tim Berners-Lee

The Semantic Web Vision is to have:

- **Interoperability of data across applications**
- **Consistent, logical, machine-accessible data enabling:**
  - **Smarter Search Engines**
  - **Efficient Automation**
  - **Powerful Web-based Services**

NATIONAL ICT AUSTRALIA

Why do we need Semantic Interoperability?

Motivation

1. There's a lot of data out there, both public and private. Significant effort has been expended collecting it, maintaining it, storing it and analysing it

2. Each data store is designed with a particular purpose in mind

Each has its own:
- vocabulary
- data formats and values
- syntax
- structure
- semantics

3. We want to share information and leverage it to obtain social, economic and environmental value.

➡ We need ways to intelligently query, infer, and reason over the amassed data

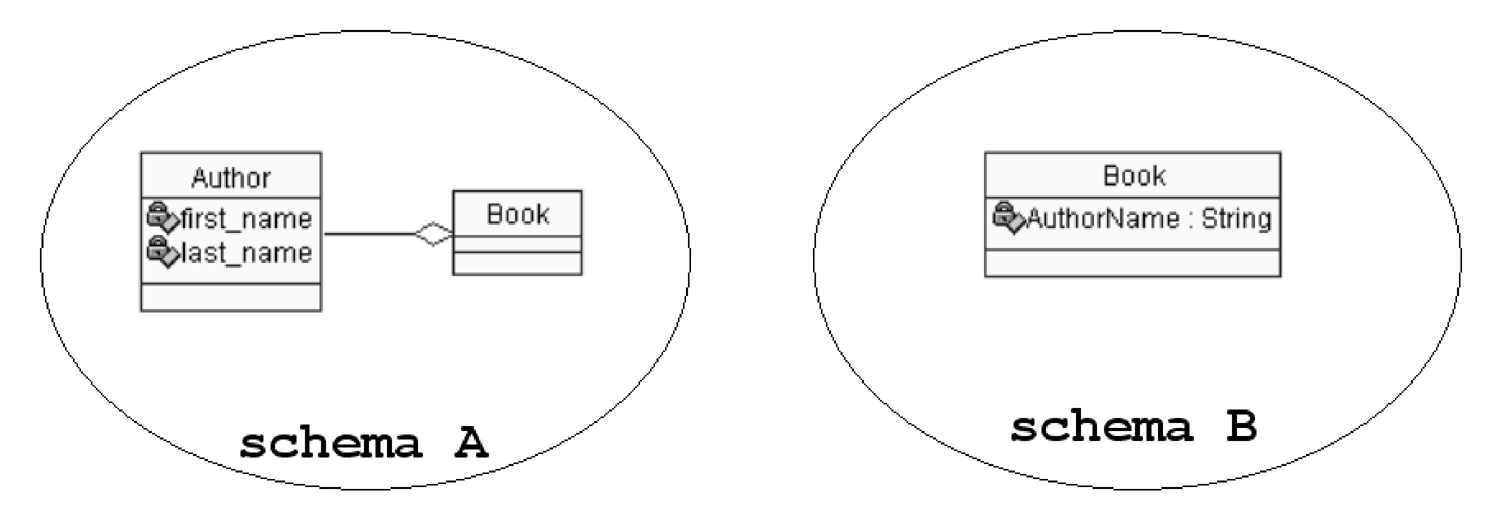
Interoperability

A key question is how to make different data sources interoperable –

Let's look at some of the challenges ….

Interoperability

Structural Incompatibility:

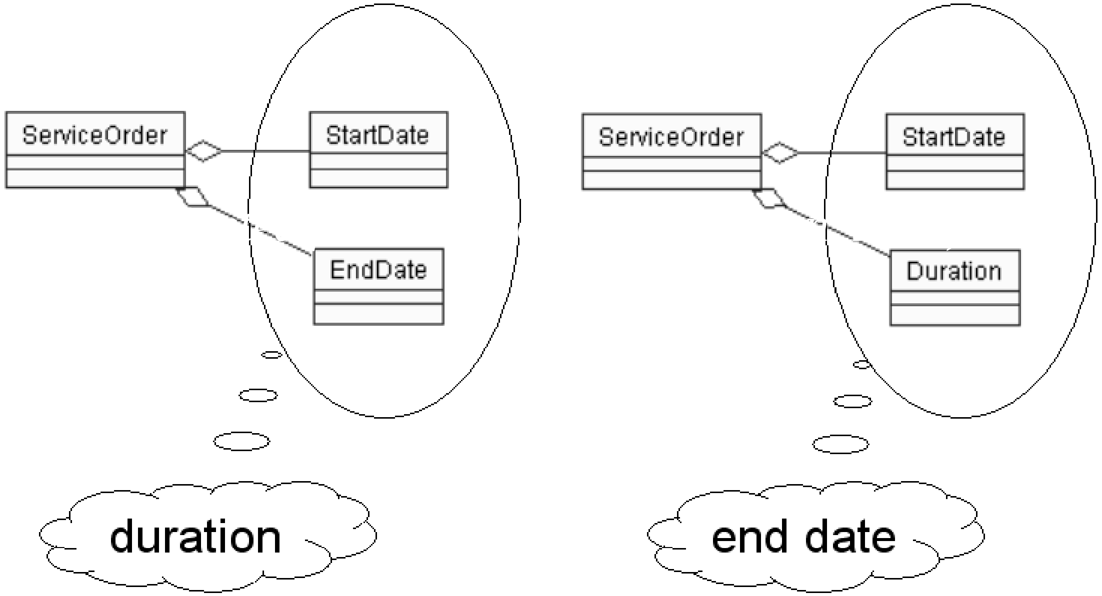


Aggregation conflict – difference in structure

(Adapted from Pollock and Hodgson, 2004)

Interoperability

Representation Incompatibility:

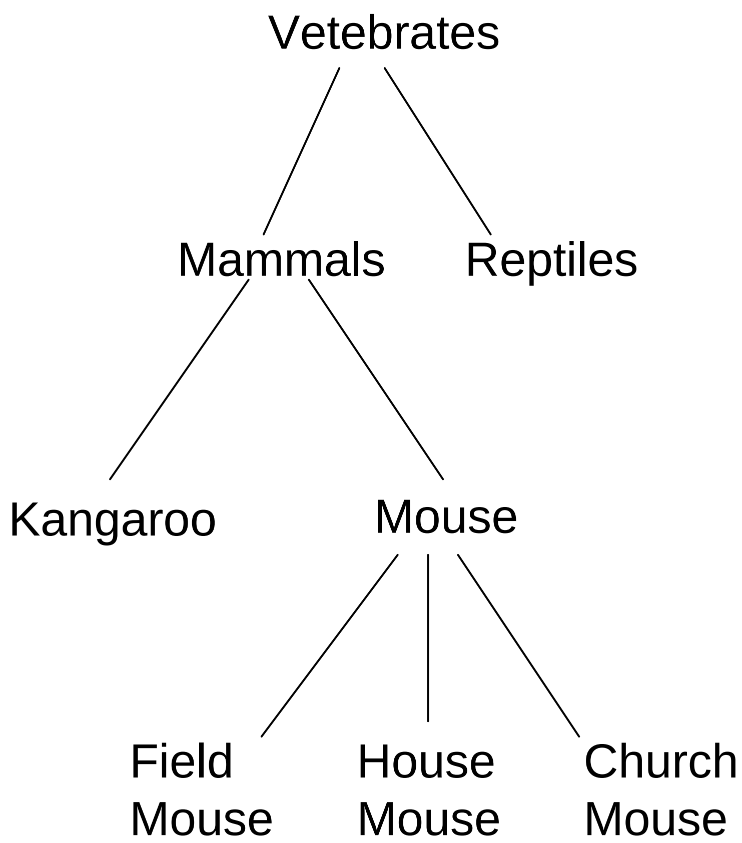


Value representation conflict

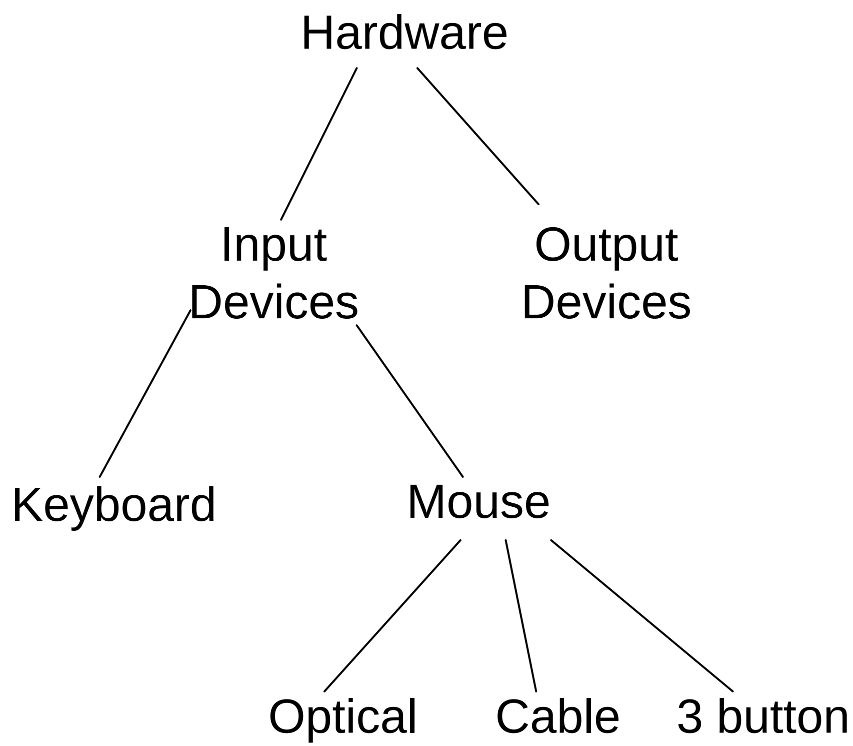
(Adapted from Pollock and Hodgson, 2004)

Interoperability

Ontology A:

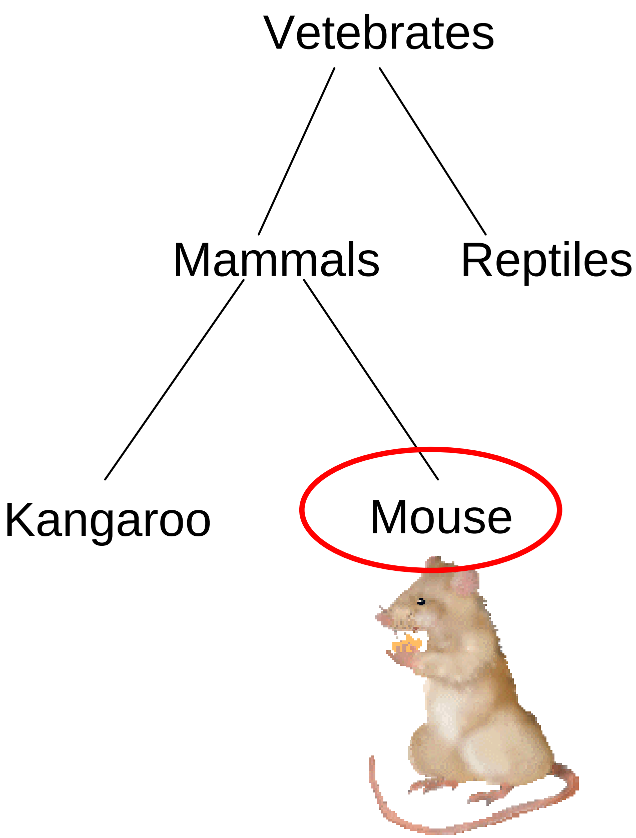


Ontology B:

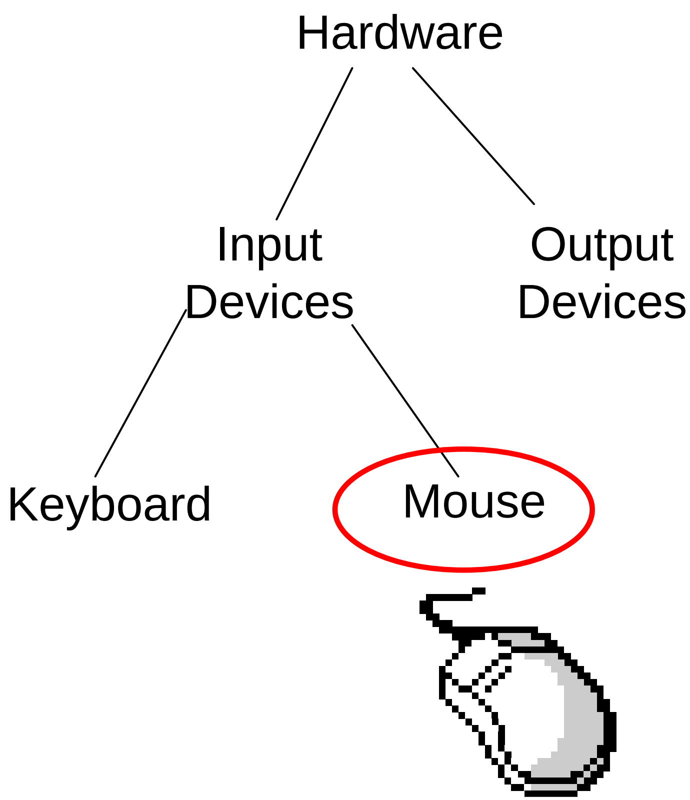


Interoperability

Ontology A:

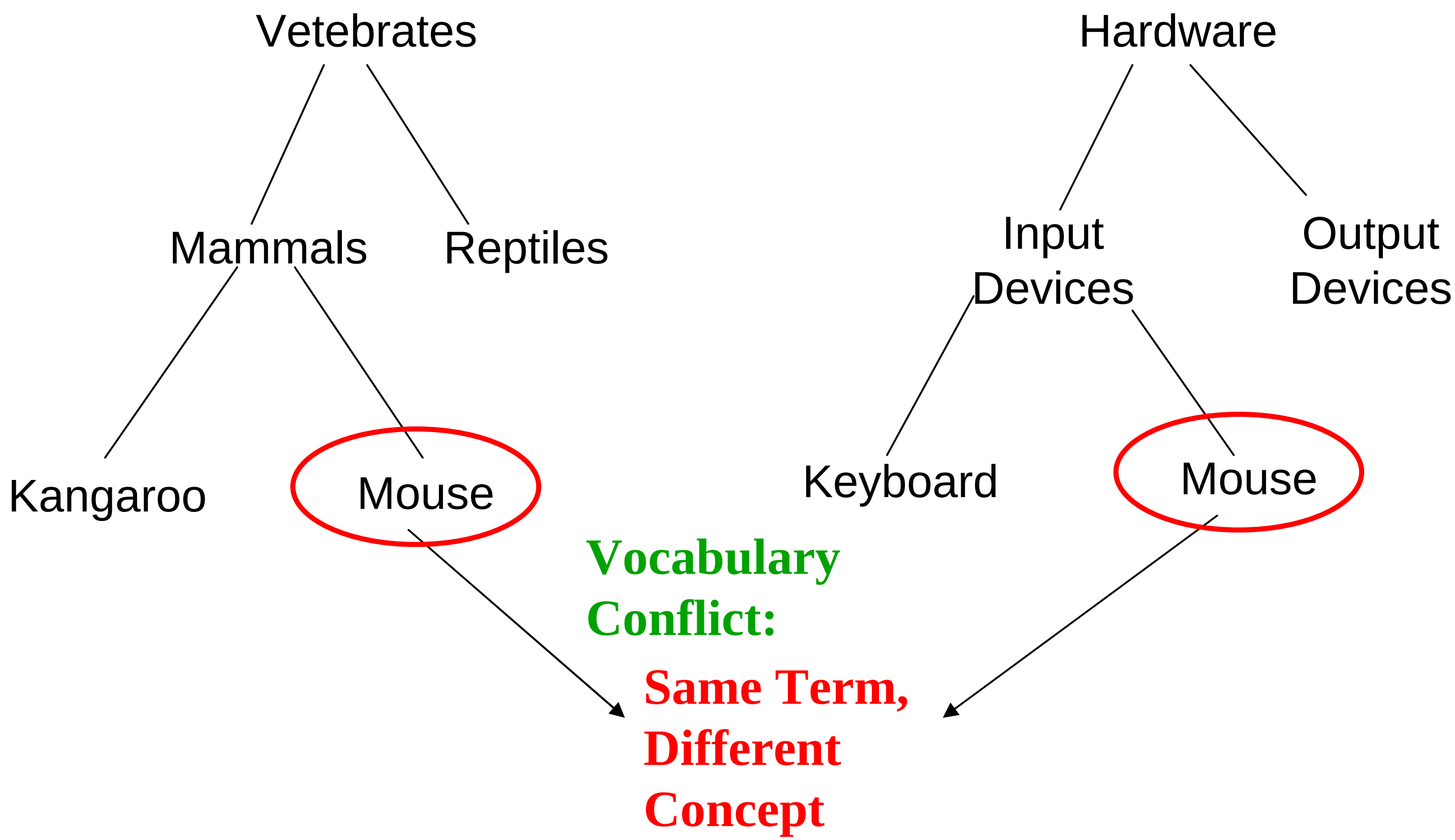


Ontology B:



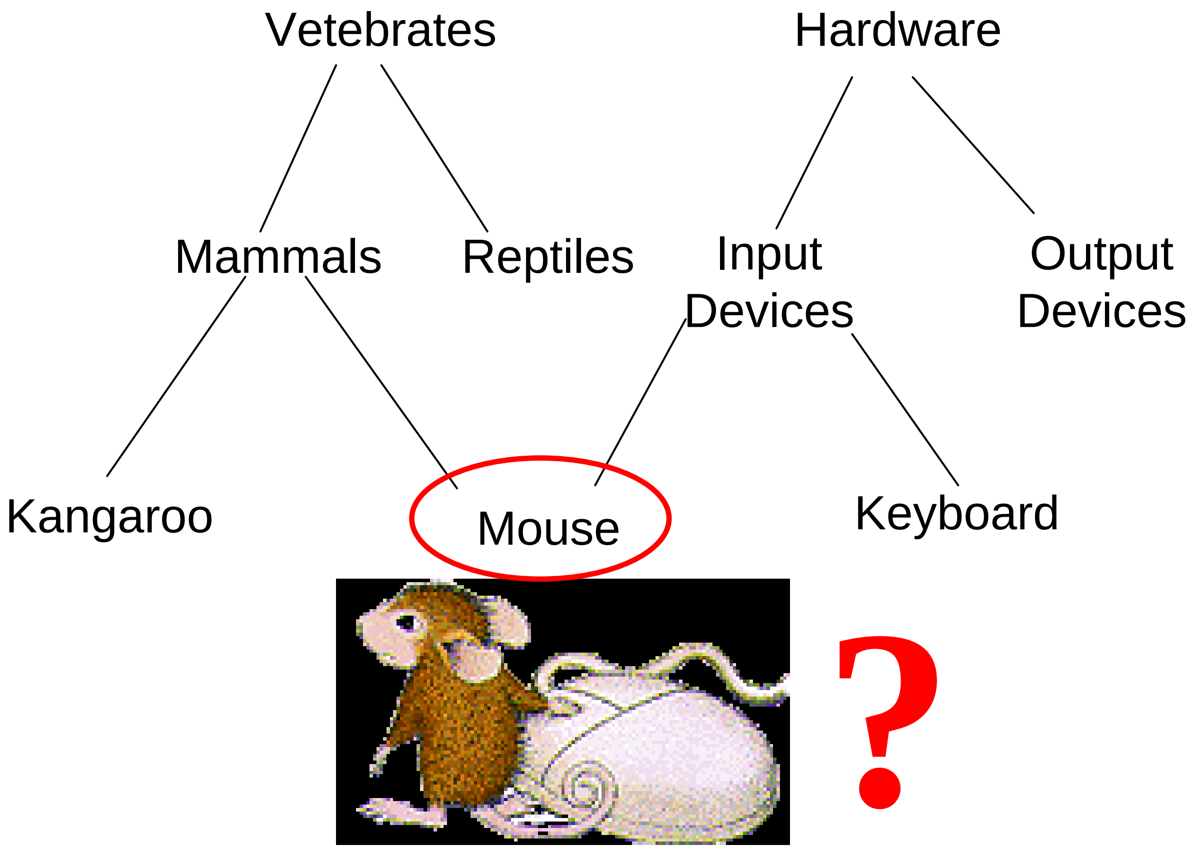
Interoperability

Ontology A: **Ontology B:**



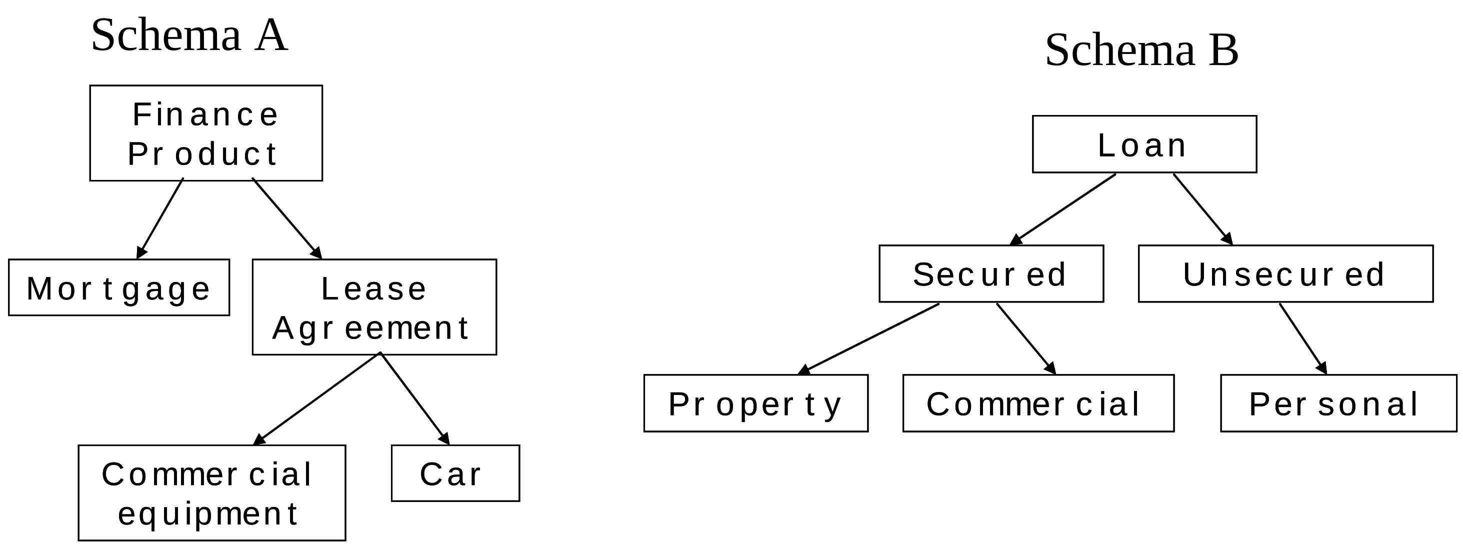
Interoperability

Merged Ontology:

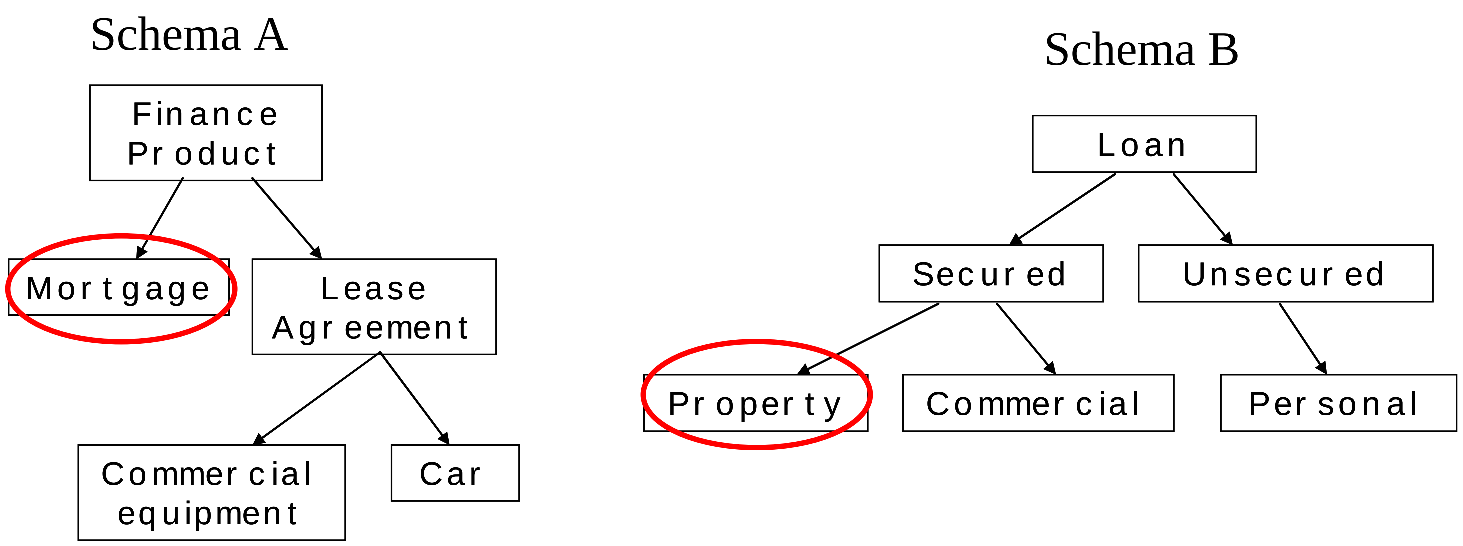


Interoperability

Vocabulary Incompatibility:



Interoperability

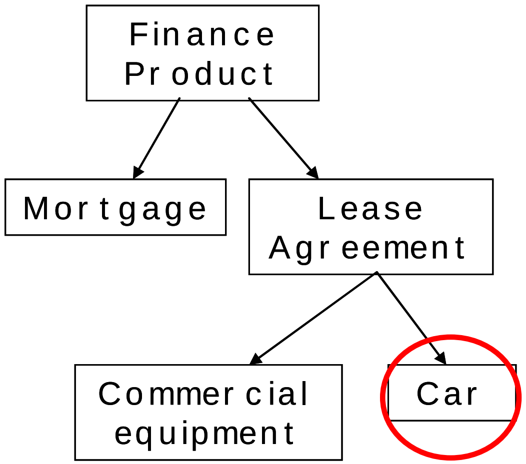


Terminology Clash
(different term, same concept)

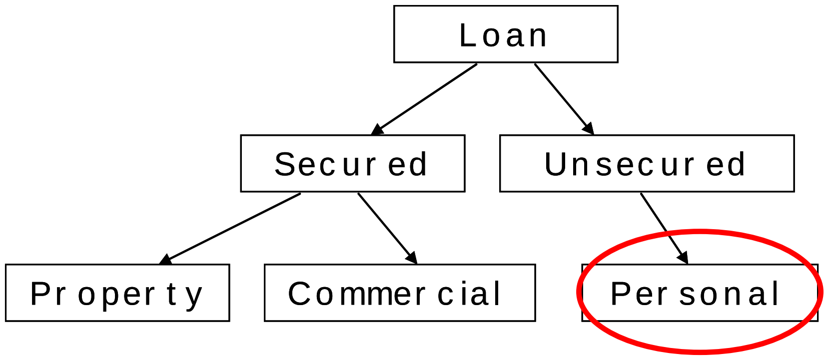
Interoperability

Conceptual Incompatibility:

Schema A



Schema B



Both can provide finance to obtain a car, but the concepts are not directly mappable

Conceptualisation Clash

- ***Leasing vs Borrowing?***
- ***Secured vs Unsecured?***
- ***Commercial vs Personal?***

Semantic Technologies

There are several strands to consider:

1. Semantic Structures
eg OWL Ontologies and Topic Maps

These are content-neutral ways of expressing relationships between data items

2. Semantic Content

What we actually want to talk about and what we want to say about it

3. Semantic Services

Which use the information to perform tasks and achieve goals

4. Logic and Reasoning

Fundamental basis to enable knowledge representation and reasoning

Semantic Technologies

1. Semantic Structures

- **W3C Recommendations :**
  - **RDF and RDF Schema**
  - **SKOS**
  - **OWL Web Ontology Language**
    - **OWL Lite**
    - **OWL DL**
    - **OWL Full**
    - **Extensions?**
  - **Rule Language to be adopted (SWRL is one candidate)**

- **ISO Standards – Topic Maps : (SC34 WG3)**
  - **Developed independently of W3C Semantic Web standards**
  - **Similar to OWL, but different**
  - **Do not have a formal semantics**

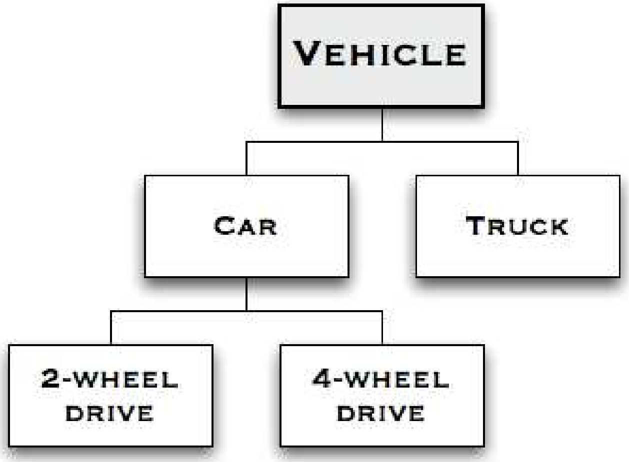
- Relational Databases, UML Models and so on can also be considered to be semantic structures – some have an explicitly specified semantics whilst others do not

What are Ontologies for?

- **Ontologies are explicit concepual models, which structure information about a domain**
  - Basic concepts or terms are chosen to suit the domain
  - Relationship between these concepts are then explicitly specified
  - Types of relationships specified in Ontologies typically include:
    - Class/subclass relations
    - Class/instance relations
    - Logical restrictions

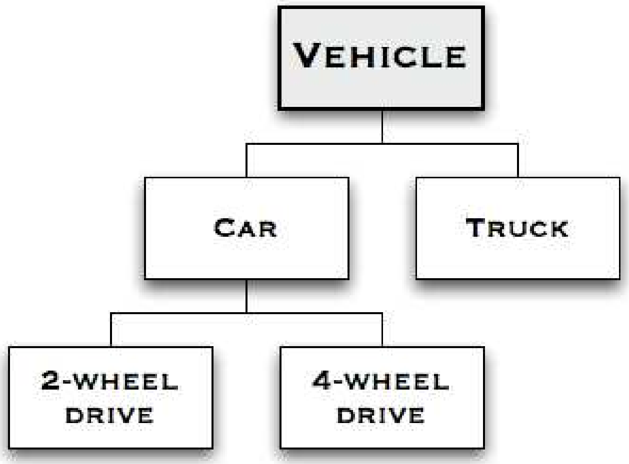
- **Example: Vehicle Ontology:**

  This very simple ontology designed for a vehicle domain, contains this information:
  - Cars and Trucks are Vehicles
  - 4-wheel drives and 2-wheel drives are cars



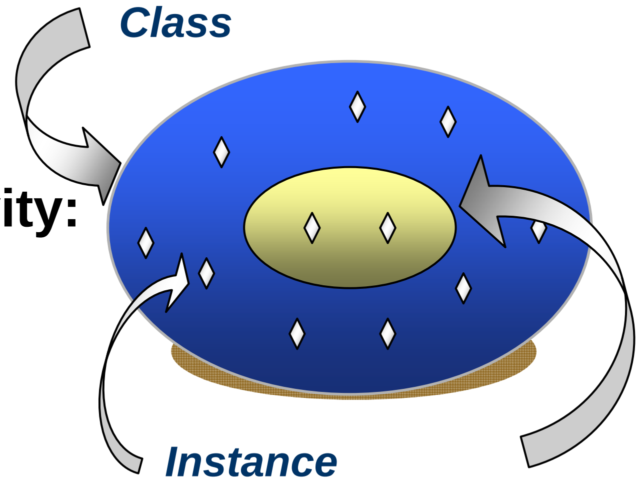
Benefits of using Ontologies

- Information becomes **interoperable**, by mapping it to a common ontology
- **Implicit** information becomes available
  - Inheritance: A 2-wheel drive is a car, and a car is a vehicle, so a 2-wheel drive is a vehicle
- Facilitates **automation**
- Enables **smart search**:
  - Finding all blue **cars** also includes blue 4-wheel drives, since a 4-wheel drive is recognised as a kind of car
- Can be used in **all** information technology channels – not only for public info on the Web, but for internal company information, B2B information management and more



OWL Ontologies: Overview

- Based on two key Ontological Relations:
  - SuperClass-SubClass
  - Class-Instance



- Two Layers Built on XML for extra expressivity:
  - RDF & RDFS (Resource Description Framework + RDF Schema)
  - OWL (Web Ontology Language)
- Constraints are Built into the Language:
  - Domain, Range, Cardinality, Transitivity, and more
  - OWL has more ability to express these than RDF
  - Three flavours of OWL provide increasing levels of expressivity
- OWL-Lite and OWL-DL have a Description Logic Semantics
  - Can do automated reasoning to determine things like:
    - Is class A a subclass of class B?

Different Emphases:

Topic Maps (ISO):

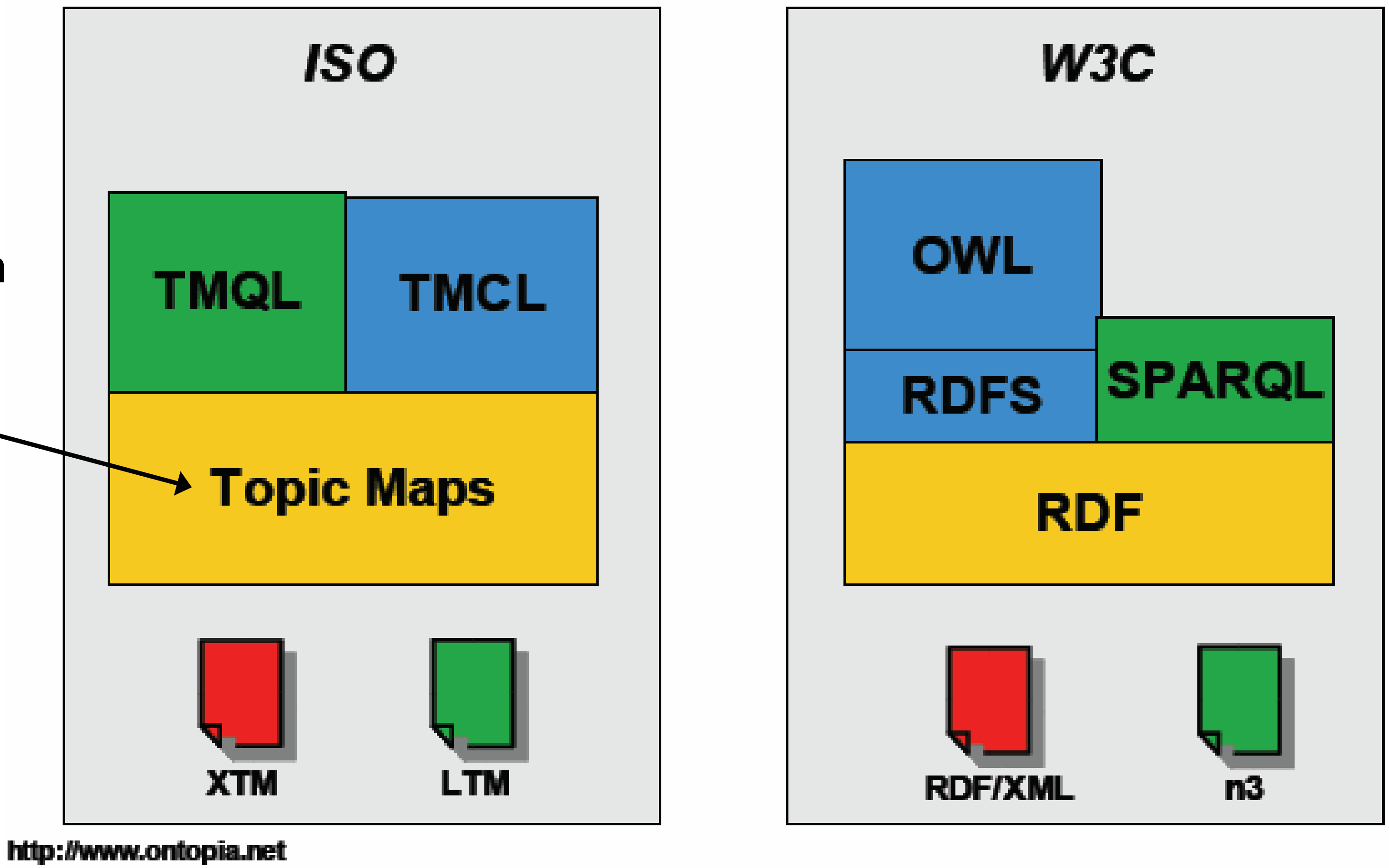
- Help humans find online information
  - Provide a way for humans to navigate information, by providing a meta-map of information resources
- Emphasis is on Human Understanding :
  - Freedom to capture any associations between subjects
  - Formal semantics are less important
- Grounded to Published Subject Indicators (URIs)

Semantic Web (W3C):

- Give machines access to semantic information
  - For meaning-based processing
  - supports new technologies for search and navigation, web services, intelligent agents, etc
- Emphasis is on Automated Processing:
  - Formal semantics are very important, to enable…
  - Algorithms for reasoning

Functional Overlap?

Based On the Topic Map Data Model (TMDM)



Source Acknowledgement: "**RDF/Topic Maps Interoperability"**
XTech 2005 presentation by Pepper & Garshol, Ontopia

An Opportunity?

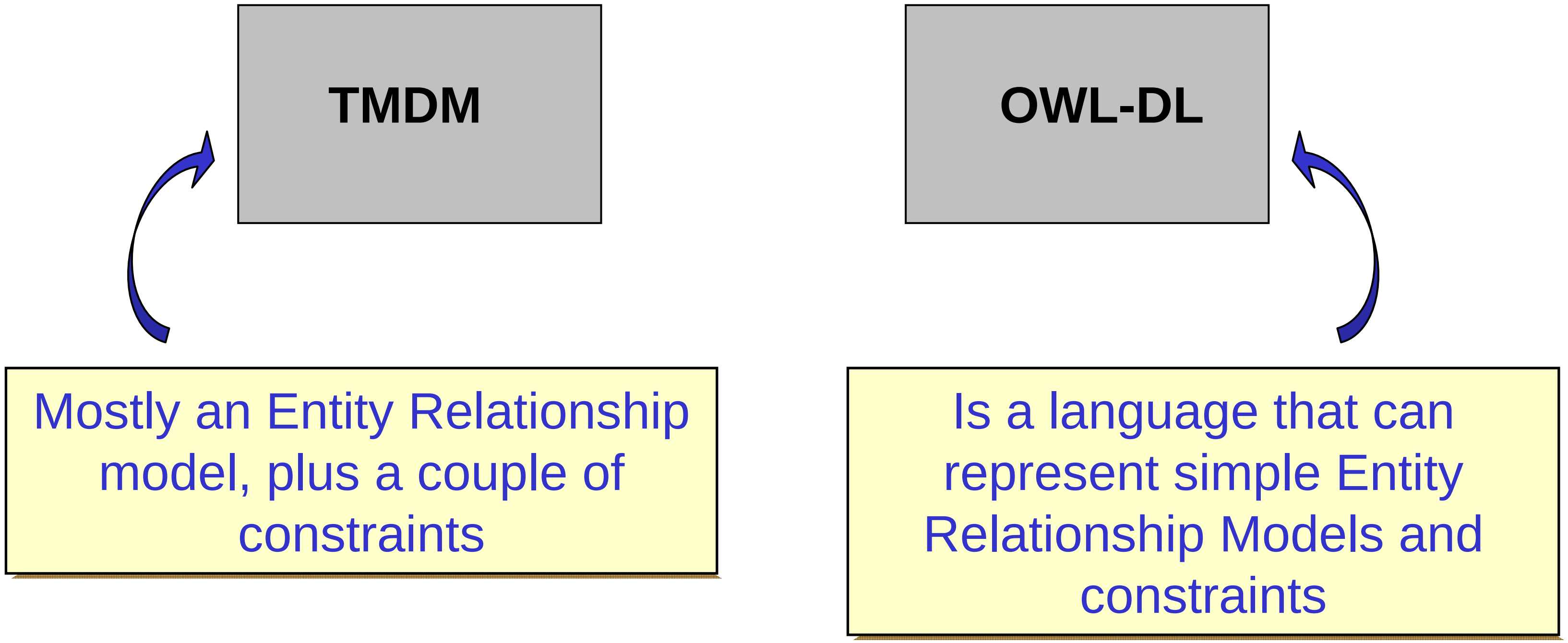
An Opportunity?

TMDM

OWL-DL

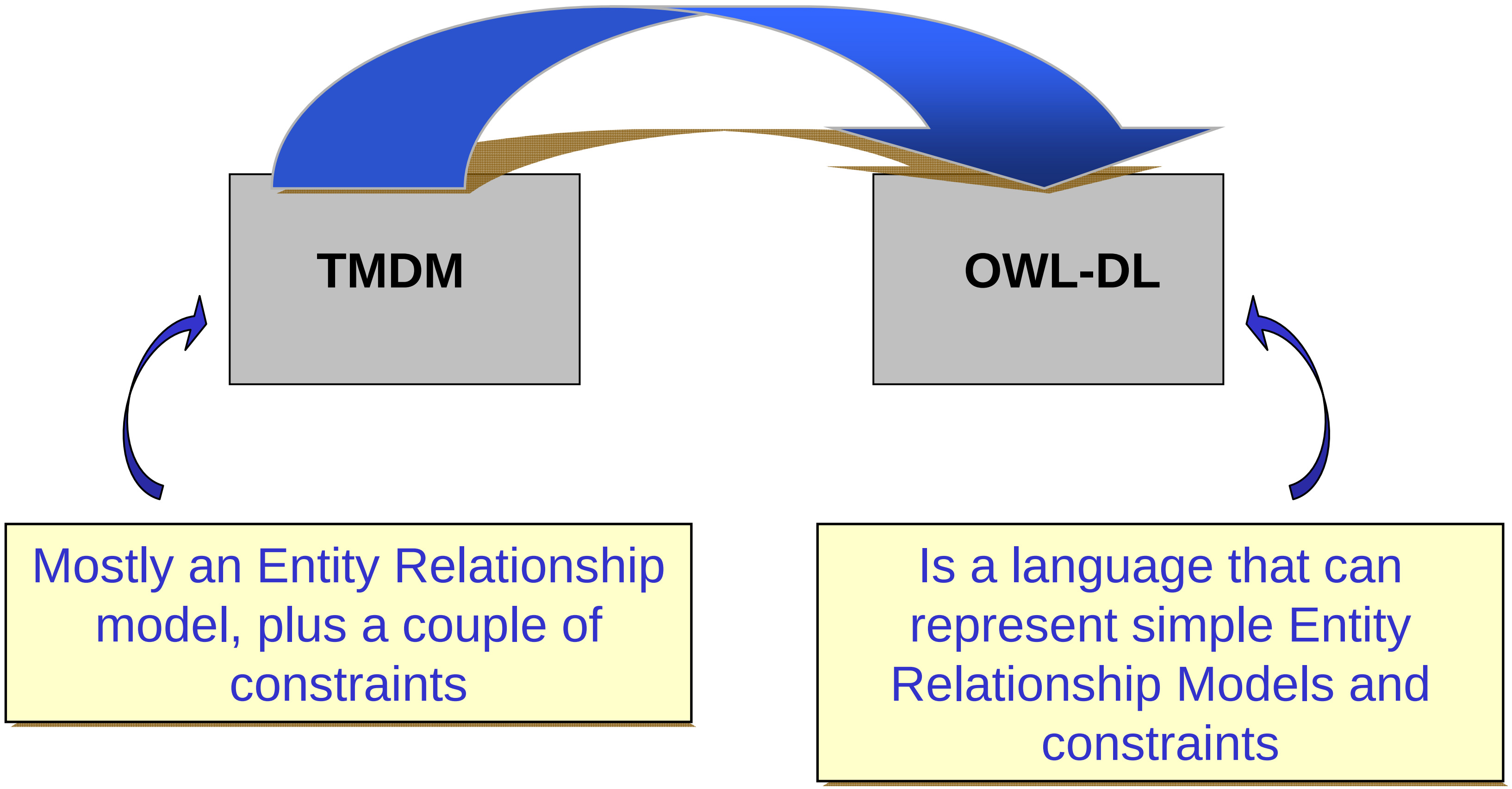
Mostly an Entity Relationship model, plus a couple of constraints

An Opportunity?



An Opportunity?

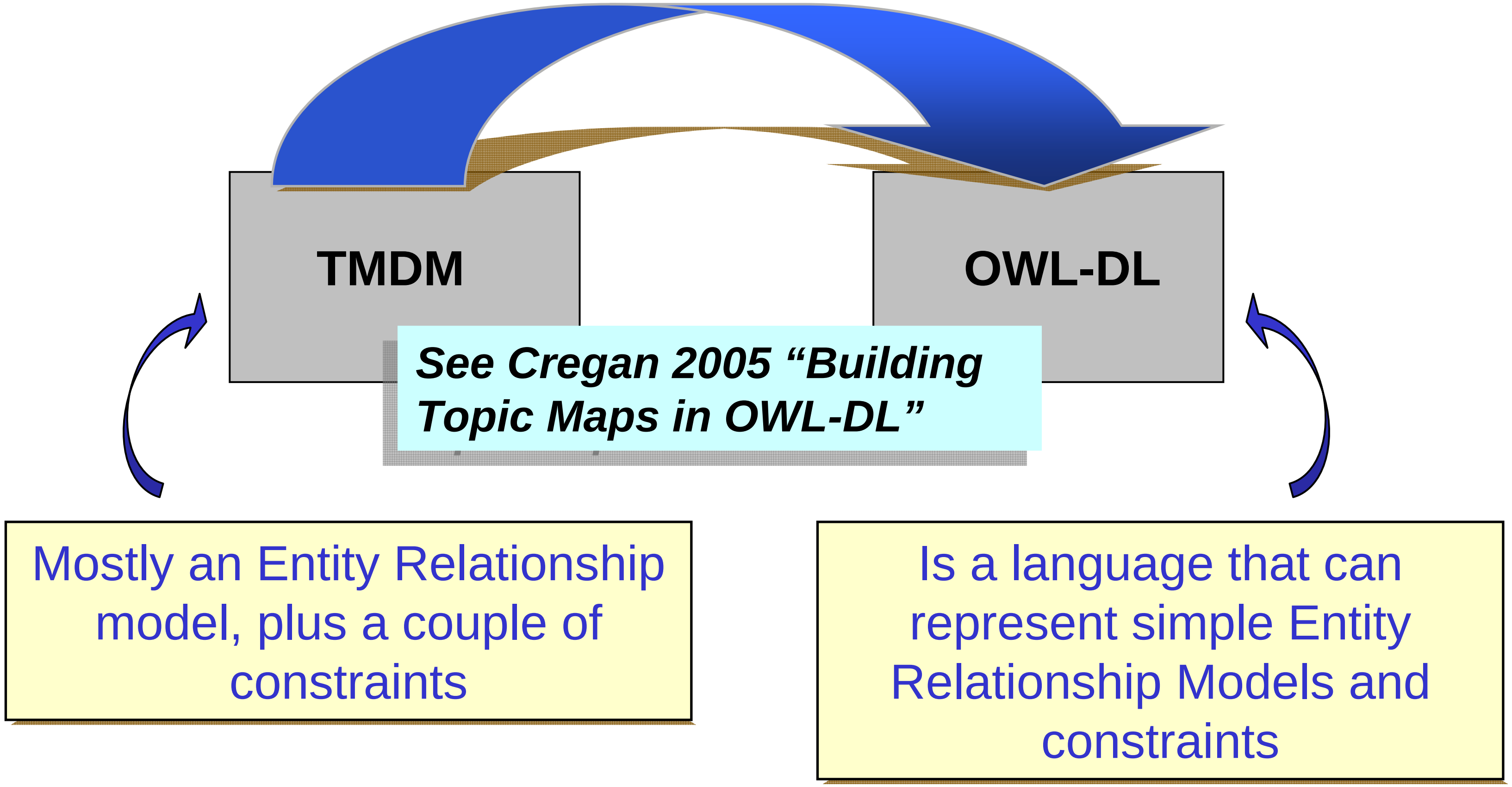
We can build the TMDM as an OWL-DL ontology and build Topic Maps by instantiating the base TMDM ontology



NATIONAL ICT AUSTRALIA

An Opportunity?

We can build the TMDM as an OWL-DL ontology and build Topic Maps by instantiating the base TMDM ontology



Interoperability

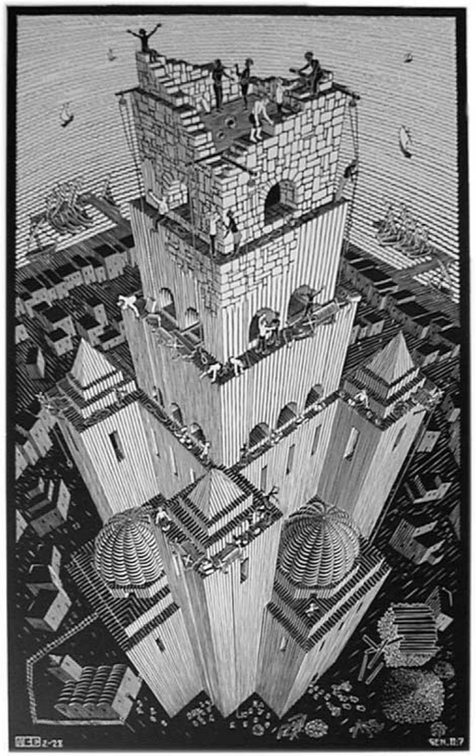
There are several strands to consider:

2. Semantic Content

Information becomes **interoperable**,
by mapping it to a common ontology

- Many people are building ontologies for many reasons
- Even if they use the same semantic structures, what if they don't use the same principles, and the same vocabulary?

Are we building
a Tower of Babel?



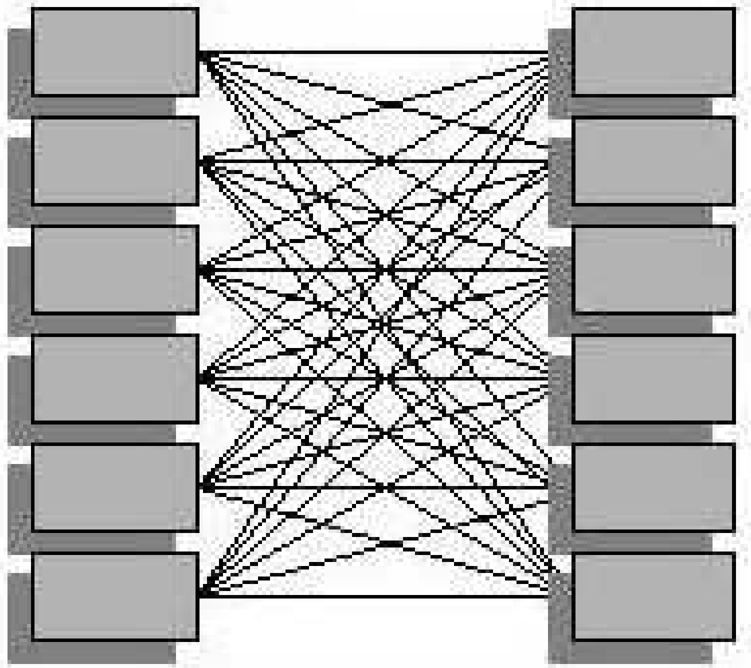
MC Escher's *Tower of Babel*

Semantic Content

Current Approach for Semantic Content Interoperability:

- Requires Point to Point mappings between every ontology / data store = n(n-1) mappings

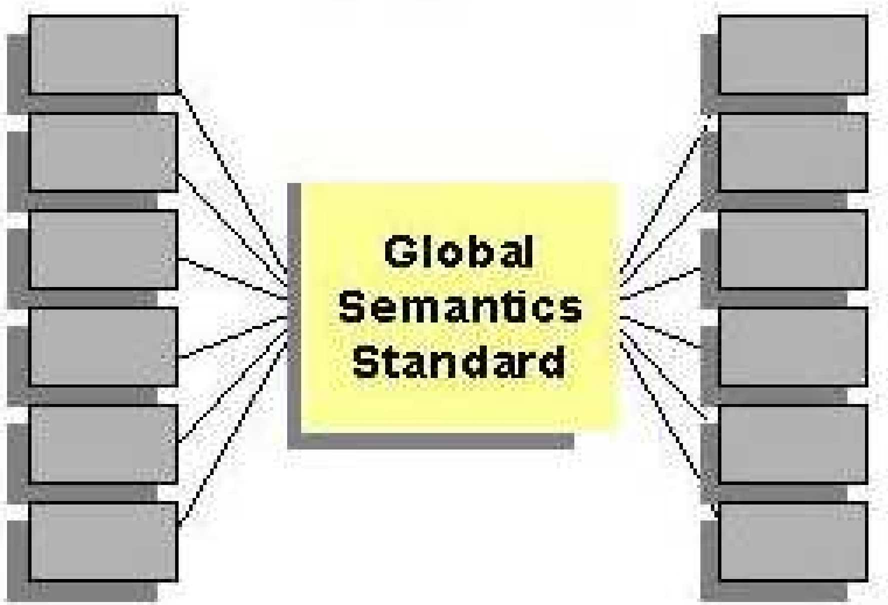
- Generally these mappings involve intensive human negotiation between designers to ascertain whether concepts align and how they align



Global Semantics Standard

Proposed Global Semantic Standards Approach:

- Uses a canonical approach to semantics: Semantic concepts are formally defined and registered
- Requires only n mappings
- Once mapped to the Semantic Standard, each ontology is automatically interoperable with every other aligned ontology



Semantic Content

Global Semantics Standard

- a conceptual naming convention framework with associated structured unique identifiers for indexing and aligning semantically equivalent concepts.
- Is a "Semantic Name Registry" that parallels the way that URIs give unique names to resources, but in the semantic space
- It provides a means to associate different data element names (i.e. vocabulary terms) that semantically refer to the same concept, to a standard data element concept name
- conformant with the relevant international standard on metadata registries and naming conventions, ISO/IEC 11179.
- Independent of technology or device

Semantic Content

Global Semantics Standard

- Proposals being developed in this space include:
  - Upper Ontologies:
    - SUMO (IEEE)
    - COSMOS (SiCoP)
  - UDEF (Universal Data Element Framework)

- Critical Factors for success are:
  - a rigorous Grounding methodology for specifying meaning precisely and resolving ambiguities and conflicts
  - Consensus from stakeholders, so they will adopt and use the recommended standards once they are developed

NATIONAL ICT AUSTRALIA

Semantic Content

Global Semantics Standard

- NICTA is working on
  - Grounding Methodology:
    - Definition Methods
    - Alignment of conceptual models
    - Conceptual Decomposition
    - Semantic Dimensions
    - Ontological Separation of Domains

Semantic Content

Global Semantics Standard

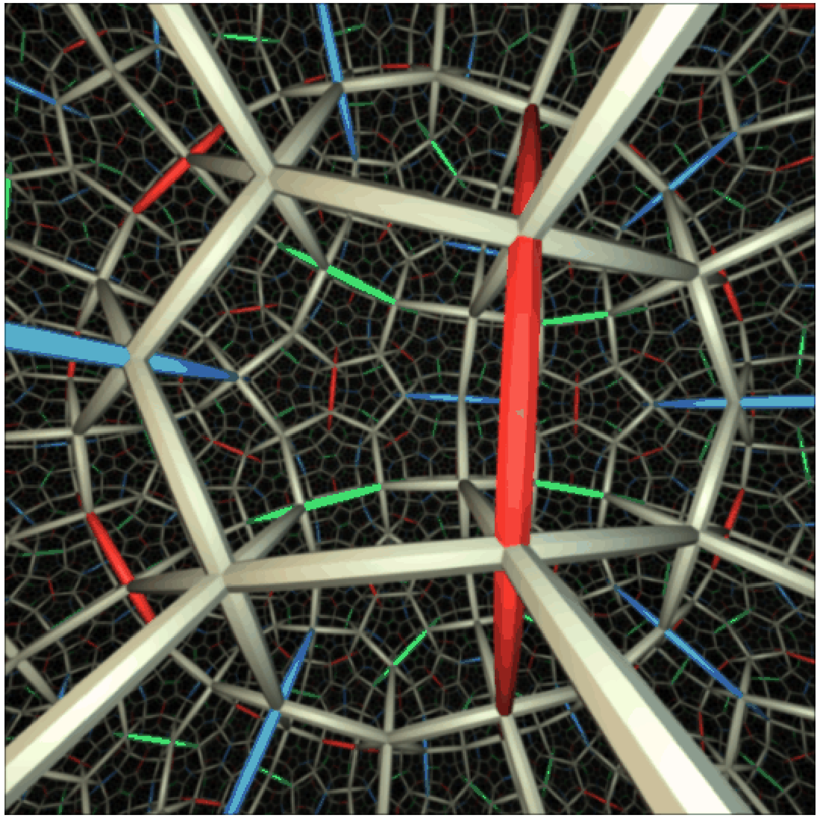
- NICTA is working on
  - Facilitating Consensus within the Australian space:
    - In liaison with Standards Australia, engaging and leading government, academic and commercial organisations across many subject domains to foster discussion and create groups to lead semantic standards formation
    - Online Semantic Techologies Forum just launched! www.semantictechnologiesforum.net
  - Whilst interfacing and harmonising with Global Standards Organisations:
    - ISO
    - W3C
    - OASIS

Semantic Services

There are several strands to consider:

3. Semantic Services

- Semantic Services also require semantic standards to ensure:
  - Appropriate Input
  - Meaningful, properly interpreted output
  - A Robust Basis for Services Composition



Logic and Reasoning

There are several strands to consider:

4. Logic and Reasoning

- What is the logical relationship between different kinds of data models eg UML and OWL?
  - Some may be isomorphic or subsets of others, being addressed by the Ontology Definition MetaModel
  - Different Semantic Structures have different logical properties
- Can the available structures provide all the expressivity needed to represent and reason with a conceptual model?
  - OWL is to be supplemented with a Rule language – SWRL is one candidate
  - Logical extensions to OWL are under consideration:
eg Role Composition eg uncle = parent + sibling (+ spouse)
  - Integrating logical reasoning with geometrical and numerical reasoning ?.....
- There's a classic tradeoff between expressivity and tractability of reasoning
  - When merging, you get the highest common denominator
eg merging an RDF structure with an OWL-Lite and OWL-Full ontology will give you an OWL-Full ontology
  - This makes reasoning intractable

NATIONAL ICT AUSTRALIA

Semantic Interoperability

NICTA invites you to participate:

- **Discuss and contribute to Semantic Interoperability at the online Semantic Technologies Forum:**

  www.semantictechnologiesforum.net/
  Sponsored by NICTA and Standards Australia

- **Come to the Australasian Ontology Workshop (AOW 2006) December 4-5, 2006, Hobart, Tasmania, Australia**

  http://www.comp.mq.edu.au/conferences/aow
  CALL FOR PAPERS has commenced - papers due 1 September 2006

  In Conjunction with the 19th Australian Joint Conference on Artificial Intelligence (AI2006) http://www.comp.utas.edu.au/ai06/