SEMANTIC WEB

INTRODUCTION

"The idea of surfing the net -- I don't know who called it that -- it's more like slogging through the net."

Anonymous

"Information on the internet is subject to the same rules and regulations as conversation at a bar."

George Lundberg

AIM

TO DESCRIBE THE ‘SEMANTIC WEB’

PREVIEW

- PHASE I - WHAT IS SEMANTIC WEB
- PHASE II - EVOLUTION & ARCHITECTURE
- PHASE III - SEMANTIC WEB TECHNOLOGIES & APPLICATION
- PHASE IV - CHALLENGES

WHAT IS SEMANTIC WEB

- The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation.

  Tim Berners-Lee, founder W3C

- The Semantic Web is a vision: the idea of having data on the web defined and linked in a way that it can be used by machines not just for display purposes, but for automation, integration and reuse of data across various applications.

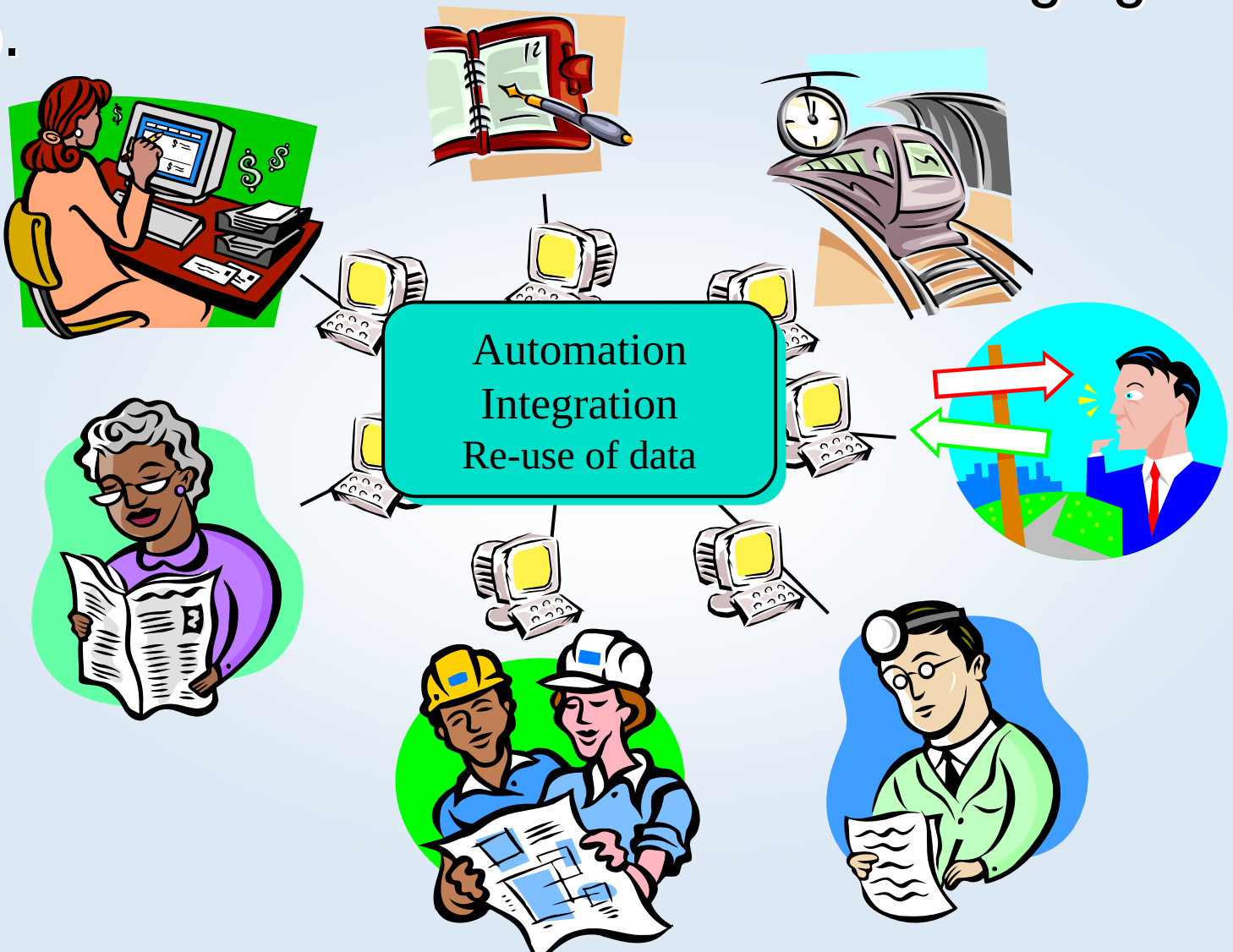
  http://www.w3.org/sw/

WHAT IS SEMANTIC WEB

- Semantic Web is an initiative of World Wide Web Consortium (W3C) .
- Semantic Web is a set of Languages and Tools for machine processing of information stored in the WWW.
- It is an efficient way of representing data on the World Wide Web, or as a globally linked Knowledge Base (KB).
- Semantic Web is about an efficient Knowledge Representation (KR) mechanism for Artificial Intelligence (AI).
- Semantic Web is about efficient and automated Reasoning Systems required for integration of distributed data .

SEMANTIC WEB

- Realization of various information exchanging via Web.



WHY SEMANTIC WEB?

- Most information on the Web is designed for human consumption, and even if it was derived from a database with well defined meanings for its columns, the structure of the data is not evident to a robot browsing the web.
- Leaving aside the Artificial Intelligence(AI) problem of training machines to behave like people, the Semantic Web approach instead develops languages for expressing information in a machine processable form.

SEMANTIC WEB VISION

The Semantic Web provides a common framework that allows **data** to be shared and reused across application, enterprise, and **community** boundaries.

Problem ,Business-to-consumer electronic commerce

- Manually retrieving the best offers from different online-shops is too time-consuming.
- Tools for shoping are available in the form of shop bots.
- For every online shop a wrapper is needed (information is extracted through text analysis).

Vision ,Business-to-consumer electronic commerce

- The user asks a autonomously acting piece of software for a certain product.
- The software retrieves all offers and compares them with the user‘s preferences.
- If needed the sofware also negotiates with the shop.

PHASE II

EVOLUTION AND ARCHITECTURE

"It shouldn't be too much of a surprise that the internet has evolved into a force strong enough to reflect our greatest hopes and fears . After all it was designed to withstand nuclear war."

Denise Caruso

TODAY´S WEB

- It is designed for human consumption.
- Information retrieval is mainly supported by keyword-based search engines.
- Some problems with information retrieval:
  - High recall, low precision.
  - Low or no recall.
  - Results are highly sensitive to vocabulary.
- Web content is not machine-proccessable
  - Tolstoy

EVOLUTION OF THE WEB

- Web made possible through established standards.
- TCP/IP for transporting bits down a wire.
- HTTP & HTML for transporting and rendering hyperlinked text.
- Applications able to exploit this common infrastructure.
- Result is the WWW as we know it.

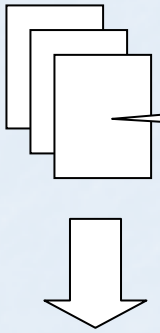
EVOLUTION OF THE WEB

- 1st generation web mostly handwritten HTML pages.
- 2nd generation (current) web often machine generated/active.
- Both intended for direct human processing/ interaction.
- In next generation web, resources should be more accessible to automated processes.
- To be achieved via semantic markup.
- Metadata annotations that describe content/function.

EVOLUTION OF WEB STANDARDS

Presentation-Oriented Markup

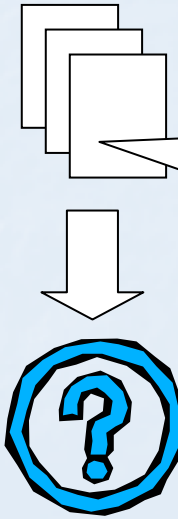
HTML



```
<tr><td><b>Charlotte's Web</b> -
E.B. White, Garth Williams.
<font color="Red">$6.99</font>
</td></tr>
```

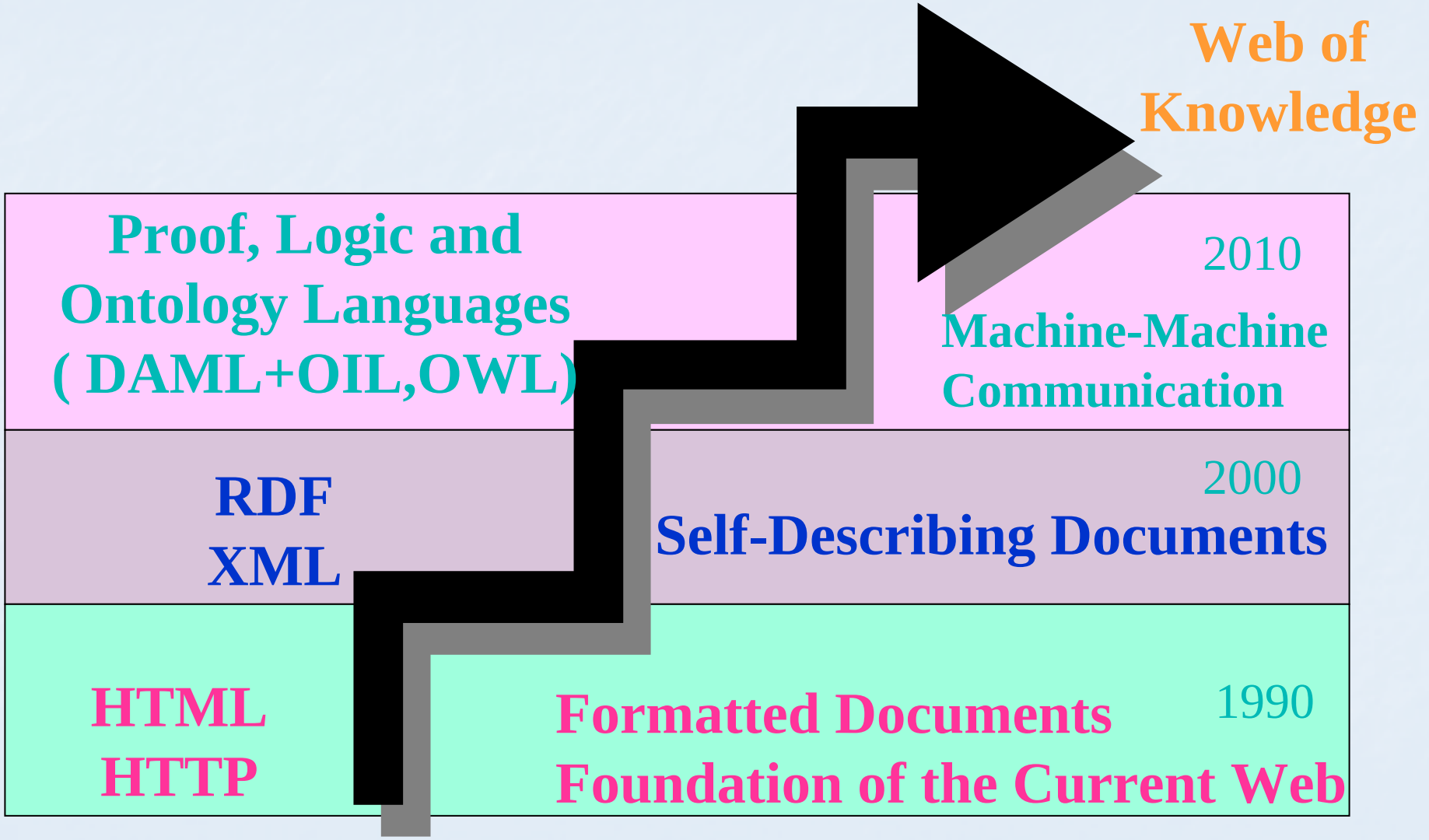
Content-Oriented Markup

XML

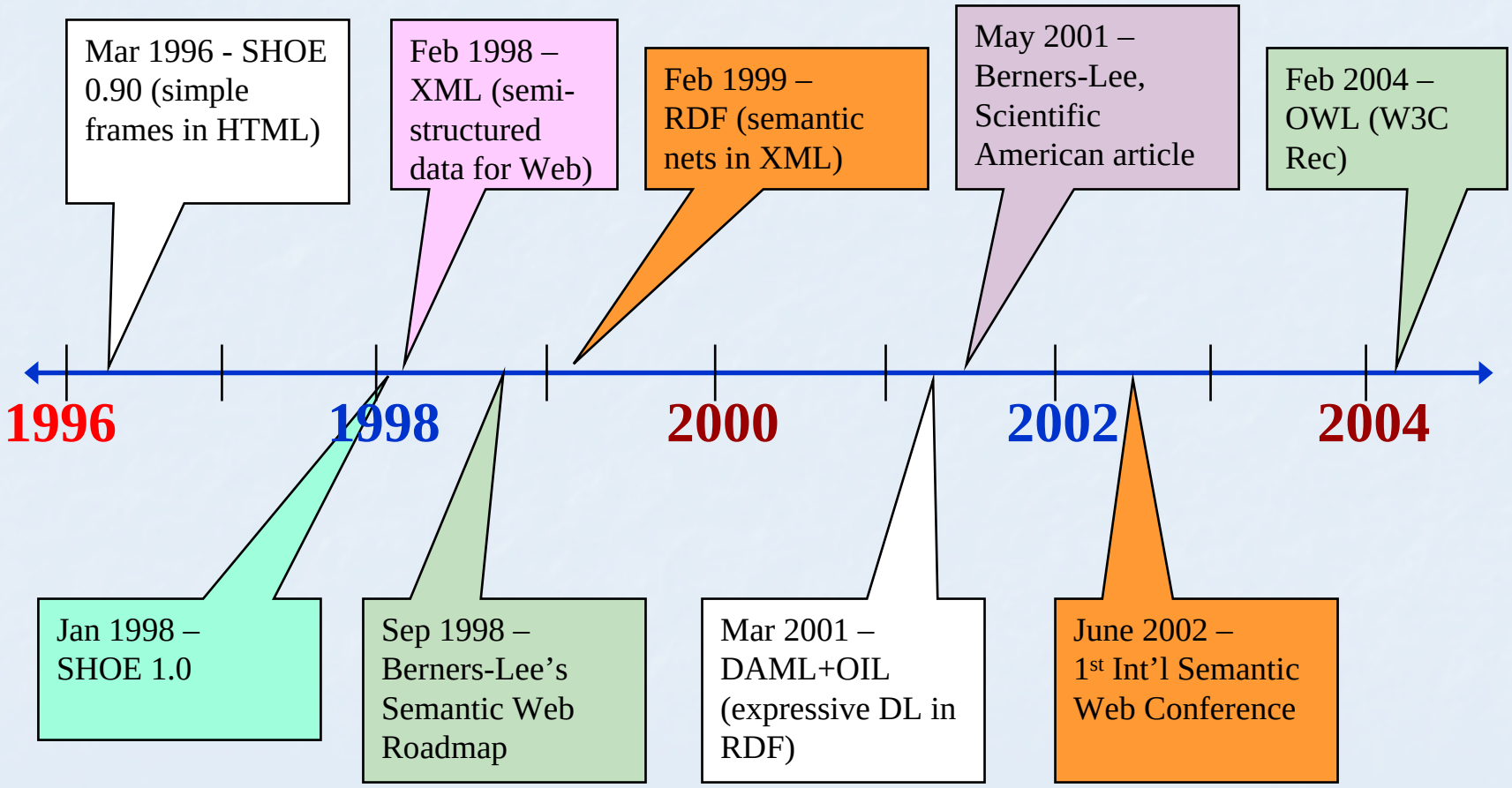


```
<book>
<title>Charlotte's Web</title>
<author>E.B. White</author>
<author>Garth Williams</author>
<price units="USD">6.99</price>
<subject>Children's Fiction</subject>
</book>
```

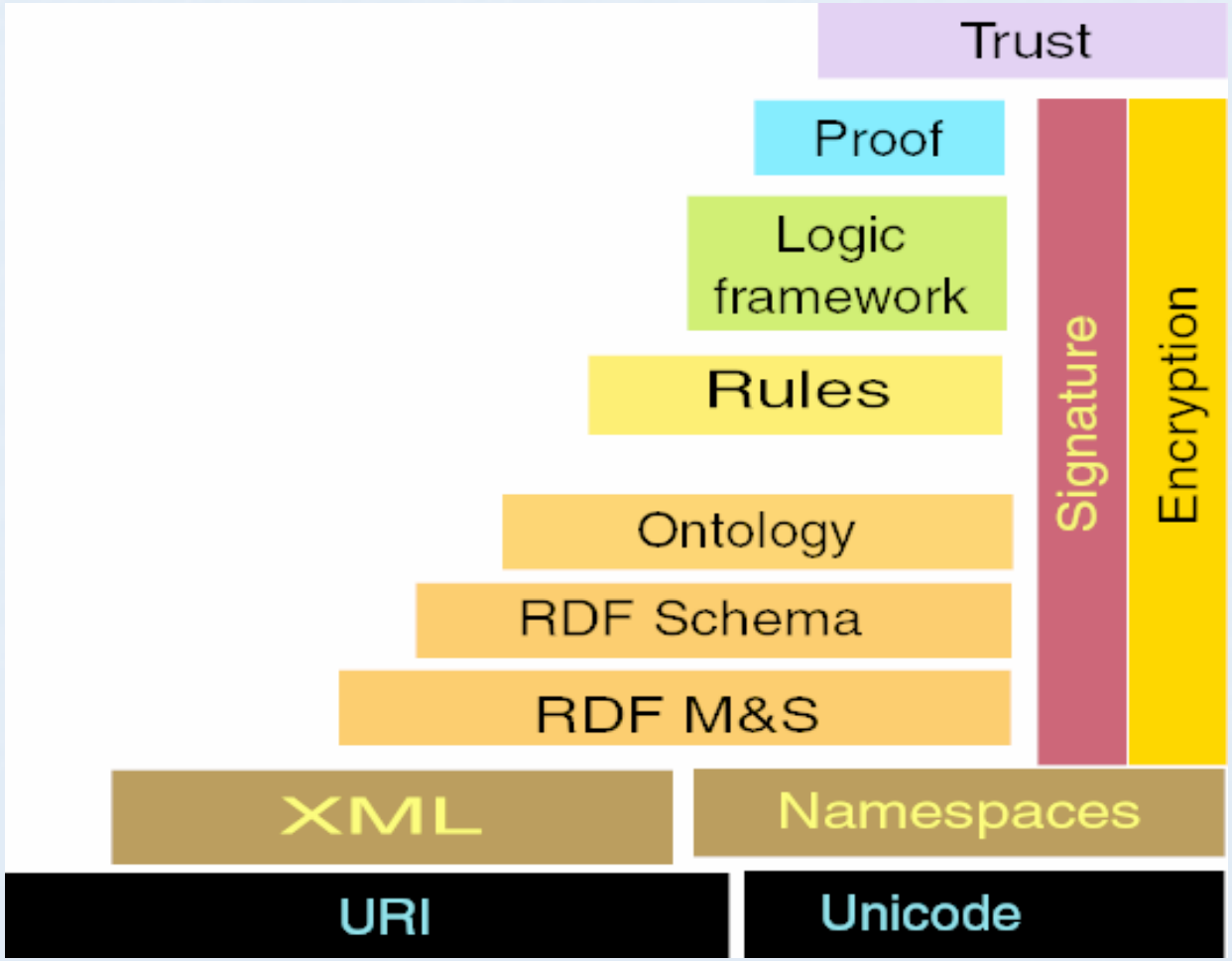
THE EVOLVING WEB



SEMANTIC WEB TIMELINE



A LAYER MODEL FOR SEMANTIC WEB

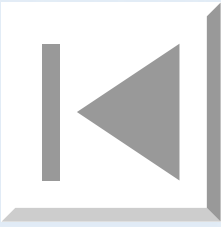


XML(eXTENSIBLE MARKUP LANGUAGE)

- A standard by W3C (1998).
- A subset of Standard Generalized Markup Language(SGML).
- Open Technology for Data Exchange.
- Data marked up using tags.
- Represent logical structures of data using Doc Type Definitions (DTD) and Schemas.
- XML doc rep as a hierarchical tree structure called a Document Object Model (DOM) Tree.

LIMITS OF XML

- XML provides syntactic interoperability. There is a need for semantic interoperability.
- How does one comp know what the other means by <price> ?
  - Does it include tax? shipping? surcharges?
- Critical in B2B e-commerce.
- If the computers of two companies are negotiating, they need to truly understand each other.
- The Semantic Web provides this added layer of interoperability through the use of shared ontologies.



RESOURCE DESCRIPTION FRAMEWORK (RDF)

- RDF is a language for representing resources in the World Wide Web.
- RDF is intended for situations in which this information needs to be processed by applications, rather than being only displayed to people.
- RDF is based on the idea of identifying things using Uniform Resource Identifiers (URIs).

RDF BASIC CONCEPTS

- Imagine trying to state that someone named John Smith created a particular Web page.

http://www.example.org/index.html has a **creator** whose value is **John Smith**

- The thing the statement describes (the web page`s URL).
- A specific property of the thing (eg creator).
- The concrete message the statement wants to give, in other words the value of the property (John Smith).

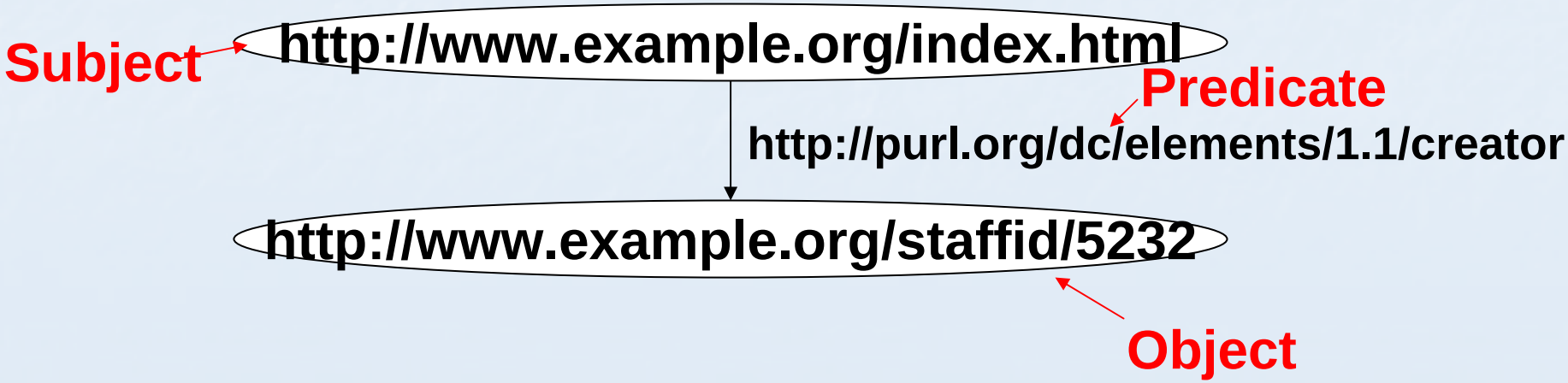
RDF BASIC CONCEPTS

- RDF Terminology
  - The part that identifies the thing the statement is about is called subject.
  - The part that identifies the property is called predicate.
  - The part that identifies the value of the property is called object.

RDF MODEL

- As mentioned:
  - RDF makes statements about resources
  - Each statement consists of a subject, a predicate and an object

http://www.example.org/index.html has a **creator** whose value is **John Smith**



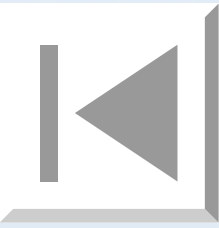
RDF BASIC CONCEPTS

- To make these statements machine-proccessable two things are needed:
  - A system of machine-processable identifiers (for subjects, predicates and objects) without any possibilty of confusion between similar looking identifiers .

    Uniform Resource Identifiers (URIs) allow to identify and uniquely name things - even if they have no network-accessible location.

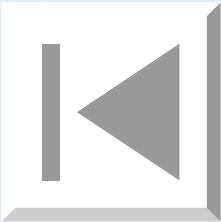
  - A machine-processable language for representing these statements and exchanging them between machines.

    RDF defines a XML markup language, named RDF/XML, which allows representation of RDF statements.



ONTOLOGIES

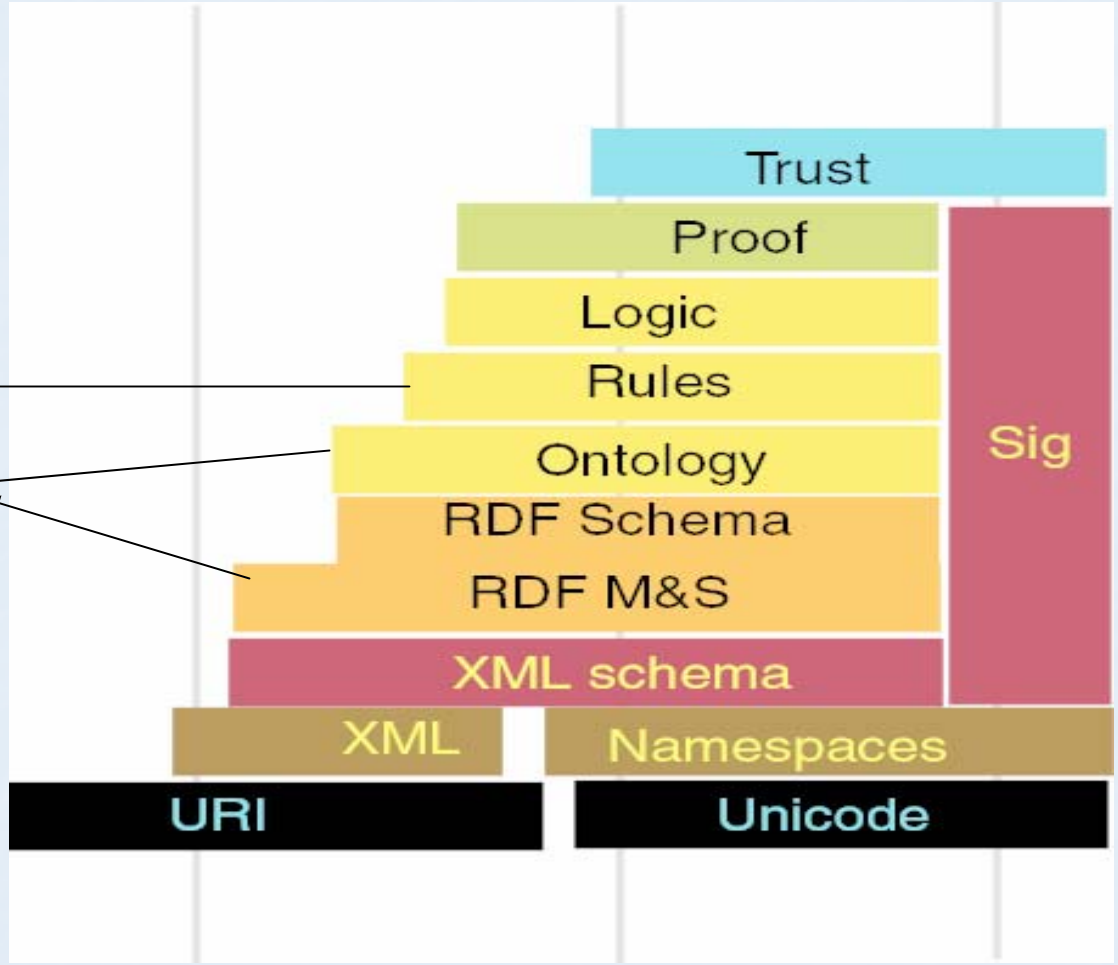
- In the context of web an ontology provides a shared understanding of a domain.
- Today, in computer science, an ontology is typically a hierarchical collection of classes, permissible relationships amongst those classes, and inference rules.
- For a computer program to reason, it must have a conceptual understanding of the world. This understanding is provided by an ontology.



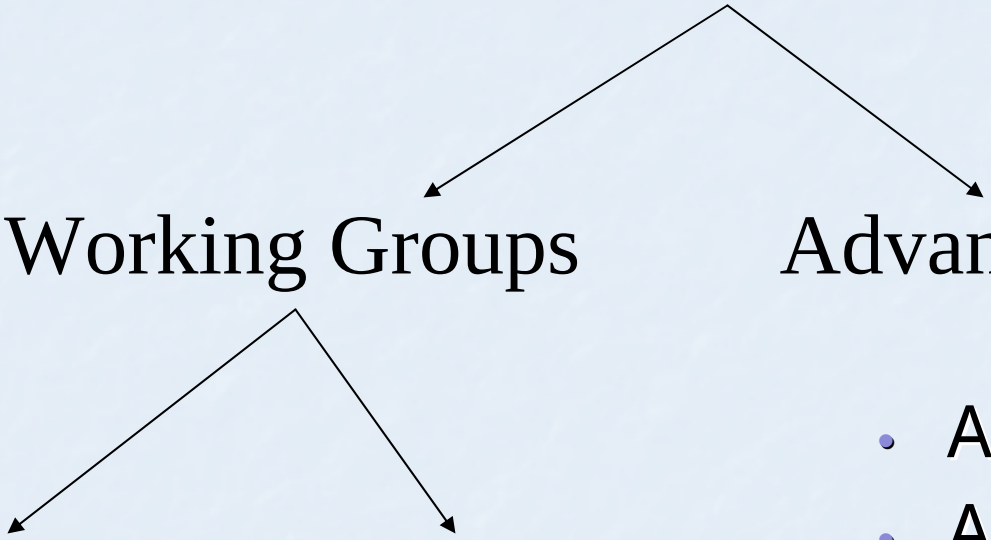
SEMANTIC WEB ACTIVITY

Research Phase

Standardization Phase



SEMANTIC WEB ACTIVITY



Working Groups

RDF Core Working Group

Web Ontology Working Group (WOW-G)

Advanced Development

- Annotation
- Access control
- Calendaring
- Collaboration
- Logic
- Rules
- Workflows

PHASE III

SEMANTIC WEB TECHNOLOGIES

"Technology is supposed to make life simpler!!!"

so they say

AGENTS

- Agents are pieces of software that work autonomously and proactively.
- A personal agent would recieive some task and preferences from the user, communicate with other agents, compare information and select certain choices.

WEB SERVICES

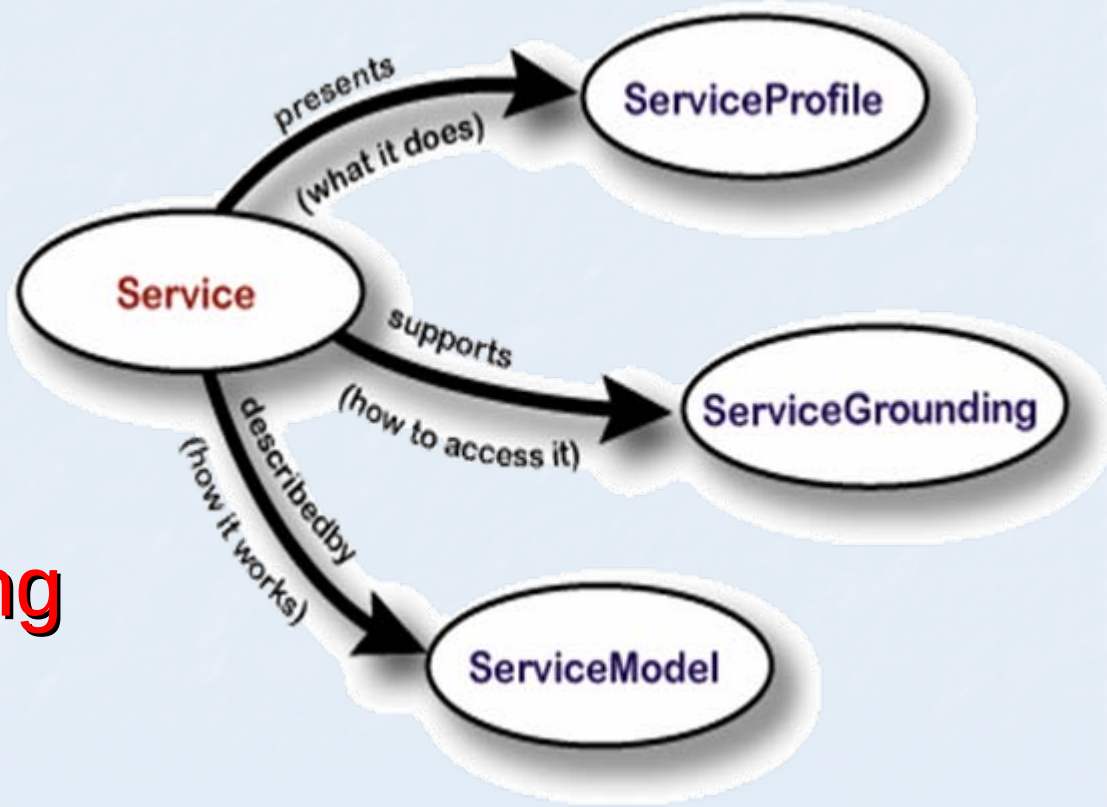
- Web services are a new breed of Web applications. They are self-contained, self-describing, modular applications that can be published, located, and invoked across the Web.
- Collection of protocols and standards for exchanging data between various applications
- Web services perform functions, which can be anything from simple requests to complicated business processes.

WEB SERVICES

- Once a Web service is deployed, other applications (and other Web Services) can discover and invoke the deployed service.
- Computer-interpretable description of the service is needed.
- The Semantic Web should enable users to locate, select, employ, compose, and monitor Web-based services automatically.
- SOAP (Simple Object Access Protocol) is a message layout specification that defines a uniform way of passing XML-encoded data.

WEB SERVICES

- What does the service provide?
  Service Profile
- How is it used?
  Service Model
- How to interact with it?
  Service Grounding



WEB PROCESSES

- Web processes describe how Web Services are connected to create reliable and dependable business solutions.
- Web processes allow businesses to describe sophisticated processes that can both consume and provide Web Services.
- The role of Web Processes within the enterprise is to simplify the integration of business and application processes across technological and corporate domains.

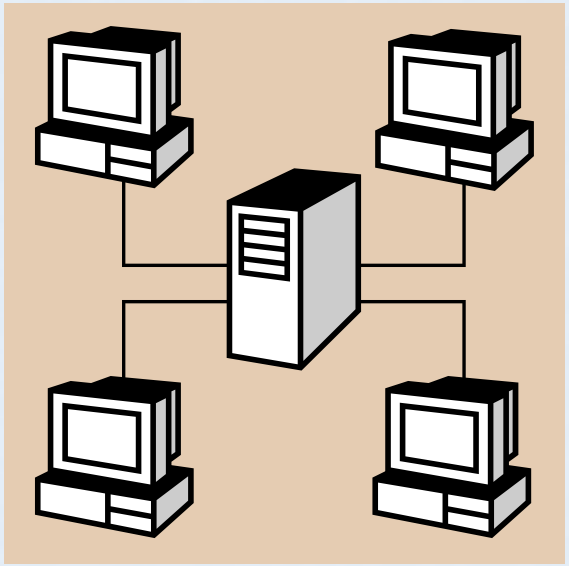
WEB PROCESSES

- Web Processes facilitate the interaction of organizations with markets, competitors, suppliers, customers etc supporting enterprise-level and core business activities.
  - Encompass the ideas of both intra and inter organizational workflow.
  - Created from the composition of Web Services.
- When all the tasks involved in a Web Process are semantically described, we may call such process as Semantic Web Processes.

PROCESSES DRIVING THE NETWORKED ECONOMY

B2B

E-Services



Workflows

Distributed Workflows

Web Processes

Enterprise | Inter-Enterprise | Global

HOW IT FITS TOGETHER?

- Onotologies can be used to represent knowledge, interpret the retrieved information and communicate with other agents.

  => RDF / RDF Schema / OWL
- Logic can be used for processing the retrieved information and for drawing conclusions.
- Agent / Web Service technologies allow the communication between different systems and the composition of complexer services from simpler ones.

  => OWL-S

SAMPLE WINE AGENT

- Wine Agent receives a meal description and retrieves a selection of matching wines available on the Web, using an ensemble of emerging Semantic Web standards and tools:
  - OWL for representing a domain ontology of foods, wines, their properties, and relationships between them.
  - Embedded Logical Reasoner (Java Theorem Prover) for deriving appropriate pairings.
  - DQL for querying a Knowledge Base consisting of the above.
  - Web Services for interfacing with vendors.
  - Utilities for conducting and caching the above transactions.

WINE ONTOLOGY

```
< <rdfs:Class rdf:ID="WINE">
    <rdfs:subClassOf rdf:resource="#POTABLE-LIQUID"/>
    <rdfs:subClassOf>
      <daml:Restriction>
        <daml:onProperty rdf:resource="#MAKER"/>
        <daml:cardinality>
          1
        </daml:cardinality>
      daml:Restriction>
        <daml:onProperty rdf:resource="#MAKER"/>
        <daml:toClass rdf:resource="#WINERY"/>
      </daml:Restriction>.....
```

(Wine has four properties – colour,sugar,body,flavour)

WINE ONTOLOGY

```
<rdfs:Class rdf:ID="MEAL-COURSE">
    <rdfs:subClassOf rdf:resource="#CONSUMABLE-
  THING"/>
    <rdfs:subClassOf>
      <daml:Restriction>
        <daml:onProperty rdf:resource="#FOOD"/>
        <daml:cardinality>
          1
        </daml:cardinality>
      </daml:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <daml:Restriction>
        <daml:onProperty rdf:resource="#FOOD"/>
        <daml:toClass rdf:resource="#EDIBLE-
  THING"/>
```

WINE ONTOLOGY

```
<rdfs:Class rdf:ID="PASTA-WITH-SPICY-RED-SAUCE">
    <rdfs:subClassOf rdf:resource="#PASTA-WITH-RED-
  SAUCE"/>
PASTA-WITH-RED-SAUCE"/>
  </rdfs:Class>
  <rdfs:Class rdf:ID="   </rdfs:Class>
  <rdfs:Class rdf:ID="PASTA-WITH-RED-SAUCE">
    <rdfs:subClassOf rdf:resource="#PASTA"/>
  </rdfs:Class>
  <rdfs:Class rdf:ID="PASTA-WITH-SPICY-RED-SAUCE-
  COURSE">
    <daml:intersectionOf rdf:parseType="daml:collection">
      <daml:Restriction>
        <daml:onProperty rdf:resource="#FOOD"/>
        <daml:hasClass rdf:resource="#PASTA-WITH-SPICY-
  RED-SAUCE"/>
      </daml:Restriction>
      <rdfs:Class rdf:about="#MEAL-COURSE"/>
```

REASONING SYSTEM

- Embedded Reasoner infers pairings within given framework to create a course

  Assert:

  (|http://www.w3.org/1999/02/22-rdf-syntax-ns#|::type
  |http://www.wine.com/wkb#|::|NEW-COURSE9|
  |http://www.wine.com/wkb#|::|PASTA-WITH-SPICY-RED-SAUCE-COURSE|)

  Assert:

  (|http://www.wine.com/wkb#|::|DRINK|
  |http://www.wine.com/wkb#|::|NEW-COURSE9|
  |http://www.wine.com/wkb#|::|WINE9|)

QUERYING KNOWLEDGE BASE FOR PROPERTIES

Query:
(|http://www.wine.com/wkb#|::|**COLOR**|
|http://www.wine.com/wkb#|::|WINE9|
?x)
---------->
X = |http://www.wine.com/wkb#|::|**RED**|
Query:
(|http://www.wine.com/wkb#|::|**SUGAR**|
|http://www.wine.com/wkb#|::|WINE9|
?x)
---------->
X = |http://www.wine.com/wkb#|::|**DRY**|

QUERYING KNOWLEDGE BASE FOR PROPERTIES

Query:
(|http://www.wine.com/wkb#|::|**BODY**|
|http://www.wine.com/wkb#|::|WINE9|
?x)
---------->
X = |http://www.wine.com/wkb#|::|**FULL**|
Query:
(|http://www.wine.com/wkb#|::|**FLAVOR**|
|http://www.wine.com/wkb#|::|WINE9|
?x)
---------->
X = |http://www.wine.com/wkb#|::|**STRONG**|

QUERYING KNOWLEDGE BASE FOR WINES

Query:

(AND (|http://www.w3.org/1999/02/22-rdf-syntax-ns#|::type ?x
|http://www.wine.com/wkb#|::|WINE|)
(|http://www.wine.com/wkb#|::|**COLOR**| ?x
|http://www.wine.com/wkb#|::|**RED**|)
(|http://www.wine.com/wkb#|::|**SUGAR**| ?x
|http://www.wine.com/wkb#|::|**DRY**|)
(|http://www.wine.com/wkb#|::|**BODY**| ?x
|http://www.wine.com/wkb#|::|**FULL**|)
(|http://www.wine.com/wkb#|::|**FLAVOR**| ?x
|http://www.wine.com/wkb#|::|**STRONG**|))--------
-->X = |http://www.wine.com/wkb#|::|CHATEAU-LAFITE-ROTHSCHILD-PAUILLAC|

WEB SERVICE

- Wine Agent also accesses information about pricing and availability of wines through www.wineinfo.com, via a rudimentary on-site search engine.
- www.wineinfo.com used as information providing Web Service.
- Markup enables automated invocation of the Web Service, facilitating interoperability, semantic translation of wine information, and enabling more sophisticated searching and filtering of wine information using the richer wine ontology provided by the Wine Agent.

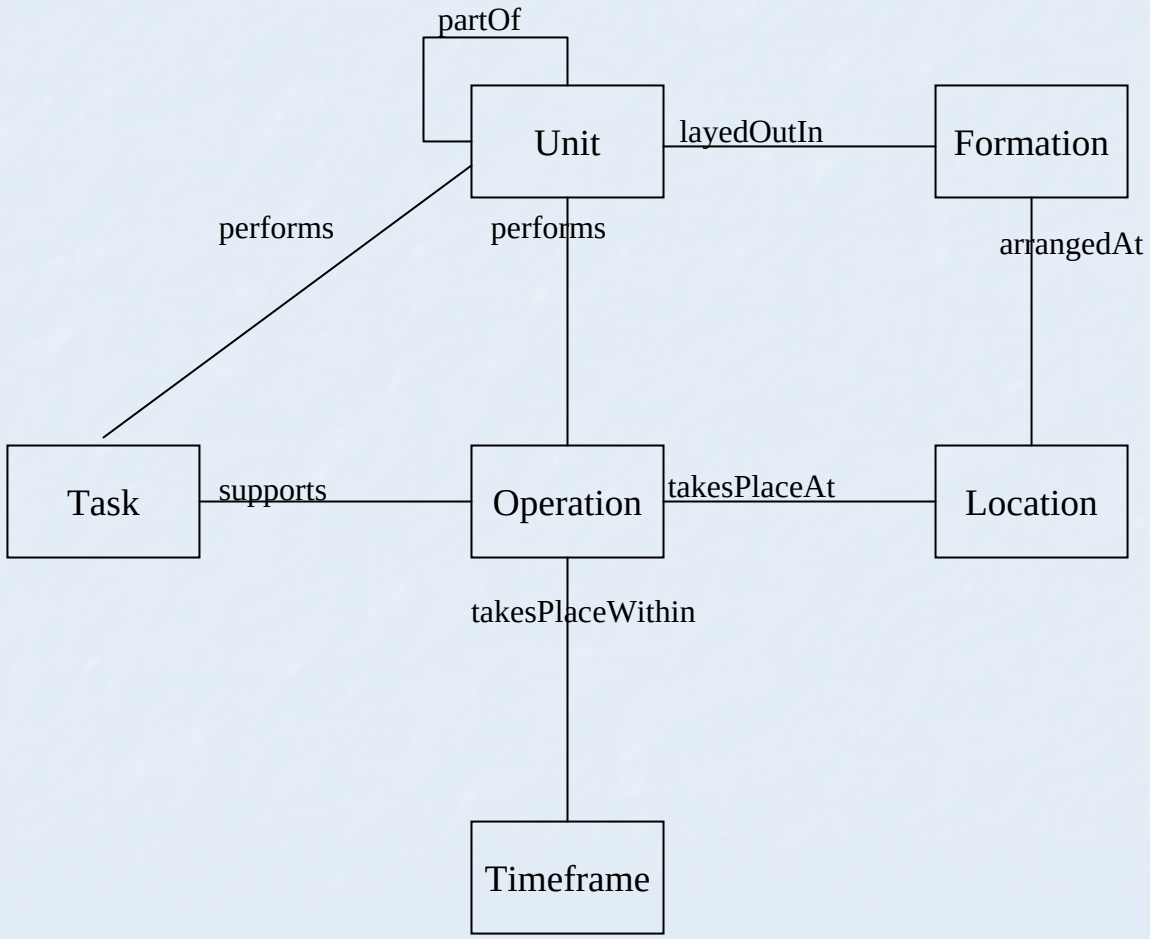
MEANS TO ACHIEVE THE VISION

- Explicit Ontologies
  - Needed to understand each others data (eg, joint notion about what a schedule is).
- Web Services
  - Required to actively interconnect systems (automatically make an appointment).
- Agents

BATTLE MANAGEMENT LANGUAGE

- Standard vocabulary used by simulation programmers or workstation controllers to describe planning and executing military functions currently performed by human controllers.
- Key factors supporting automated Course of Action (COA) analysis:
  - Unit identification (who)
  - Operation / Task (what)
  - Operation time (when)
  - Operation location (where)
  - Operation purpose / mission (why)

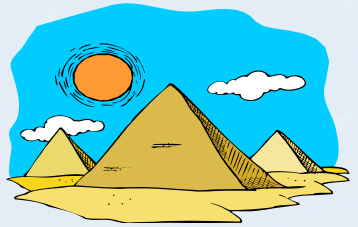
BML ORDER ONTOLOGY

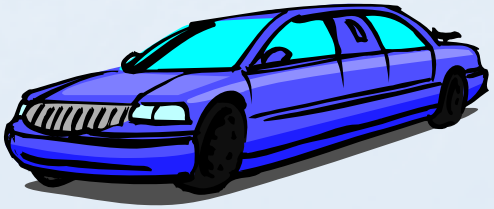


HUMINT REPORT REPRESENTATION

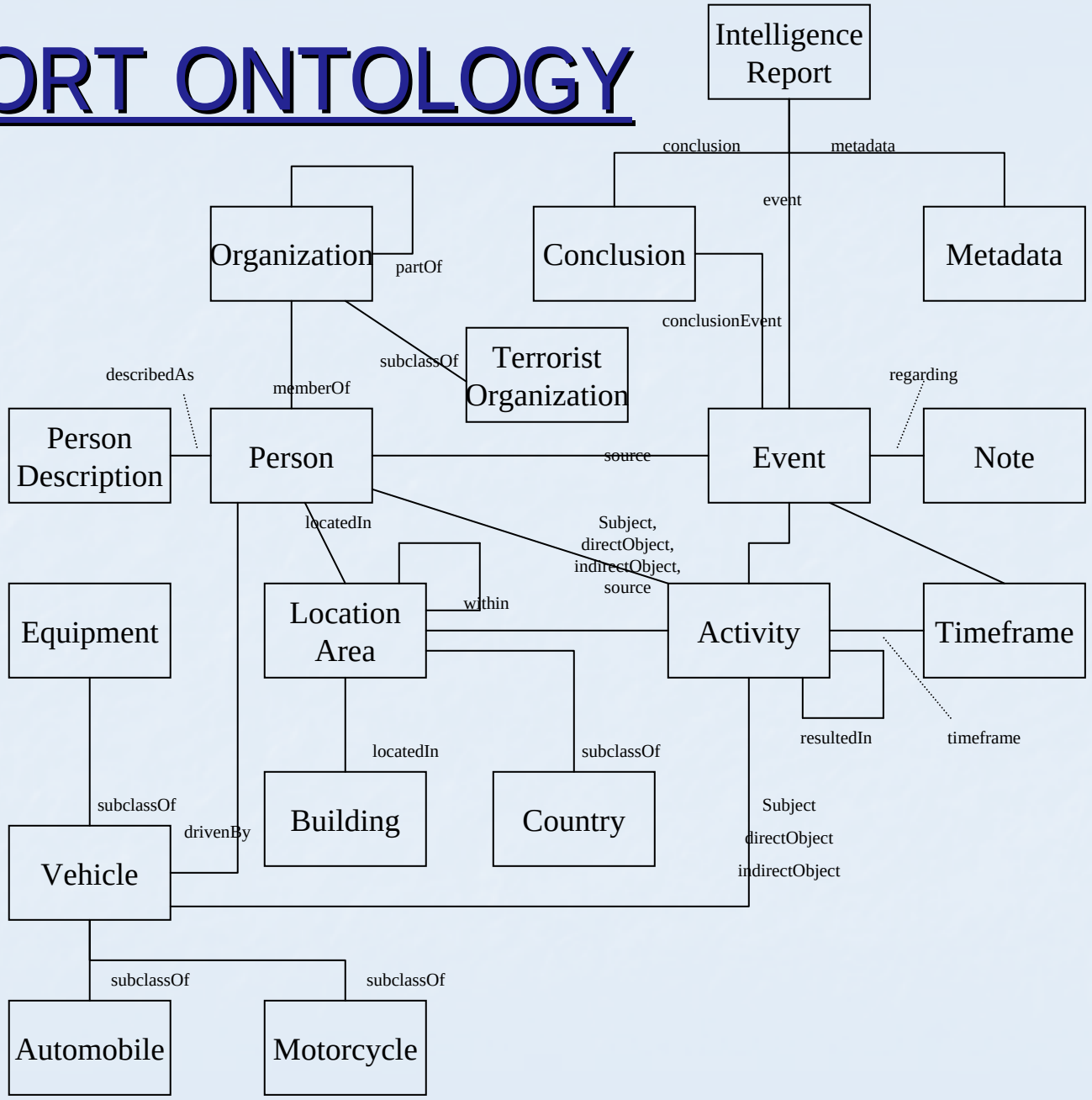
Classes identified include:

- Organization, Person (who)
- Activity, Event (what)
- Timeframe (when)
- Location Area (where)
- Conclusion (why)





INT REPORT ONTOLOGY



PHASE IV

CHALLENGES

" Dreams can often become challenges, but challenges are what we live for."

Travis White

TECHNICAL CHALLENGES

- Interoperability
  - Unreliable, ill-defined, evolving services.
  - Inaccurate, incomplete, heterogeneous data.
- Natural Language Processing(NLP), Data Mining...
  - make information explicit.
- Scalability
  - Subsecond performance.

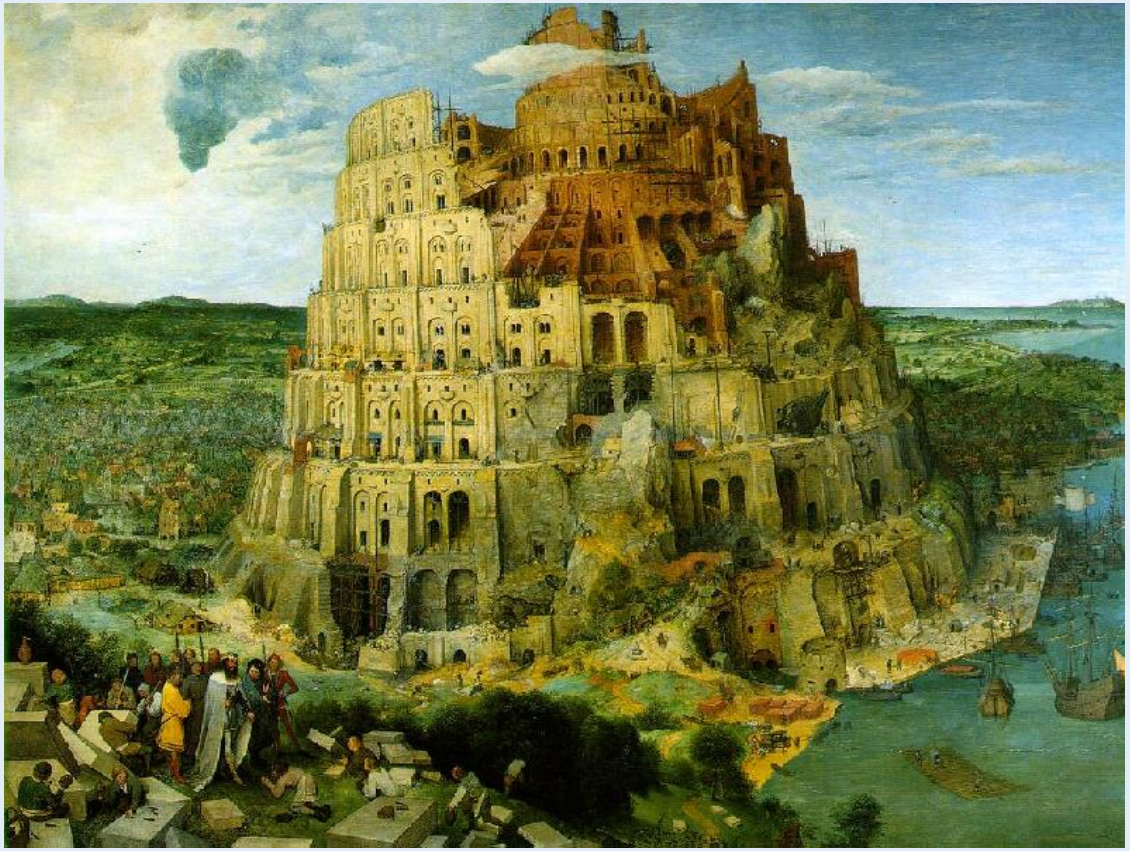
SOCIAL CHALLENGES

- Standardization is hard.
  - Dublin Core
- Bogus or inaccurate metadata.
  - Physician rating, profile.
- Competition and commoditization.
- Economic incentive.
- Complexity: developers and users.

MANY PREVIOUSLY UNKNOWN COMMUNICATION PARTNERS

HETEROGENOUS DATA

- Too many data formats/languages.



CONCLUSION

- Today‘s web and its problems.
- A vision for Semantic Web applications.

- The basic technologies exist *but* standards have to mature *and*
- More practical problems have to be solved (tool support, ontology matching, ...).