


**Semantic Web Enabled  
Software Engineering**

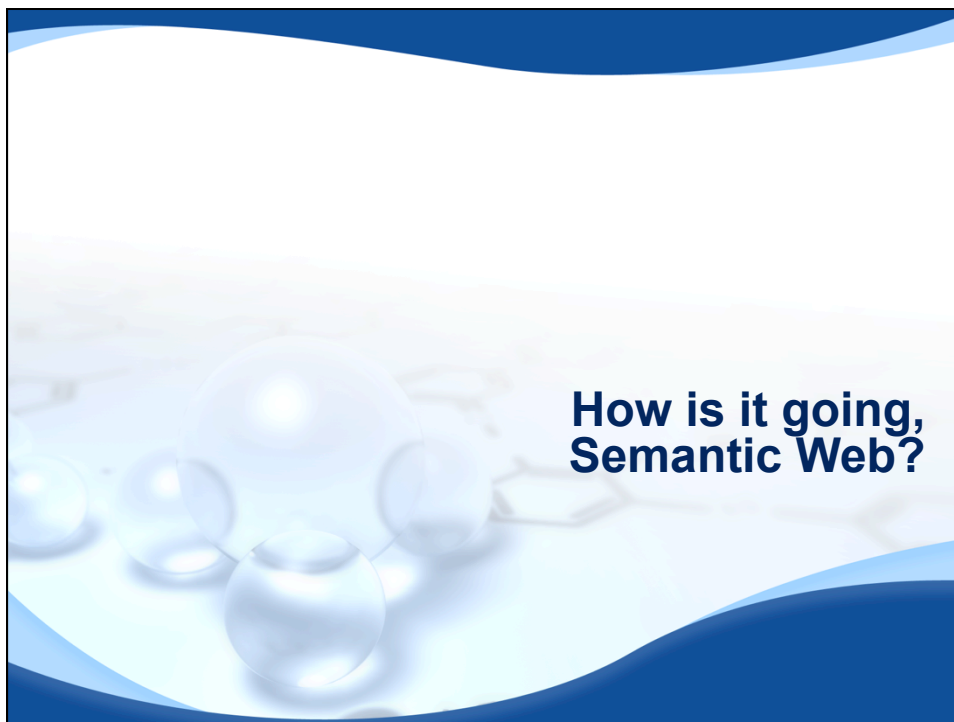
WebIST2012

**Jeff Z. Pan**

 UNIVERSITY  
OF ABERDEEN

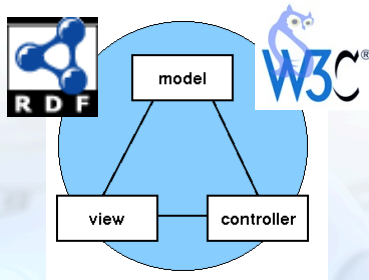
**TrOWL**  
www.trowl.eu

Department of Computing Science  
University of Aberdeen, UK



**How is it going,  
Semantic Web?**

## What's new in the Semantic Web



[Photo source: javaworld.com]

- **RDF**: a new **data model** for the Web
- **OWL**: standard **web ontology language** to define the meaning of vocabulary
- **SPARQL**: query language

4

## SW became widely accepted



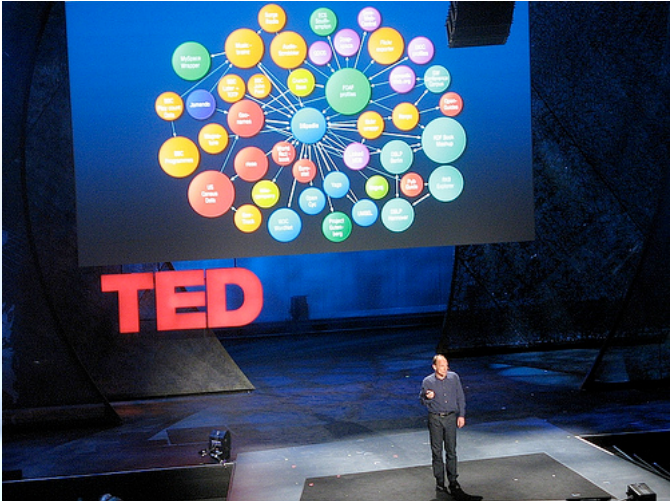
[Photo source: socialmeteor.com]

Two key steps:

1. **Building reusable ontologies**
2. **Using ontological vocabulary to annotate data**

5

# Linking open data



[Photo source: talis.com]

# Linked open data

data.gov.uk

BETA

Opening up government data

Home

Data

Apps

Who does what in Whitehall?

Who ministers are meeting

BBC MUSIC

HOME SHOWCASE

Most Played Artists

23-29 JANUARY 2012

The latest version of information displayed on this page

THIS WEEK'S

ALBUMS & SINGLES

Portico Quartet

RECOMMENDED BY CHRIS PARSONS

an anthology absorbing that LP from the inspired jazz soul.

BROWSE BBC REVIEWS 42 OF 48

DBpedia

DBpedia Blog | Get Involved | Get Help

About / News

Applications

Use Cases

Datasets

Online Access

DBpedia Live

Downloads

Interlinking

Development

Support

Publications

Credits

Contact / Imprint

DBpedia is a community effort to extract structured information from Wikipedia and to make this information available on the Web. DBpedia allows you to ask sophisticated queries against Wikipedia, and to link other data sets on the Web to Wikipedia data. We hope this will make it easier for the amazing amount of information in Wikipedia to be used in new and interesting ways, and that it might inspire new mechanisms for navigating, linking and improving the encyclopaedia itself.

News

DBpedia Spotlight has been selected for Google Summer of Code. Please apply now!

The Google Summer of Code (GSoC) is a global program that offers student developers (BSc/MSc/PhD) depends to write code for open source software projects. It has had thousands of participants since the first edition in 2005, connecting prospective students with mentors from open source communities such as Debian, KDE, Gnome, Apache Software Foundation, Mozilla, etc. For [...]

DBpedia 3.7 released, including 15 localized Editions

Hi all, we are happy to announce the release of DBpedia 3.7. The new release is based on Wikipedia dumps dating from late July 2011. The new DBpedia data set describes more than 3.64 million things, of which 1.83 million are classified in a consistent ontology, including 416,000 persons, 526,000 places, 106,000 music albums, 60,000 films, 17,500 [...]

Official DBpedia Live Release

We are pleased to announce the official release of DBpedia Live. The main objective of DBpedia is to extract structured information from Wikipedia, convert it into RDF, and make it freely available on the Web. In a nutshell, DBpedia is the Semantic Web mirror of Wikipedia. Wikipedia users constantly revise Wikipedia articles with updates happening almost [...]

BBC RADIO 1

Fearne Cotton

20/01/2012

LISTEN NOW

Fearne chats to The Ting Tings about BBC Innovating.

BBC RADIO 2

Ann Clearey Spins The Globe

19/01/2012

7

3



The screenshot shows the Schema.org homepage. At the top left, there is a logo with the Google, Bing, and Yahoo! logos and the text 'schema.org'. To the right of this is the 'Schema.org' title. Below the title is a search bar with the text 'schema.org' and a 'Search' button. Underneath the search bar is a navigation menu with links for 'Home', 'Schemas', and 'Documentation'. The main content area is titled 'What is Schema.org?' and contains several paragraphs explaining the purpose of the site. At the bottom, there are two links: 'Official OWL ontology: <http://schema.org/docs/schemaorg.owl>' and 'HTML microdata: <http://www.w3.org/TR/microdata/>'. A small number '8' is visible in the bottom right corner of the slide.

**Google**  
schema.org  
**bing YAHOO!**

# Schema.org

schema.org Search

Home Schemas Documentation

## What is Schema.org?

This site provides a collection of schemas, i.e., HTML tags, that webmasters can use to markup their pages in ways recognized by major search providers. Search engines including Bing, Google, Yahoo! and Yandex rely on this markup to improve the display of search results, making it easier for people to find the right web pages.

Many sites are generated from structured data, which is often stored in databases. When this data is formatted into HTML, it becomes very difficult to recover the original structured data. Many applications, especially search engines, can benefit greatly from direct access to this structured data. On-page markup enables search engines to understand the information on web pages and provide richer search results in order to make it easier for users to find relevant information on the web. Markup can also enable new tools and applications that make use of the structure.

A shared markup vocabulary makes easier for webmasters to decide on a markup schema and get the maximum benefit for their efforts. So, in the spirit of sitemaps.org, search engines have come together to provide a shared collection of schemas that webmasters can use.

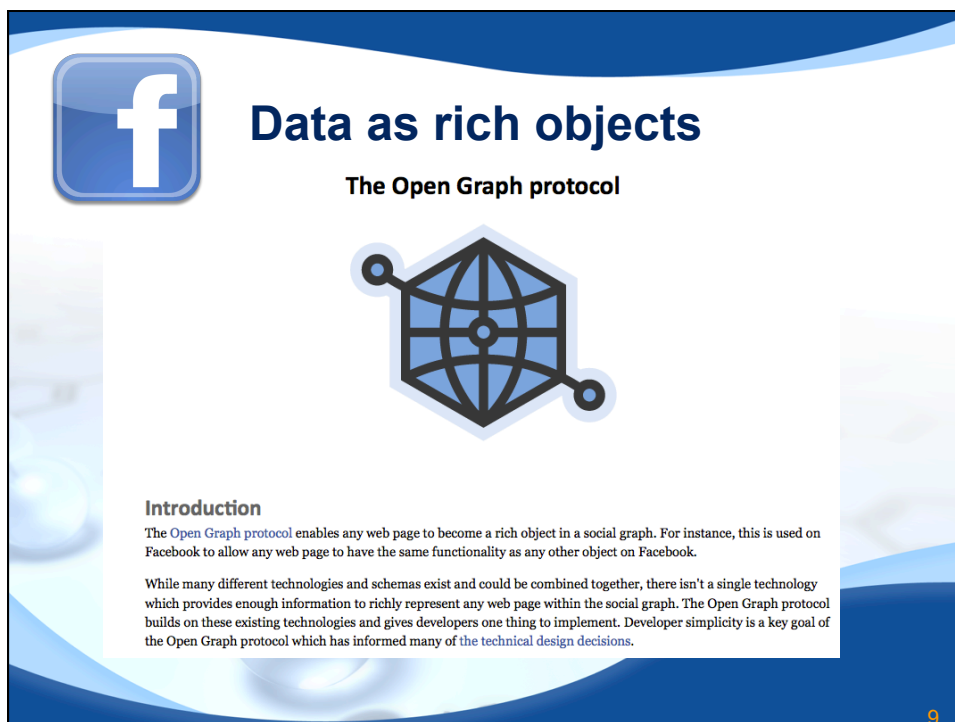
We invite you to [get started!](#)

View our blog at [blog.schema.org](http://blog.schema.org).

Official OWL ontology: <http://schema.org/docs/schemaorg.owl>

HTML microdata: <http://www.w3.org/TR/microdata/>


8



The slide features a Facebook 'f' logo in a blue square on the left. To its right is the title 'Data as rich objects' in a large, bold, dark blue font. Below the title is the subtitle 'The Open Graph protocol' in a smaller, bold, dark blue font. In the center of the slide is a stylized icon of a globe with a grid of lines, representing a network or graph. Below the icon is a section titled 'Introduction' in bold. The text in this section explains that the Open Graph protocol enables any web page to become a rich object in a social graph, specifically mentioning its use on Facebook. It also states that while many different technologies and schemas exist, there isn't a single technology that provides enough information to richly represent any web page within the social graph. The Open Graph protocol builds on these existing technologies and gives developers one thing to implement, with developer simplicity being a key goal. The slide is numbered '9' in the bottom right corner.

# Data as rich objects

## The Open Graph protocol



### Introduction

The Open Graph protocol enables any web page to become a rich object in a social graph. For instance, this is used on Facebook to allow any web page to have the same functionality as any other object on Facebook.

While many different technologies and schemas exist and could be combined together, there isn't a single technology which provides enough information to richly represent any web page within the social graph. The Open Graph protocol builds on these existing technologies and gives developers one thing to implement. Developer simplicity is a key goal of the Open Graph protocol which has informed many of the technical design decisions.

9



## Creating semantic data with mobiles



Your mobile knows “everything” about you:

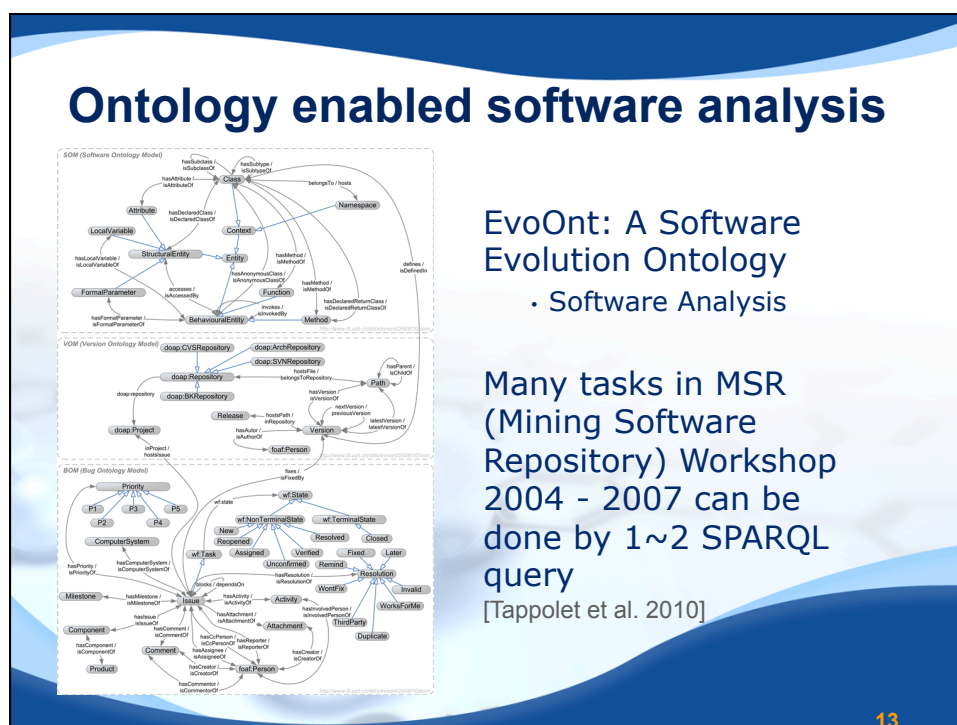
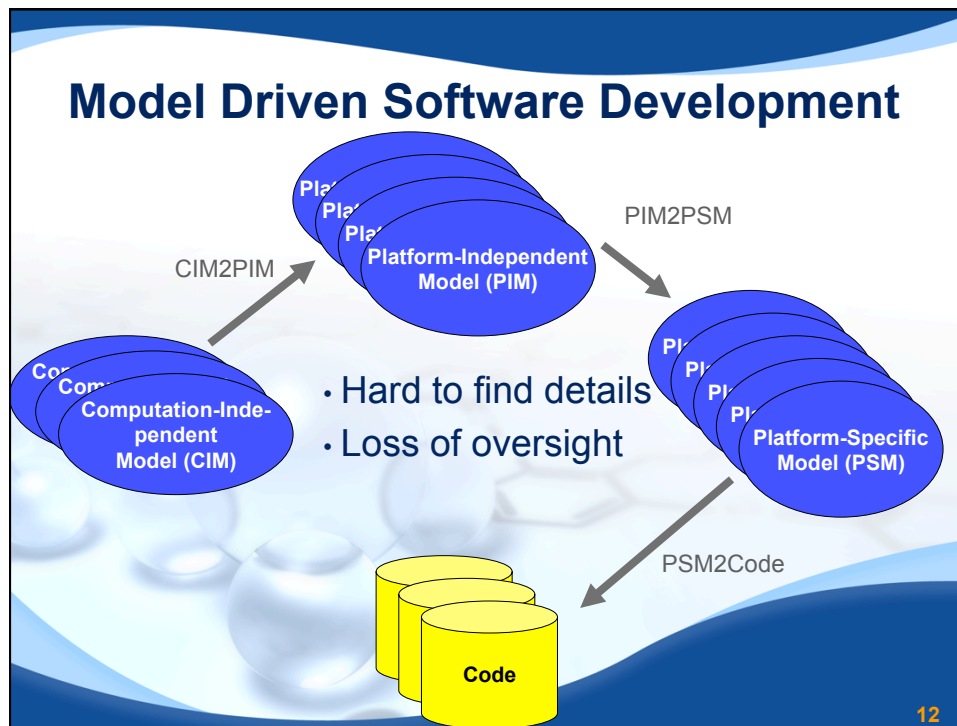
- Where you are
- Who your social network contacts are

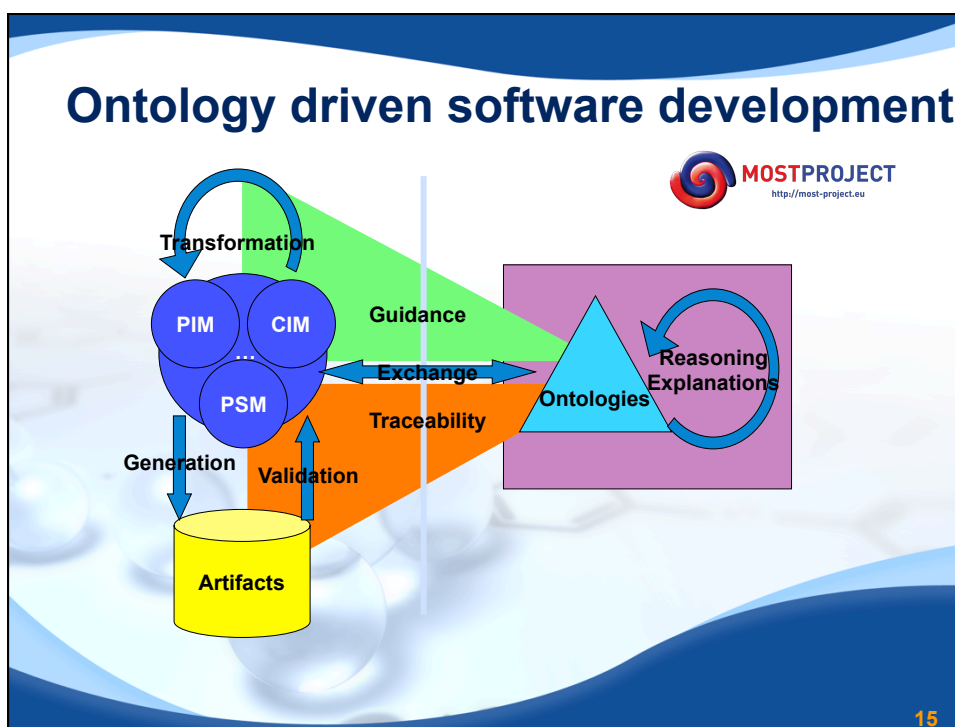
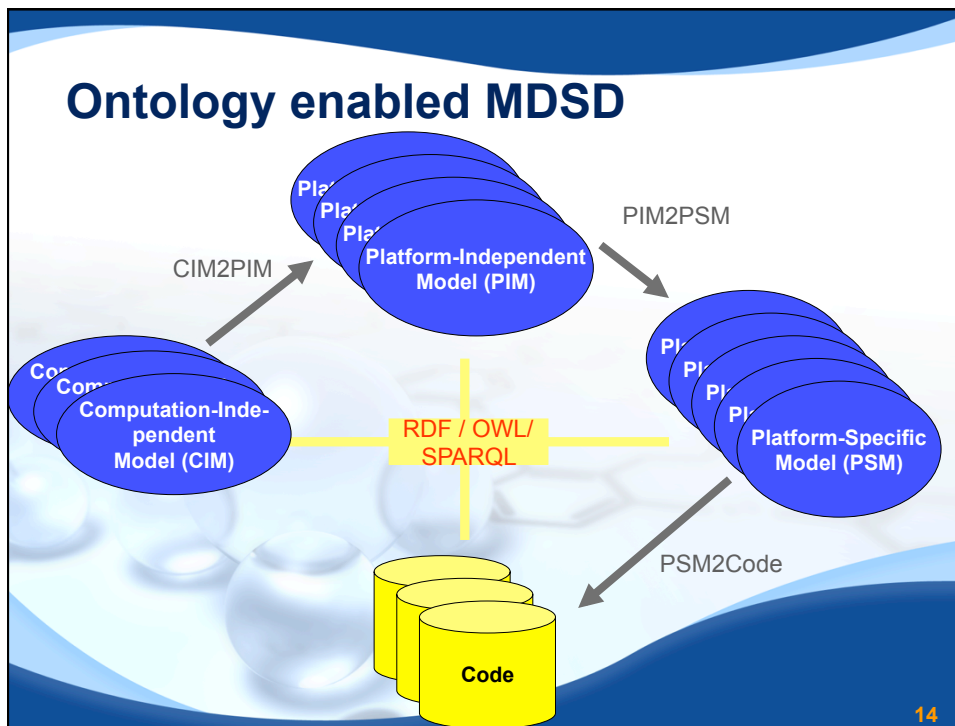
10

## Semantic Web Enabled Software Engineering: Why and How

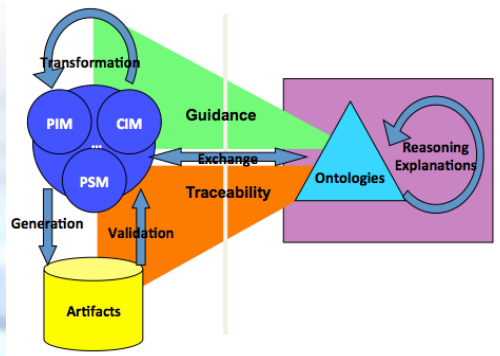


11





## How to bridge SE models and ontologies

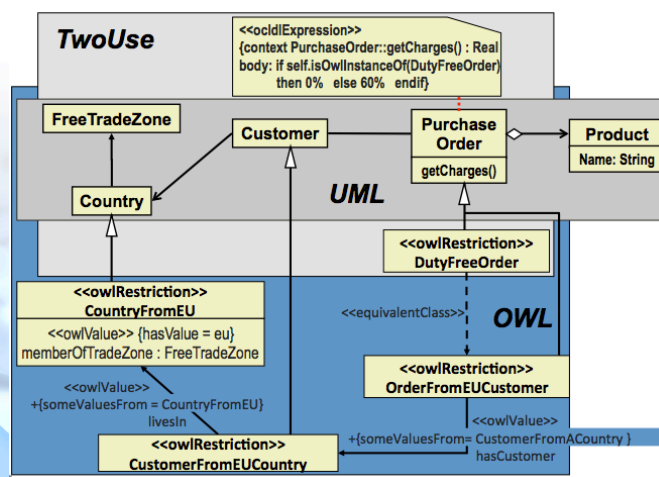


### Approach 1: Integrated models

- Static SE models
- Ontologies

16

## Integrated model: example 1



### Two components

- UML class diagrams
- Ontologies

17

## Integrated models: example 2

Two components

- Physical Device DSL (PDDSL) models
- Ontologies

```

DeviceType "Cisco_7603"
SubClassOf: pd_hasConfiguration some ( pd_hasSlot some ( pd_hasCard some Cisco_7600_SIP ))
longName : "CISCO 7603 CHASSIS" description : "The Cisco® 7603 Router is a high-performance ..." allowed : {
PossibleConfiguration "Cisco_7603_Configuration" {
Slot "1" allowed : "Supervisor_Engine_2" "Supervisor_Engine_720" required : true
Slot "2" allowed : "Supervisor_Engine_2" "Supervisor_Engine_720" "Catalyst_6500_Module"
required : false
Slot "3" allowed : "Catalyst_6500_Module" required : false
}

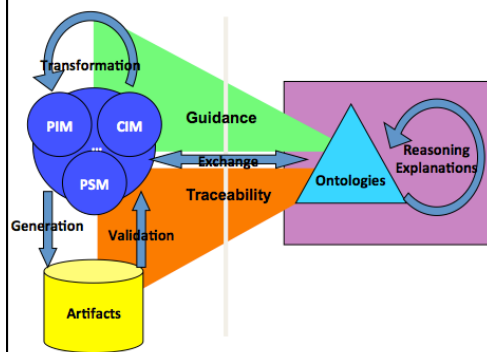
Device serialNumber : "cisco_7603" hasType : "Cisco_7603"
configuration :
{
Slot id : "1" : Card serialNumber : "supervisor_2_5" hasType : "WS-X6K-S2U-MSFC2"
Slot id : "2" :
Slot id : "3" :
}

```

[Miksa et al. 2009]

18

## How to bridge SE models and ontologies



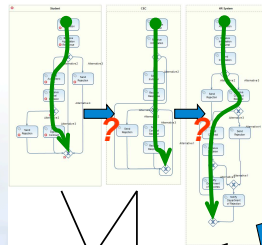
### Approach 2: problem reduction to task ontologies

- To reduce an SE problem
- Into ontology reasoning problem

19



## Bussiness process refinement



“Traceability”:  
Which activities  
cause the  
problem?

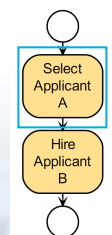
Justification:  
Which axioms  
cause the  
problem?

- To reduce business process refinement validation
- Into ontology consistency/coherency checking

[Ren et al. 2009b; Groener and Staab 2009]

20

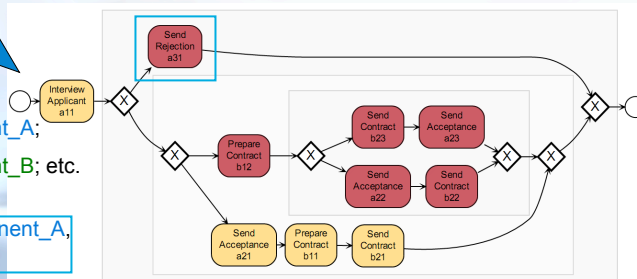
## Bussiness process refinement: example



1. In pre-refinement process:  
Component\_A subclassOf to only (Component\_A or Component\_B);  
Component\_B subclassOf from only Component\_B or Component\_A;
2. In post-refinement process:  
a31 subclassOf to some End  
b12 subclassOf (to some b23) and (to some a22), etc.

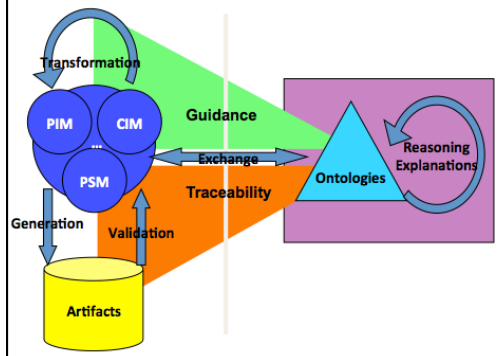
**Invalid**

3. Mapping:  
a31 subclassOf Component\_A;  
b23 subclassOf Component\_B; etc.
4. Uniqueness:  
Disjoint(Start, End, Component\_A,  
Component\_B), etc.



21

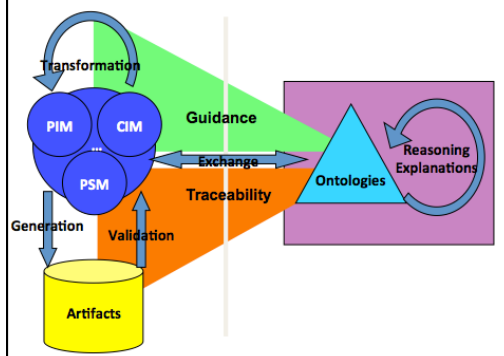
## Traceability: exploiting the integrated models and task ontologies



- Help understand the associations and dependencies
- By exploiting reasoning and query answering

22

## Dynamics and complete of knowledge



- Two types of knowledge**
- Some we have complete understanding (DB)
  - Some we only know partially (SW)

23

## Semantic Web supports BOTH

NBox (Negation As Failure Box) enabled ontologies [JTS]

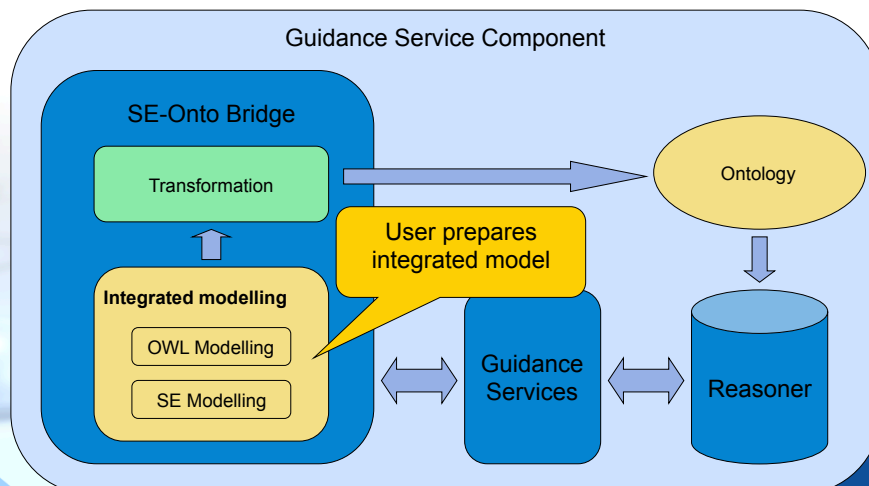
- TBox: **schema** axioms
- ABox: **data** axioms
- **NBox**: a set of **closed vocabulary**
  - NAF of NBox concepts can be simulated by full negation
    - Such as **Spicy**
  - Require **incremental reasoning**

Food	Note
Curry Chicken	Minor Spicy
Salmon Fillet	
Spicy Grilled Shrimp	Spicy
Pepper Salad	Vege

Name	Vegetarian
Jeff	No
Yuting	Yes
Jek	
Yuan	No

24

## Architecture



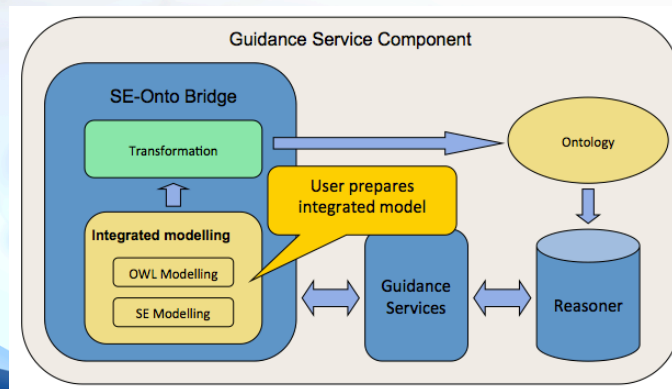
[Ren et al. 2009]

25

## Guidance services

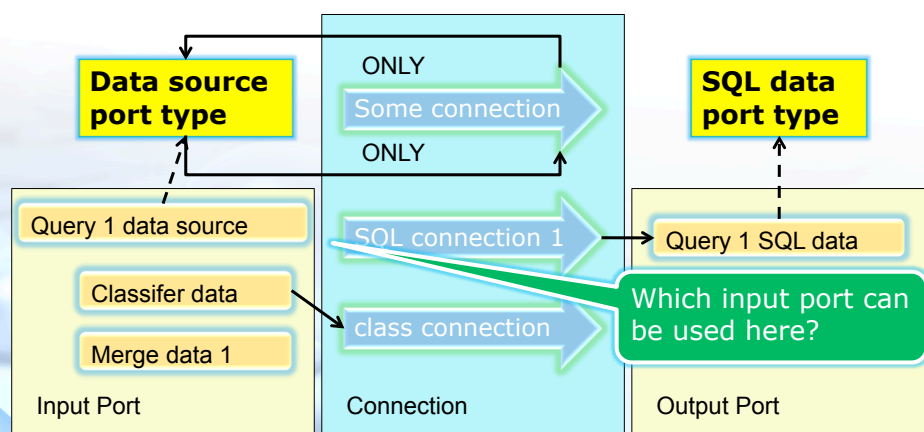
### Todo list services for developers:

- To recommend the next available actions/options
- To validate the current action/option



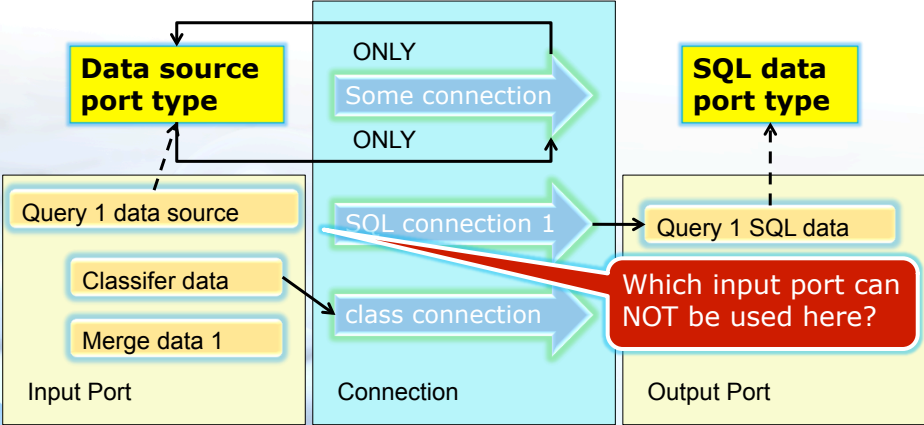
26

## Guidance services: example



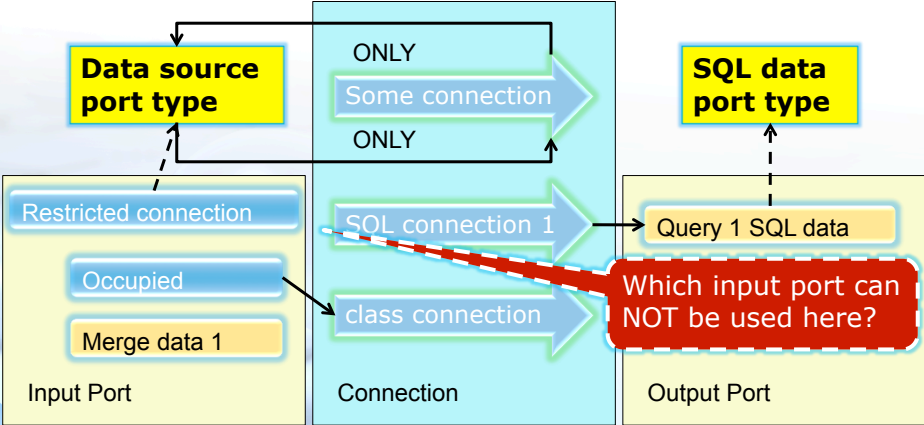
27

# Guidance services: example



28

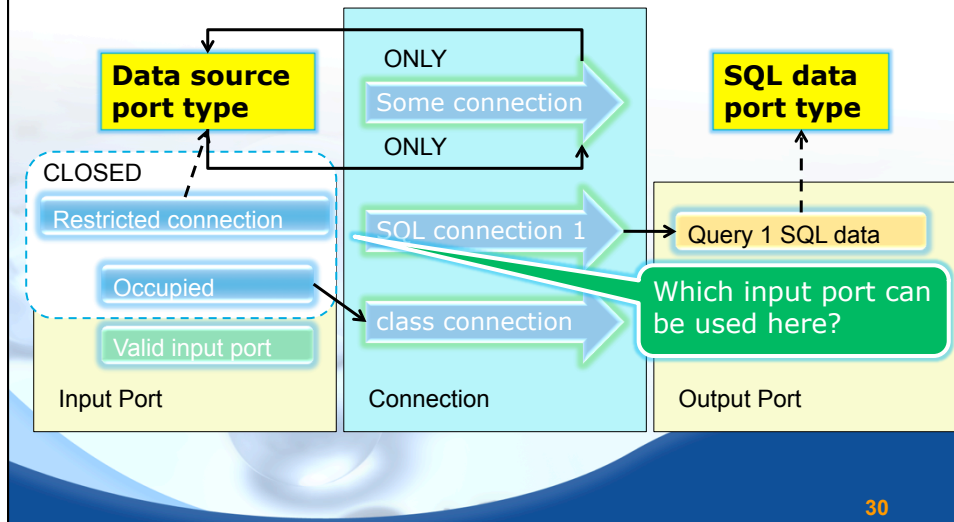
# Guidance services: example



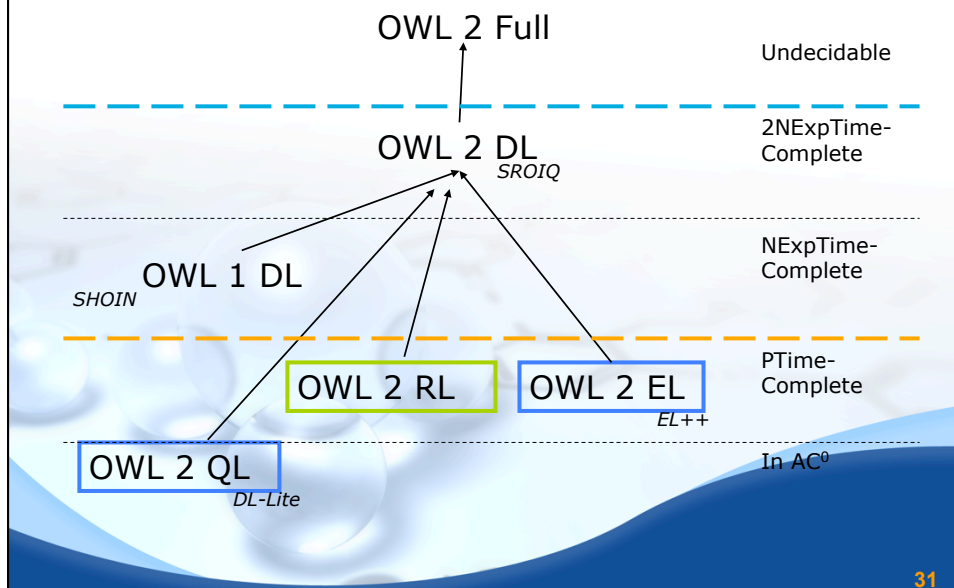
29



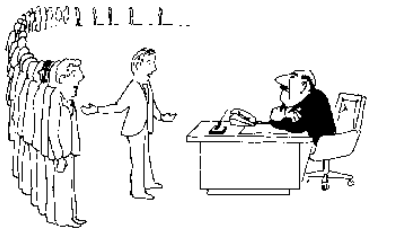
## Guidance services: example



## OWL: Standard Ontology Language



## Approximate reasoning plays a key role



### Inputs

- Ontology
- Queries

### Outputs

- soundness preserving
- completeness preserving

I can't find any algorithms, but neither can all these famous people.

[Garey & Johnson. Computers and Intractability: A Guide to the Theory of NP-Completeness. Freeman, 1979.]

32

## Faithful Syntactic Approximate Reasoning [AAAI2010]

### Approximation

- Directly represent non-OWL2-EL concepts with fresh named concepts
  - E.g.,  $\forall r.C \text{ subClassOf } D \rightarrow A_{\forall r.C} \text{ subClassOf } D$
- Maintain semantic relations for these named concepts
  - *complementary relations*
  - *cardinality relations*

### Reasoning

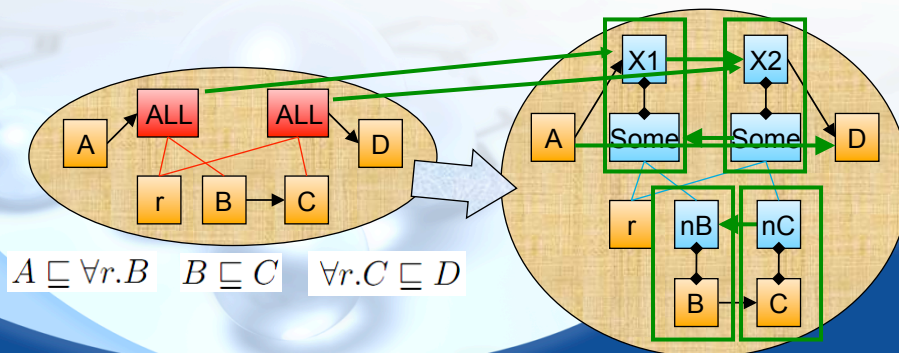
- Using additional **tractable** completion rules to recover the semantics

33

## Example: How Does it Work

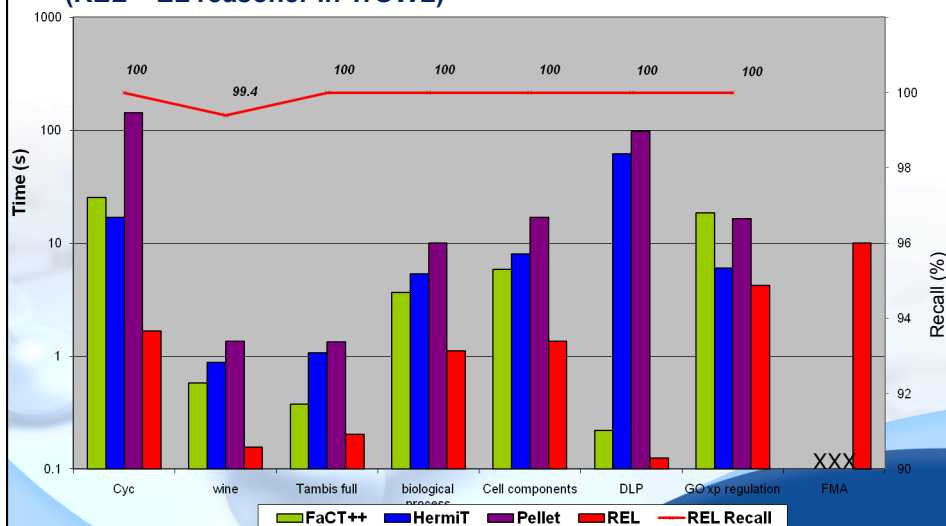
Additional completion Rules (on top of the EL ones), e.g.

- Handling complement
  - E.g.  $B \text{ subClassOf } C \Rightarrow \neg C \text{ subClassOf } \neg B$
- Handling cardinality
  - E.g.  $A \text{ subClassOf } \geq 3 r. B \Rightarrow A \text{ subClassOf } \geq 2 r. B$
- **Soundness preserving and tractable**



34

## Evaluations for the Oxford Benchmarks (REL – EL reasoner in TrOWL)



35

## TrOWL: a tractable semantic reasoning infrastructure

Approximate reasoning

[AAAI2007, AAAI2010]

Parallel reasoning

[JIST2011]



Stream reasoning

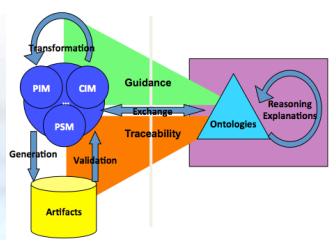
[CIKM2011]

Local closed world reasoning

[JTS]

36

## Benefits for Software Engineering



- Integrated modelling and querying
- Guidance services for developers:
  - To recommend the next available actions
  - To validate the current action
- Ontologies available at deployment time

37

## Vision: Programs that can understand and improve themselves



[Photo source: northjersey.com]

## API4KB





## Possible impacts on the Semantic Web

© Original Artist  
Reproduction rights obtainable from  
www.CartoonStock.com



- More data
  - Linked software data
- Easier to debug, reuse and extend semantic applications
- Software becomes semantics software
- Software/service as a first class citizen like data?

40

## Where to find more information

International SWESE workshop series

<http://www.abdn.ac.uk/~csc280/event/workshop/swese2011/>

Two books coming:

- **J. Z. Pan**, S. Staab, U. Assmann, J. Ebeert, Y. Zhao (Eds.), *Ontology Driven Software Development*. Springer. 2012
- **J. Z. Pan** and Y. Zhao (Eds.), *Semantic Web Enabled Software Engineering*. AKA Verlag Heidelberg and IOS Press, Amsterdam, The Netherlands. 2012



**Semantic Web Enabled  
Software Engineering**

WebIST2012

Thank you

... questions?

 UNIVERSITY  
OF ABERDEEN

 **Trowl**  
www.trowl.eu