



Applying Semantic Technology

John Davies

Head of Next Generation Web Research

GCTO

john.nj.davies@bt.com



Overview

- What is the Semantic Web?
- Semantic Technology application areas
 - Knowledge management
 - Information integration
 - Service-oriented architectures
 - Health sector

Today's Web is...

- A place where
 - computers do the presentation (easy) and
 - people do the linking and interpreting (hard).
- *Why not get computers to do more of the hard work?*

Semantic Web

- Today's web
 - Machine-to-human – emphasis on presentation
- Semantic web vision
 - “an **extension** of the current web in which **information** is given well-defined **meaning**” (Tim Berners-Lee)
 - making web-based information **machine-processable**
 - `<bold>use bold font</>` → `<product-code>1234-6/A</>` (`>`)
 - also rules (reasoning; inheritance; business logic; conflict detection) (RDF, OWL)

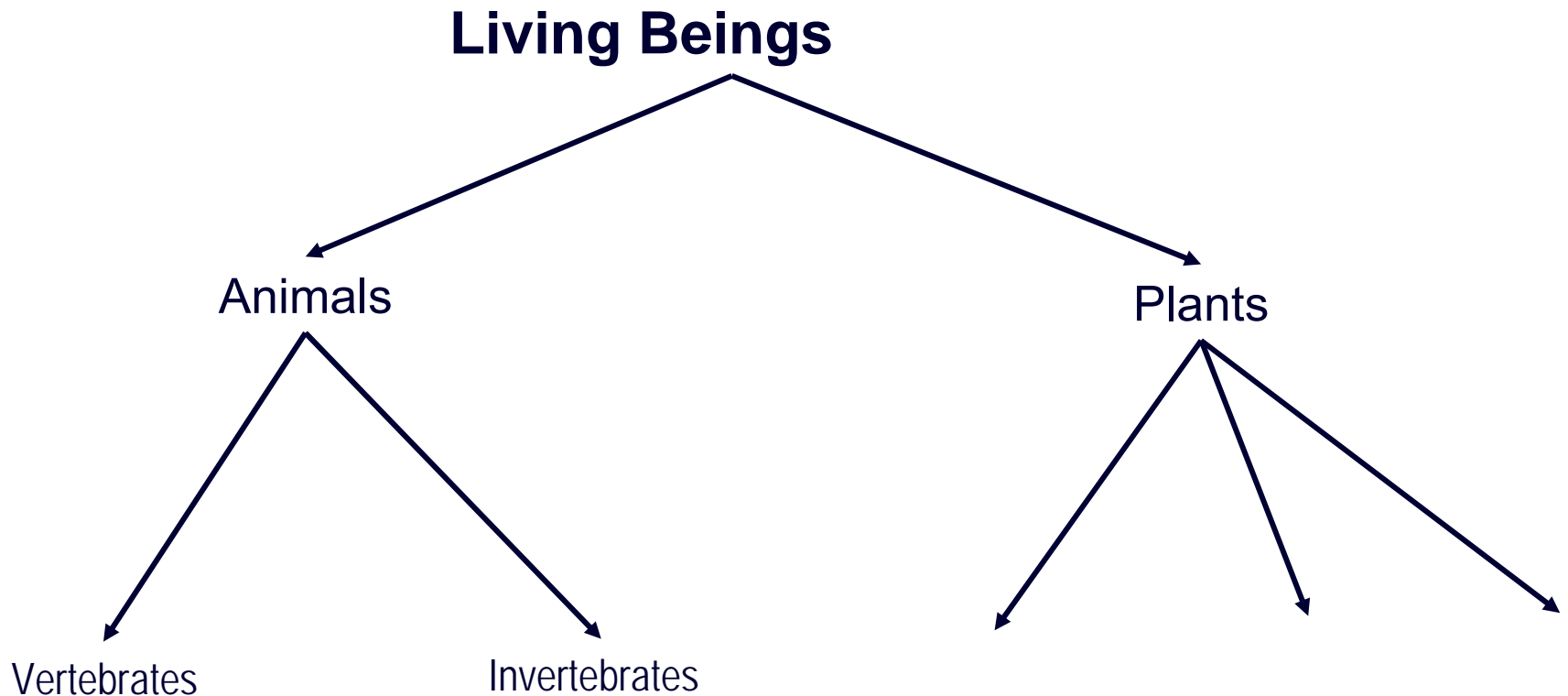


Ontologies

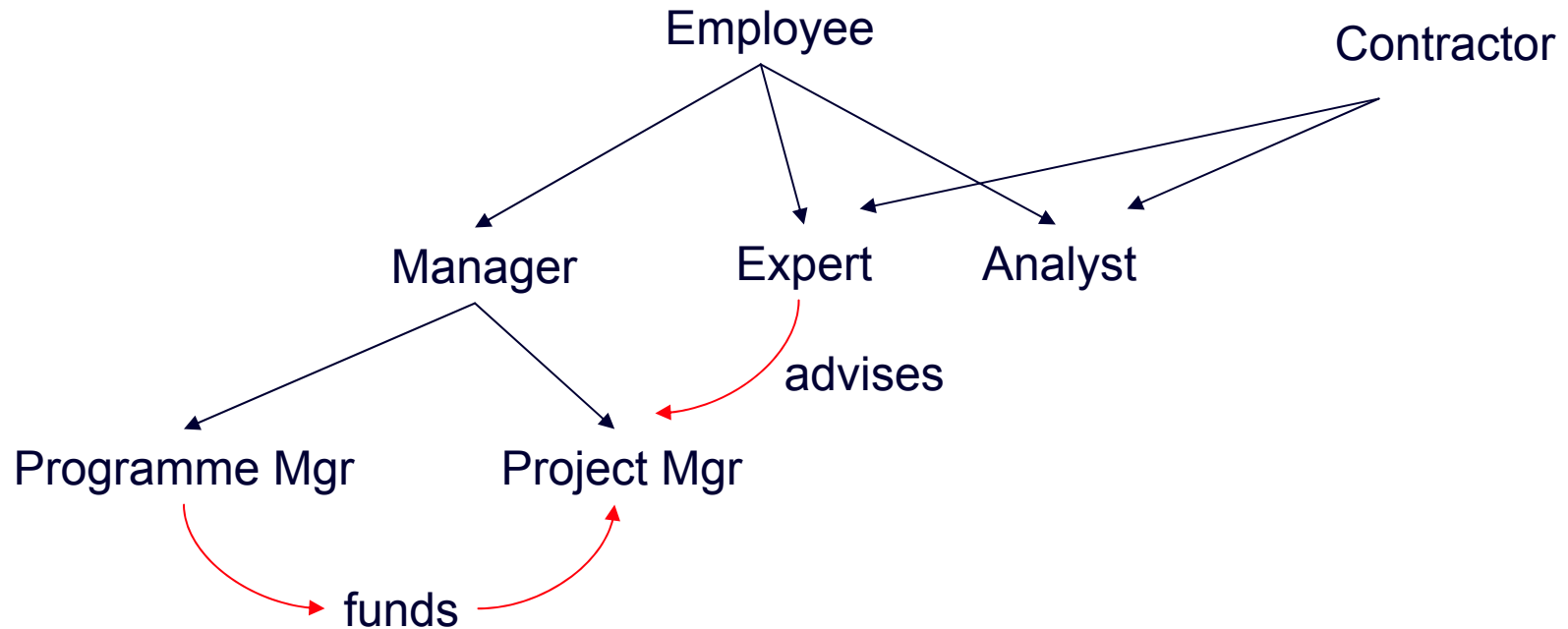
- An ontology describes a formal specification of a certain domain:
 - Shared understanding of a domain of interest
 - Formal and machine processable model of a domain of interest (telecoms systems, gene structures, public services, ...)
- W3C has developed standards for describing web-based ontologies
 - Logic-based languages RDF and OWL
 - Specify ontologies and link them to information such as (parts of) web pages

Ontologies & Taxonomies

Taxonomy is a classification system where each node has only one parent – simple ontology



Ontology of People and their Roles



Typically, we want a richer ontology with more relationships between concepts:

Why develop an ontology?

- ***To define information (e.g. web-based) more precisely and make it more amenable to machine processing***
- To make domain assumptions explicit
 - Easier to change domain assumptions
 - Easier to understand and update legacy data
- To separate domain knowledge from operational knowledge
 - Re-use domain and operational knowledge separately
- A community reference
- To share a consistent understanding of what information means
- To allow reasoning about data

Application Areas for Semantic Technologies

- Knowledge Management
- Semantic Web Services
- Systems Interoperability/Information integration
- Health sector
 - Medical Vocabularies

In three words

Semantic knowledge management classifies, finds, distributes, shares and uses knowledge based on **meaning** **not** the particular **words** used to represent meaning.

Words and meanings

- same word, different meanings

Jaguar



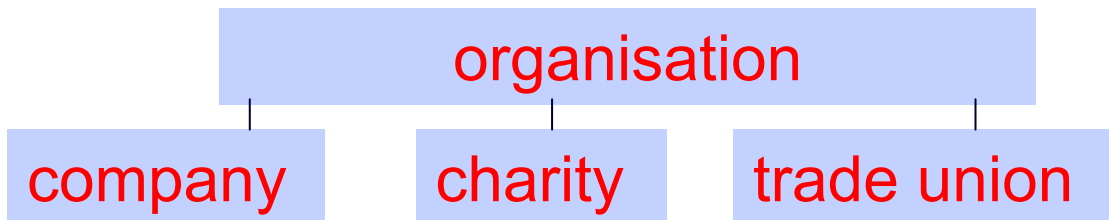
- different words, same meaning

disability legislation

equal opportunity laws

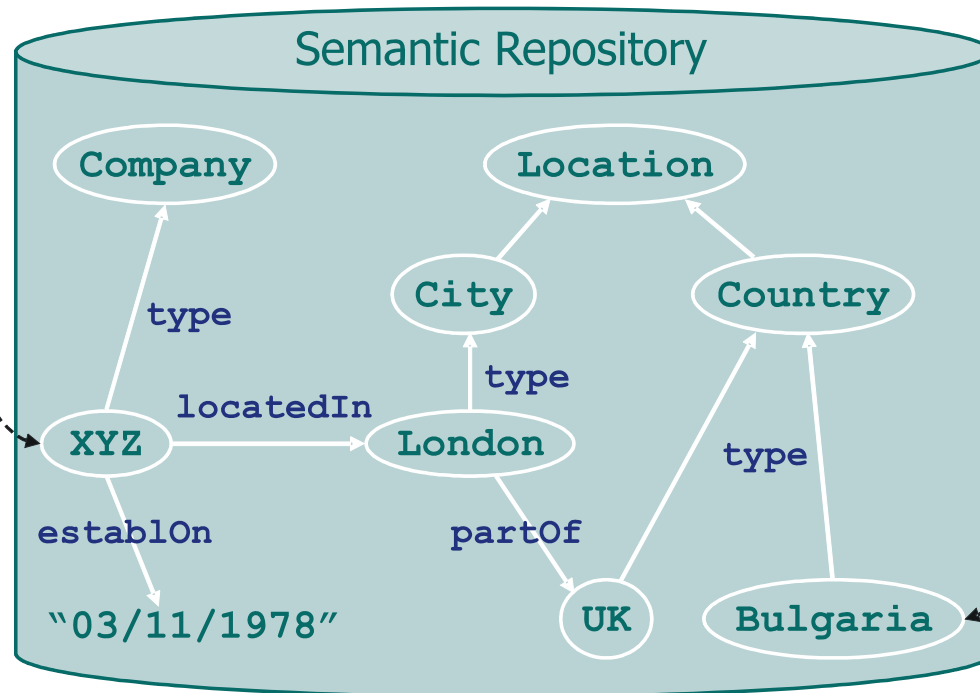
- different words, related meaning

- leading to inheritance (and other) reasoning



Semantic Annotation

XYZ announced profits in Q3, planning to build a \$120M plant in Bulgaria, and more and more and more and more text.



Semantic Browsing

The screenshot displays a Microsoft Internet Explorer browser window showing a BBC News article. The browser's address bar shows the URL: [BBC NEWS | Business | South Korea fines Microsoft \\$32m - Microsoft Internet Explorer](#). The page title is "BBC NEWS" and the article title is "South Korea fines Microsoft \$32m". The article text includes: "Microsoft has been fined 33bn won (\$32m; £18.4m) following an antitrust ruling by South Korean regulators." and "The US software giant was ordered to unbundle its messaging service from its Windows software by South Korea's Fair Trade Commission." The article also mentions that regulators ordered Microsoft to introduce a version of Windows which enables the embedding of services by other software companies, and that the news came as Microsoft announced plans to invest \$1.7bn in India. Chairman Bill Gates is quoted as saying the company planned to increase its workforce in the country "to 7,000 over the next three to four".

On the left side of the browser window, there is a "KIM Plugin" sidebar. It features a tree view under the heading "Entity" with various categories and sub-items, each with a checkbox and a small colored square icon. The categories include: Abstract, ContactInformation, GeneralTerm, Language, Number, Topic, BusinessAbstraction, NaturalPhenomenon, SocialAbstraction, TemporalAbstraction, Happening, Event, Situation, TimeInterval, Object, Agent, Location, Product, Service, Statement, Account, and Brand. Below the tree view, there are buttons for "Classes", "Entities", and "Place Links".

The browser window also shows a navigation menu on the left side of the page with categories like "News Front Page", "World", "UK", "England", "Northern Ireland", "Scotland", "Wales", "Business", "Market Data", "Your Money", "E-Commerce", "Economy", "Companies", "Fact Files", "Politics", "Health", "Education", "Science/Nature", "Technology", "Entertainment", "Have Your Say", "Magazine", and "In Pictures".

Microsoft Corporation is a Public Company located in United States and Worldwide.

Designs, develops, manufactures, licenses, sells and supports a wide range of software products.

Its webpage is www.microsoft.com. It is traded on NASDAQ with the index MSFT. Key people include:

Bill Gates - Chairman, Founder

Steve Balmer - CEO

John Conners - Chief Financial Officer

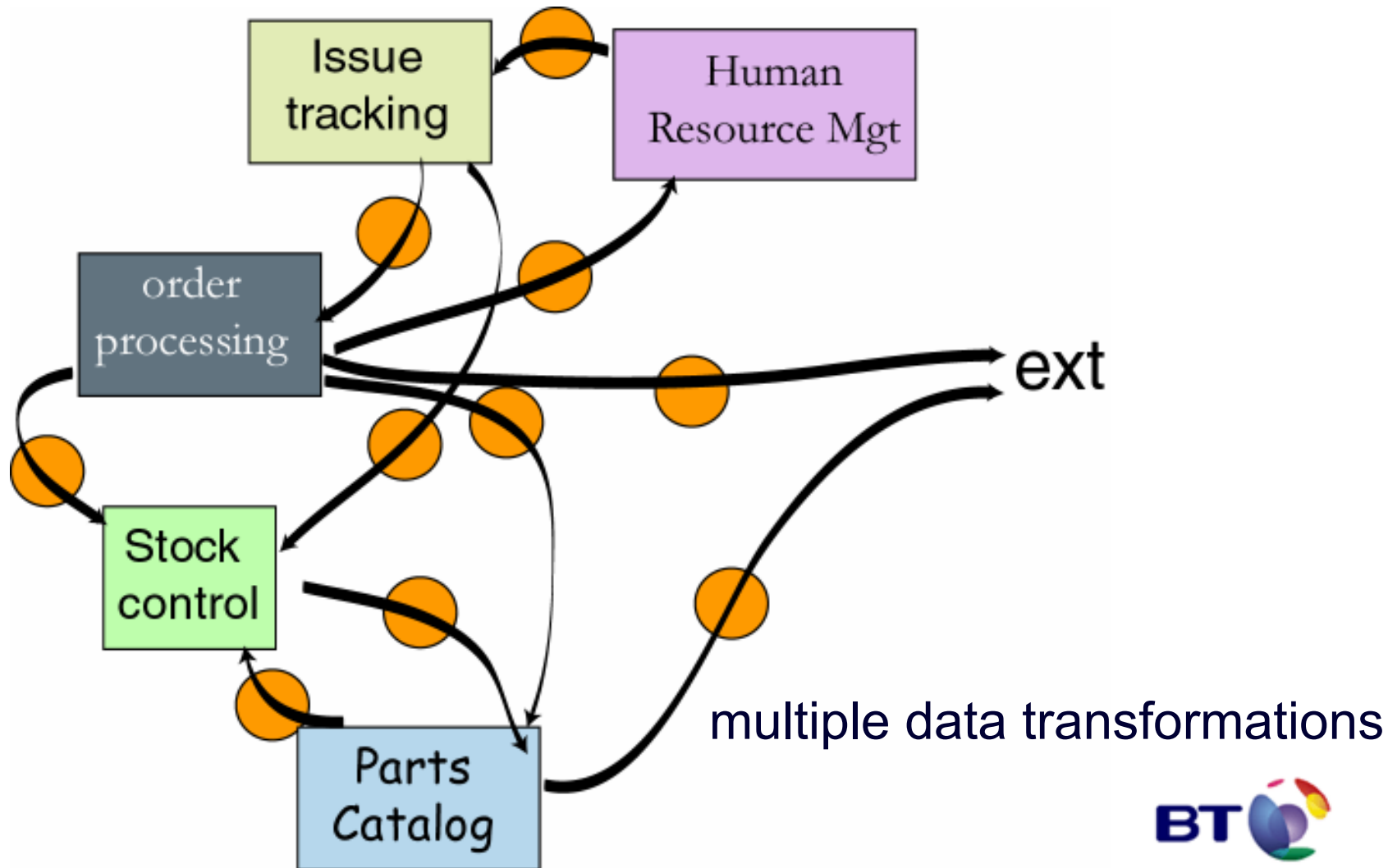
Last year its revenues were \$36.8bn and its net income was \$8.2bn.

The screenshot shows a web browser window with a news article. The article text includes: "ordered to abandon its messaging service from its Windows software by South Korea's Fair Trade Commission. Microsoft was accused of 'hurting the interest of consumers' Regulators ordered Microsoft to introduce a version of Windows which enables the embedding of services by other software companies. The news came as Microsoft announced plans to invest \$1.7bn in India. Chairman Bill Gates said the company planned to increase its workforce in the country 'to 7,000 over the next three to four'". The browser interface includes a sidebar with categories like E-Commerce, Economy, Companies, Fact Files, Politics, Health, Education, Science/Nature, Technology, and Entertainment. A search bar and navigation buttons are also visible.

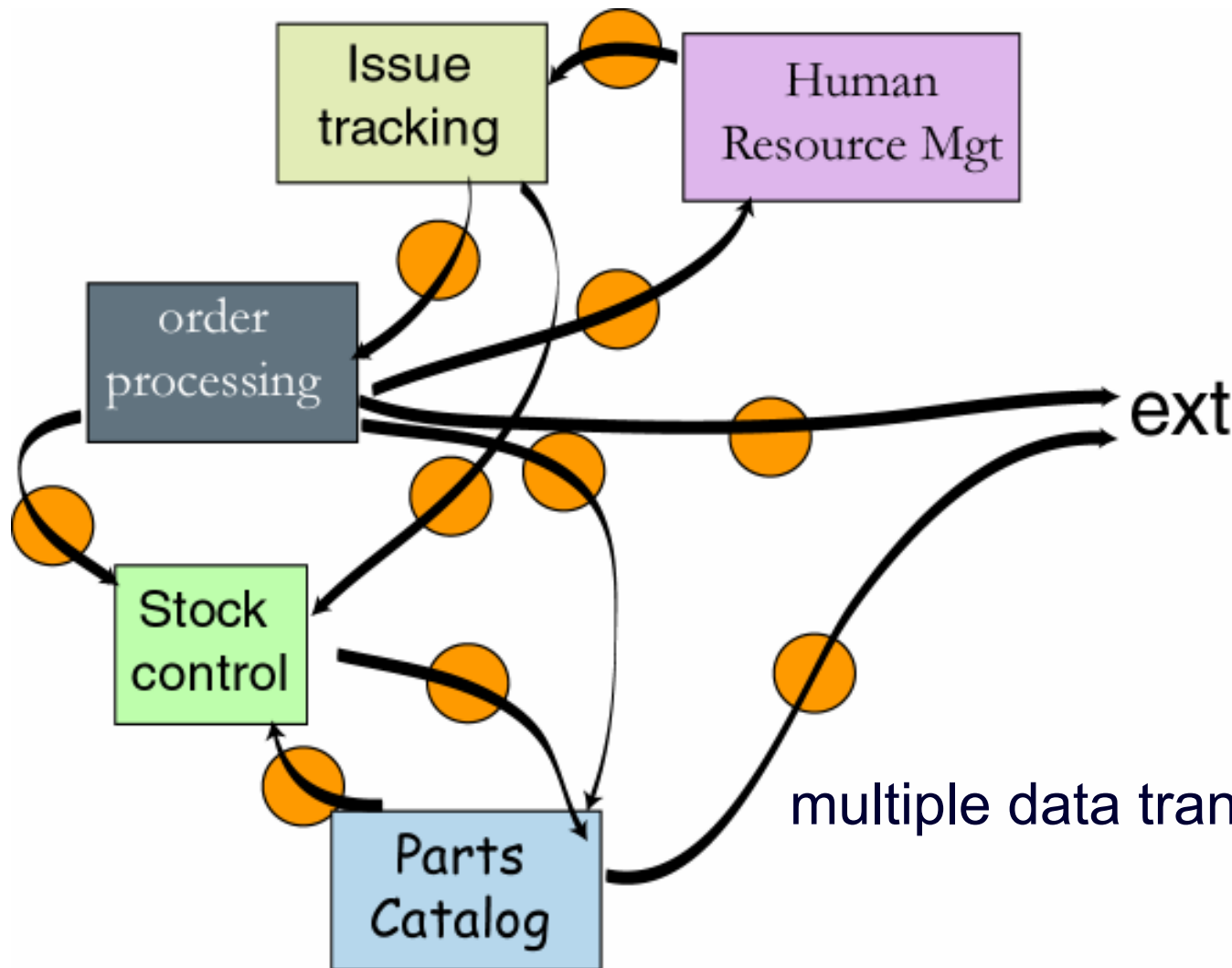
Precision in Semantic Web Search

- Semantic Search could match
 - a query: *Documents concerning a telecom company in Europe, John Smith as a director, and a date in the first half of 2002.*
 - With a document containing: *“At its meeting on the 10th of May, the board of Vodafone appointed John Smith as CTO”*
- Traditional search engines cannot do the required reasoning:
 - Vodafone is a mobile operator, which is a kind of telecom company;
 - Vodafone is in the UK, which is a part of Europe;
 - CTO is a type of director
 - 5th of May is a "date in first half of 2002”

Information Integration 2007



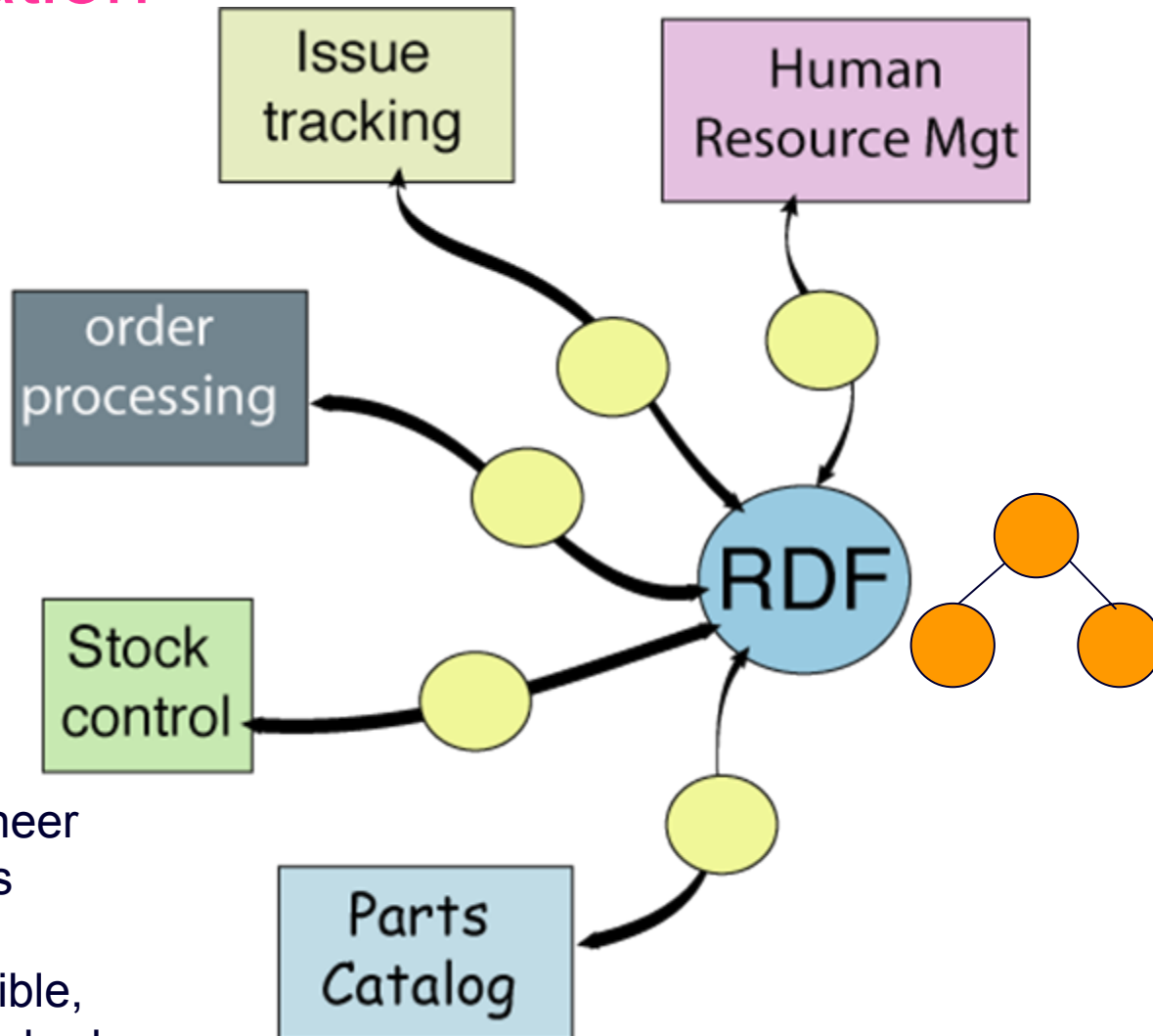
43% of businesses resort to manual processes and/or new software when integrating information for reporting



multiple data transformations



Semantic Technology & Information Integration



No need to re-engineer legacy data sources

Lightweight, extensible, based on open standards



Courtesy Tim Berners-Lee and W3C

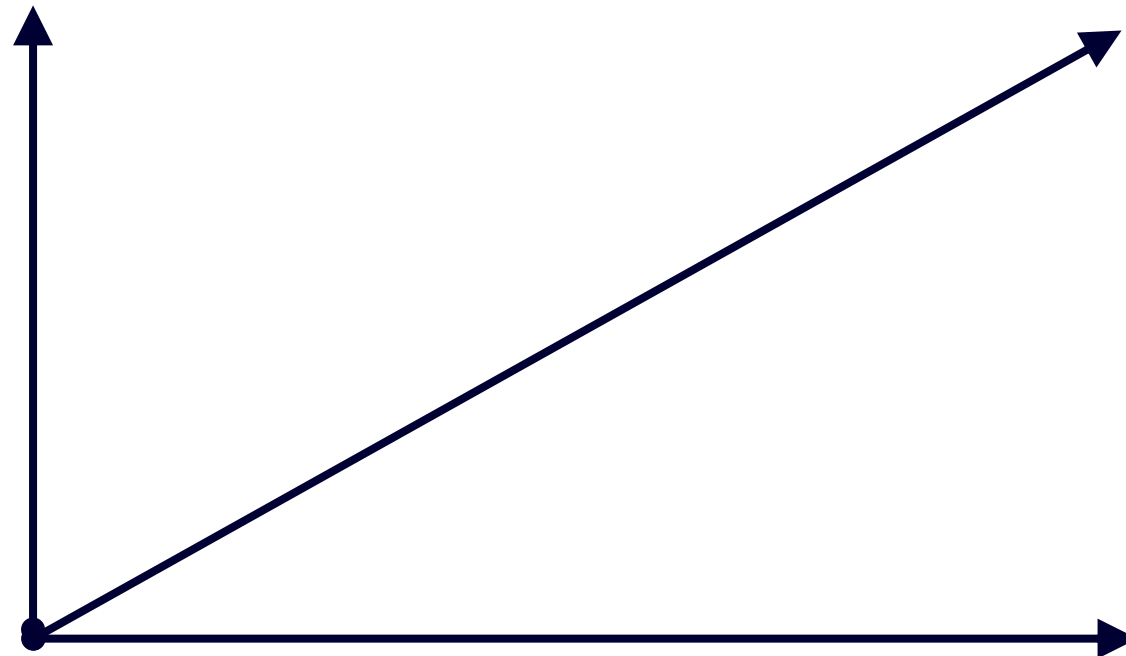
Semantic Technology & SOA

Web Services

computational objects

Semantic Web Services -

intelligent service discovery, interoperation, composition



WWW

static, unstructured info

Semantic Web

machine-processable info



Semantic Web Services

- Enabling the SOA
- The (semi-)automatic discovery and composition of services
- BT – Project DIP
 - EU collaborative project 2004 - 2006
 - dip.semanticweb.org
 - WSMO framework
 - Proposed as a W3C standard for describing services semantically

Semantic Technology & Health

- SNOMED-CT
 - Standard medical vocabulary
 - 400000+ concepts
- Expressed formally in Description Logic (formal basis of semantic web languages)
 - Reasoning
 - Consistency checking
- Semantic Technology in use today!

Semantic reasoning at the point of care

- Why do you need semantic reasoning in the live environment?
 - Post coordination (i.e. allowing users – clinicians - to create new terms)
 - Can never enumerate all concepts in advance
 - New terms are created by combining and/or extending old terms and description logic reasoning determines the correct logical place for new concepts
- Issue
 - Simple examples:
 - If user has built 'allergy + almond' how does the application work out to retrieve this information when someone else asks for patients with allergy to nut?
 - A procedure involving a laparoscope is a keyhole procedure: how is the system to determine the number of keyhole procedures if not explicitly stated to be such?
 - Post-coordination means description logic reasoning to solve these problems

Semantic deployment

- Semantic Technology has applications on and beyond the web
- First applications areas may be in information-intensive industry sectors (health, pharma, finance, ...)
- Some sectors are using the technology today
- Many others are experimenting
- Semantic Technology is going mainstream...
 - Oracle – adoption of RDF in 10g
 - Google: Froogle – RDF-based catalogues
 - Much start-up activity (Radar Networks, Metaweb, Ontoprise, Ontotext)
 - FOAF, RSS
 - SNOMED, GO – the Gene Ontology,

Thank you for your time

Questions?

