Applications

CSE 595 – Semantic Web

Instructor: Dr. Paul Fodor

Stony Brook University

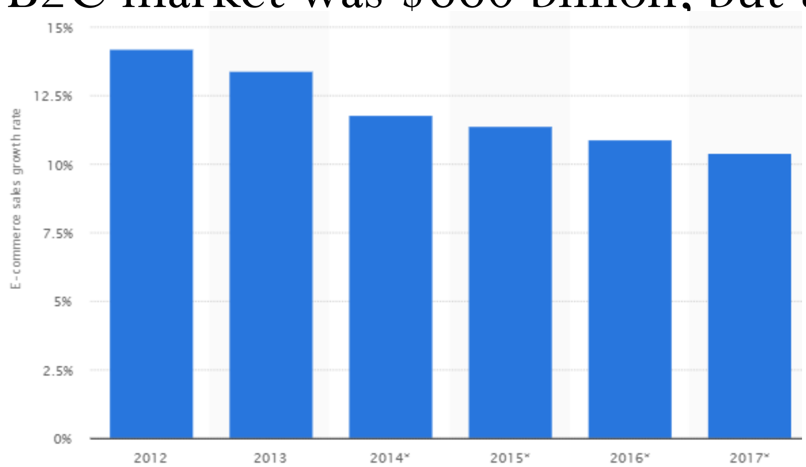
http://www3.cs.stonybrook.edu/~pfodor/courses/cse595.html

Lecture Outline

- GoodRelations
- BBC Artists
- BBC World Cup Website
- Government Data
- New York Times
- Sig.ma and Sindice
- Swoogle
- OpenCalais
- Schema.org
- data.world
- Elsevier
- Audi Data Integration
- Swiss Life
- EnerSearch
- E-Learning
- Web Services
- Multimedia Collection Indexing at Scotland Yard
- Online Procurement at Daimler
- Device Interoperability at Nokia
- Publication Management

GoodRelations

- E-commerce, and in particular Business-to-Consumer (B2C) e-commerce, has been one of the main drivers behind the rapid adoption of the World Wide Web in everyday live
  - It is now commonplace to see URLs listed on storefronts and goods vehicles
  - Taking the UK as an example, the B2C market has grown from £87 million in April 2000 to £68.4 billion by the end of 2009, a thousand-fold increase over a single decade
    - USA 2017 B2C market was $660 billion, but the growth is decreasing



statista.com

GoodRelations

- E-commerce marketplace is suffering from all the deficits of the traditional web:
  - E-commerce websites are typically generated from structured information systems, listing price, availability, type of product, delivery options, etc., but by the time this information reaches the company's web pages, it has been turned into HTML and all machine-interpretable structure has disappeared, with the result that machines can no longer distinguish a price from a product-code
  - Search engines suffer from the inability to interpret the e-commerce pages that they try to crawl and index, and are unable to correctly distinguish product-types or to produce meaningful groupings of products

GoodRelations

- GoodRelations is an OWL-compliant ontology that describes the domain of electronic commerce

http://www.heppnetz.de/projects/goodrelations/

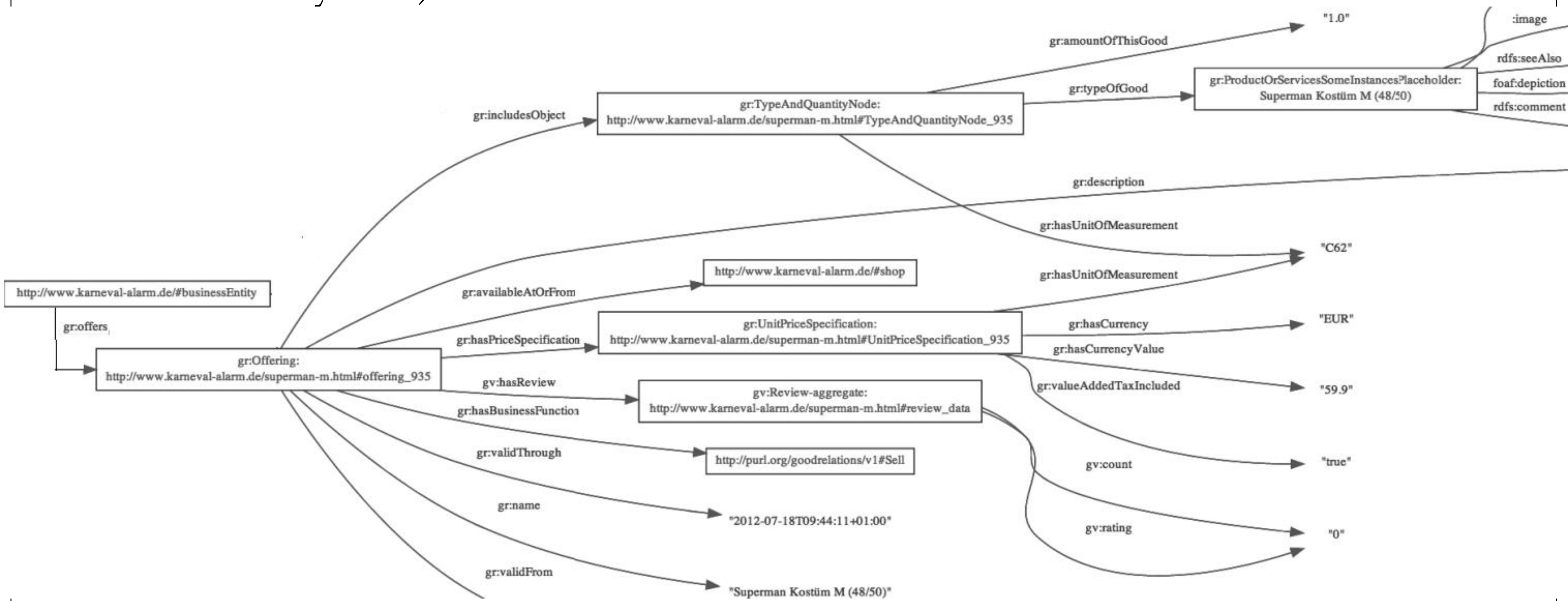
  - It can be used to express an offering of a product, to specify a price, to describe a business
  - The RDFa syntax for GoodRelations allows this information to be embedded into existing web pages so that they can be processed by other computers
  - The primary benefit of GoodRelations and the main driver behind its rapidly increasing adoption is how it improves search
    - Adding GoodRelations to webpages improves the visibility of offers in modern search engines and recommender systems
    - It allows for very specific search queries and gives very precise answers
    - Used by: Google, BestBuy, sears.com, kmart.com … and 10,000 more

GoodRelations

- The ontology also allows the expression of commercial and functional details of e-commerce transactions, such as:
  - eligible countries
  - payment and delivery options
  - quantity discounts
  - opening hours
  - The GoodRelations ontology http://www.heppnetz.de/ontologies/goodrelations/v1 contains classes such as **`gr:ProductOrServiceModel`**, **`gr:PriceSpecification`**, **`gr:OpeningHoursSpecification`**, **`gr:DeliveryChargeSpecification`**, with properties such as **`gr:typeOfGood`**, **`gr:acceptedPaymentMethods`**, **`gr:hasCurrency`**

GoodRelations

- Example: http://www.karneval-alarm.de/superman-m.html (Karneval Alarm shop, selling party costumes in Germany) describes a superman costume in size 48/50 for the price of 59.90 euros, product number 935 (represented as the RDF entity **`offering_935`** and, using the RDFa syntax)



GoodRelations

```
offering_935 gr:name "Superman Kostum 48/50" ;
  gr:availableAtOrFrom
    http://www.karneval-alarm.de/#shop ;
  gr:hasPriceSpecification UnitPriceSpecification_935 .
UnitPriceSpecification_935 gr:hasCurrency "EUR" ;
  gr:hasCurrencyValue "59.9" ;
  gr:valueAddedTaxIncluded "true" .
```

Product Ontology

- GoodRelation annotations need to mention product types and categories
  - The Product Ontology describes products that are derived from Wikipedia pages: http://www.productontology.org/

```
pto:Party_costume a owl:Class;
  rdfs:subClassOf gr:ProductOrService;
  rdfs:label "Party Costume"@en;
  rdfs:comment """A party costume is clothing..."""@en.
```

  - The Product Ontology contains several hundred thousand OWL DL class definitions of products
  - These class definitions are tightly coupled with Wikipedia: the edits in Wikipedia are reflected in the Product Ontology
    - if a supplier sells a product that is not listed in the Product Ontology, they can create a page for it in Wikipedia and the product type will appear within 24 hours in the Product Ontology

GoodRelations

- Adoption:
  - the first company to adopt the GoodRelations ontology on a large scale was BestBuy
    - BestBuy reported a 30 percent increase in search traffic for its GoodRelation-enhanced pages, and a significantly increased click-through rate
  - Google is now recommending the use of GoodRelations for semantic markup of e-commerce pages
  - Both Bing and Google are crawling RDFa statements and using them to enhance the presentation of their search results
  - Other adopters: Overstock.com retailers, the O'Reilly Media publishing house, the Peek & Cloppenburg clothing chain store, and Robinson Outdoors
  - Sindice semantic search engine lists 273,000 pages annotated with the GoodRelations vocabulary

BBC Artists

- BBC Music Beta project: http://www.bbc.co.uk/music/artists
  - Example John Lennon: http://www.bbc.co.uk/music/artists/4d5447d7-c61c-4120-ba1b-d7f471d385b9
  - It is an effort by the BBC to build semantically linked and annotated web pages about artists and singers whose songs are played on BBC radio stations
  - Collections of data are enhanced and interconnected with semantic metadata, letting music fans explore connections between artists that they may have not known existed
  - Previously, writers at the BBC would have to write (and keep up to date) interesting and relevant content on every single artist page they published
  - Instead, the BBC is pulling in information from external sites such as MusicBrainz and Wikipedia and aggregating this information to build their web pages
  - Web pages can be created and maintained with a fraction of the manpower required in the past

BBC Artists

- Example John Lennon: http://www.bbc.co.uk/music/artists/4d5447d7-c61c-4120-ba1b-d7f471d385b9
  - MusicBrainz RDF ID: **4d5447d7**

```
5c014631#artist foaf:name "John Lennon".
```

  - John Lennon died on December 8, 1980:

```
4d5447d7#artist bio:event _:node15vknin3hx2.
_:node15vknin3hx2 rdf:type bio:Death.
_:node15vknin3hx2 bio:date "1980-12-08".
```

  - he made a record entitled "John Lennon/Plastic Ono Band," and that he is also known under the URI dbpedia:John_Lennon:

```
4d5447d7#artist foaf:made _:node15vknin3hx7.
_:node15vknin3hx7 dc:title "John Lennon/Plastic
Ono Band".
4d5447d7#artist owl:sameAs dbpedia:John_Lennon.
```

  - The full content of the BBC Artist page on John Lennon contains 60 triples and 300 more can be inferred

BBC Artists

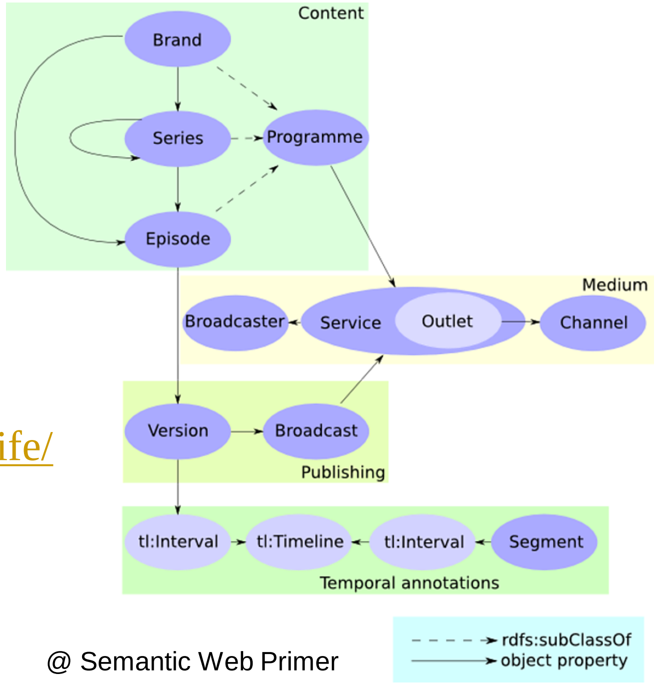
- The full site has in the order of 400,000 artist pages, 160,000 external links and 100,000 artist-to-artist relationships
- Use of Semantic Web technology: using URIs as identifiers and aligning these with external semantic data providers
- BBC is not only consuming information resources, but is also serving them back to the world by adding .rdf to the URI of any BBC Artist web page
- When using public information as input, there is always the risk of such information containing errors: BBC does not repair those errors internally, but, instead, they will repair the errors on the external sources such as MusicBrainz and DBPedia (this not only results in repairing the errors on the BBC site as well (since it is feeding of those sources), but it also repairs the error for any other user of these data sources, thereby contributing to increased quality of publicly available data sources)

BBC Artists

- Adoption:
  - The BBC Artist project is "*a part of a general movement that's going on at the BBC to move away from pages that are built in a variety of legacy content production systems to actually publishing data that we can use in a more dynamic way across the web.*"
  - http://www.bbc.co.uk/programmes
    - A vocabulary for program data: concepts such as brands, series, episodes, broadcasts

https://www.bbc.co.uk/ontologies/po

http://www.bbc.co.uk/ontologies/wildlife/



@ Semantic Web Primer

BBC World Cup

- The BBC website for the 2010 World Cup soccer event deployed semantic technologies in order to achieve more automatic content publishing, a higher number of pages manageable with a lower headcount, semantic navigation, and personalization
  - hundreds of players, dozens of teams, all groups, all matches
  - The BBC has developed small ontologies to capture the domain of soccer, including domain-specific notions concerning soccer teams and tournaments, as well as very generic notions for events and geographic locations, using well-known ontologies such as:
    - FOAF, a project devoted to linking people and information using the Web (social networks): http://xmlns.com/foaf/spec/
    - GeoNames geographical database covers all countries and contains over eleven million placenames: http://www.geonames.org
      - Stony Brook: http://www.geonames.org/maps/google_40.926_-73.141.html

BBC World Cup

- Adoption:
  - Three-tier Semantic Web architecture:
    - all information stored in an RDF triple-store
    - the triple store is organized by a number of ontologies that enable querying
    - a user interface layer that uses ontology-based queries to the triple store to obtain information to display to the user
  - BBC claims to have greatly increased their opportunities for content reuse and repurposing, reduced the journalist headcount required to maintain the site, and improved the user experience through semantically driven page-layout and multi-dimensional entry-points (player, match, group)
  - BBC Olympics 2012: 10,000 athletes from over 200 countries, real-time statistics and serving 58,000 hours of video content
  - BBC general sports ontology

Government Data

- Some of the early large-scale adoption of Semantic Web technologies in the world was a result of the Obama administration's drive towards more transparency of government
  - Publishing data sources that had traditionally been locked up inside governmental institutions and not available to the public was seen as a crucial step toward more transparent government
  - https://www.data.gov

Government Data

- 250,000 datasets from 250 different government agencies
  - economic indicators
  - health-service statistics
  - environmental data
- Full Linked Data format, exploiting RDF, URIs, ontologies, and linking to other datasets
- 5-star scale introduced by the W3C as a road toward publishing Linked Data that starts with much simpler technologies:
  - ❖put data on the web at all, in any format, anywhere
  - ❖use a machine readable format
  - ❖use an open format
  - ❖give a URL for each data item
  - ❖link out to shared vocabularies

Government Data

- 5-star scale by the W3C:
  - put data on the web at all, in any format, anywhere
    - just putting PDF documents with tables (or even scanned documents) on a website somewhere would already rate as a 1-star step in publishing data
    - Although technically trivial, in practice this already means overcoming social, organizational, and bureaucratic hurdles
  - use a machine readable format
    - avoiding formats such as PDF docs
    - Excel spreadsheets is the most typical example of a 2-star data-publishing step
  - use an open format
    - avoiding propriety formats such as Excel, and instead using open formats like .csv file
  - give a URL for each data item
    - this allows others to link to the published dataset
  - link out to shared vocabularies
    - use the DBPedia term for city
    - This use of external vocabularies truly enables the interlinking of datasets across different sites

Government Data

- Uses:
  - comparing agency budgets across three public budget datasets
  - plotting social networks of people visiting the White House
  - plotting postal service expenditure
  - plotting wilderness fires
  - covering interstate migration from information on tax forms
  - plotting family income against Medicare claims
- Although impossible to establish a causal link, the number of citizen appeals under the Freedom of Information Act have reportedly dropped substantially since the launch of http://data.gov
- Over 200 applications built by third parties have been reported, including: a safety map of bicycle routes, information for home buyers about their new neighborhoods, a school finder, a nursery finder, pollution alerts, a regional expenditure map, and WhereDidMyTaxGo

Government Data

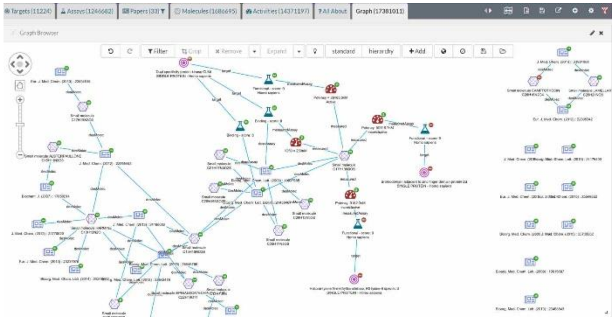
- Encouraged by this success, other governments have started to follow this example, including the governments of many states and large cities in the US, countries like Canada, Ireland, Norway, Australia, and New Zealand, and citizen-initiatives have started in countries such as France, Italy, Denmark, Austria, Germany
- UK: http://data.gov.uk is seen as not only a step towards more transparent government, but also a cost-reduction mechanism
  - Instead of building and maintaining expensive websites with governmental information, the government simply publishes the underlying data sources and encourages third parties (be it either citizens or commercial parties) to develop services on top of these published data sources.
- Intergovernmental agencies are following the same path, such as the World Bank http://data.worldbank.org , the European Commission's website for European tendering procedures http://ted.europa.eu , the European statistics office Eurostat http://ec.europa.eu/eurostat

New York Times

- Since 1913, the New York Times has been maintaining an index of all subjects on which it has published
  - grown into a collection of almost 30,000 “subject headings,” describing locations, people, organizations, and events
  - Every article in the New York Times since its first appearance in 1851 has been tagged with these subject headings, and such tags are being used to provide news-alerting services, to automate the editorial process, and to build “topic pages” that collect all info for a given topic
  - In 2009, the New York Times started to convert its entire subject headings index to Semantic Web format
  - The terms are also being linked to hubs on the Linked Data Web such as DBPedia, Freebase, and GeoNames
    - helped them to provide geographic information
    - align articles with other data sources used by the newspaper, such as the Library of Congress

Sig.ma and Sindice

- Previous search engine on the RDF graph that maked up the global Semantic Web
  - classical search engine architecture: a crawler, a very large indexed store, and a retrieval interface
  - Sindice did not retrieve and index words and phrases, but instead retrieved and indexed RDF graphs
  - 400 million RDF "documents," resulting in 12 billion RDF statements
  - available for querying through a SPARQL endpoint
- With the launch in 2012 of Schema.org, Google and others have effectively embraced the vision of the "Semantic Web"
- Sindice became commercial https://siren.io and provides services in finance, medicine, insurance



Swoogle

- http://swoogle.umbc.edu

Swoogle
semantic web search

ontology **document** **term** **more >>**

owl

Swoogle Search

list ontologies matching ontology search

http://www.w3.org/2002/07/owl
[DESC] title - The **OWL** 2 Schema vocabulary (**OWL** 2) | comment - This ontology partially describes the built
SemanticWebDocument, RDFXML, 2009-11-15, 31K, ontoRatio(1.00), metadata, cached

http://inferenceweb.stanford.edu/2004/07/iw.owl
SemanticWebDocument, RDFXML, 2009-04-22, 34K, ontoRatio(0.99), metadata, cached

http://spire.umbc.edu/ontologies/ethan_keywords.owl
SemanticWebDocument, RDFXML, 2006-10-19, 57K, ontoRatio(1.00), metadata, cached

http://morpheus.cs.umbc.edu/aks1/ontosem.owl
SemanticWebDocument, RDFXML, 2005-08-25, 3M, ontoRatio(1.00), metadata, cached

http://lod.taxonconcept.org/ontology/txn.owl
SemanticWebDocument, RDFXML, 2013-05-19, 165K, ontoRatio(0.98), metadata, cached

http://inference-web.org/2.0/pml-justification.owl
SemanticWebDocument, RDFXML, 2008-05-08, 20K, ontoRatio(1.00), metadata, cached

http://inference-web.org/2.0/pml-provenance.owl
SemanticWebDocument, RDFXML, 2007-06-28, 49K, ontoRatio(1.00), metadata, cached

OpenCalais

- Thompson Reuters: http://www.opencalais.com
  - The press agency Thompson Reuters produces thousands of news items every day, covering wide areas of science, business, politics, and sports.
  - Although originally intended to be read by humans, many of the consumers of their news items are no longer people, but instead computers reading the news flowing out of Reuters, analyzing it and producing summaries, reading lists, financial investment advice
    - Computers producing information for computers
  - Using unstructured natural language is unsuited communication medium
  - Reuter's OpenCalais service is intended to help computers process data captured in natural language texts by recognizing named entities such as people, locations, companies, etc., and annotating the text with RDF to identify these named entities
  - Processes five million documents per day

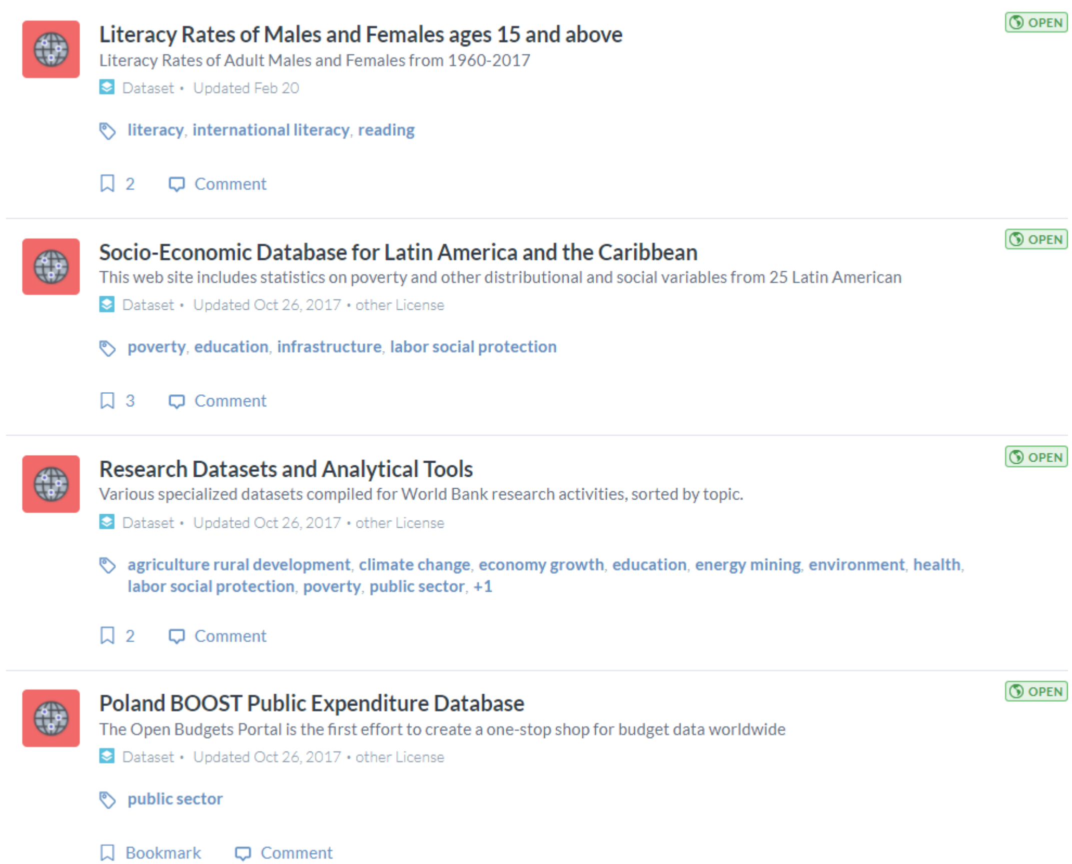
Schema.org

- http://schema.org
  - Effort by search engine providers (Google, Yahoo, Microsoft, Yandex) to encourage the semantic mark up of pages
  - A common schema or vocabulary that covers common things that people search for, such as products, jobs, and events
    - By using a common vocabulary, search engines are able to better index pages but also show richer information in search results.
  - Getting started:

```
<div itemscope itemtype ="http://schema.org/Movie">
  <h1 itemprop="name">Avatar</h1>
  <div itemprop="director" itemscope
    itemtype="http://schema.org/Person">
  Director: <span itemprop="name">James Cameron</span>
    (born <span itemprop="birthDate">August 16, 1954</span>)
  </div>
```

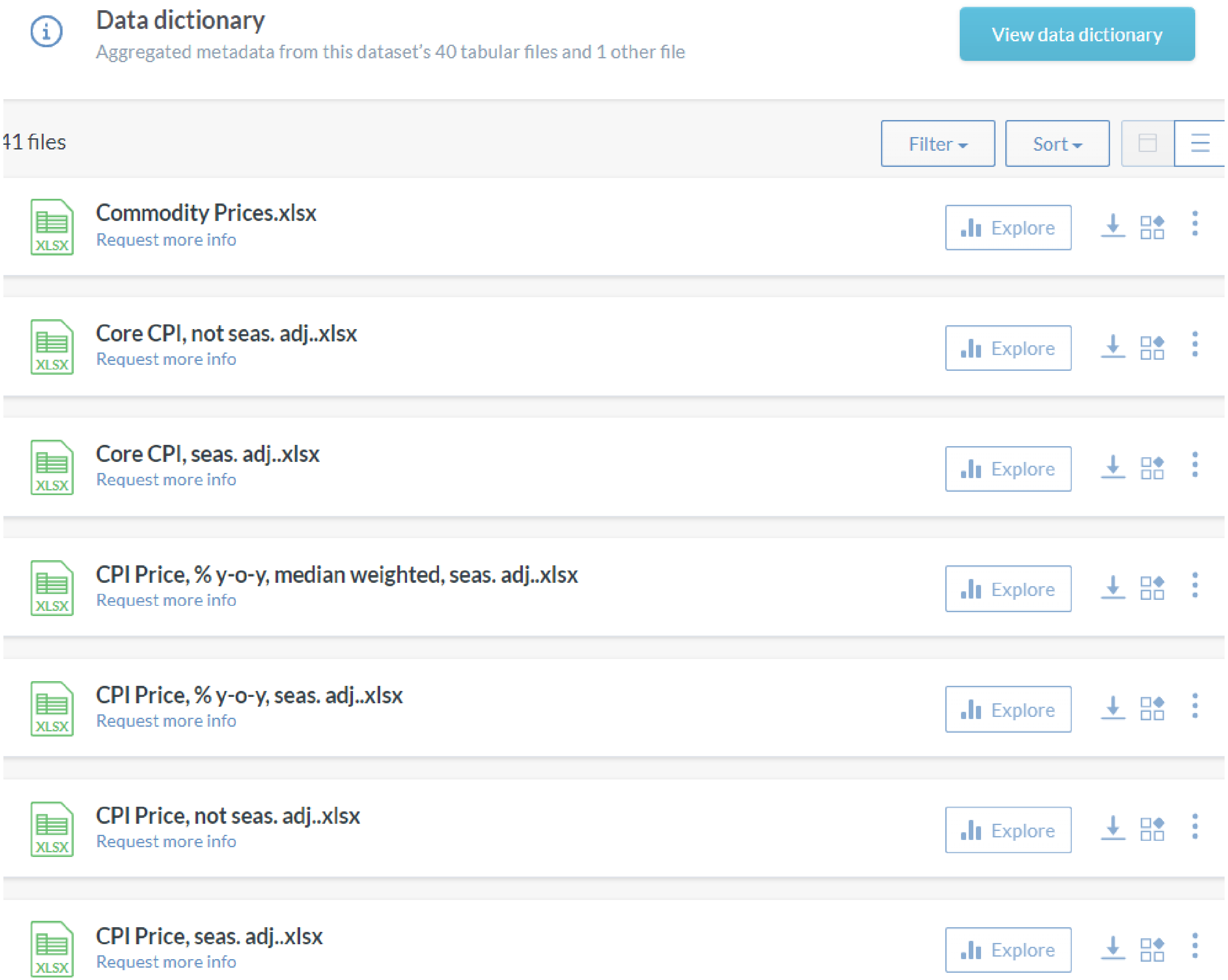
data.world

- https://data.world
  - Podcast: https://www.dataengineeringpodcast.com/wp-content/uploads/Episode-9-Data.world_.mp3
  - https://data.world/worldbank



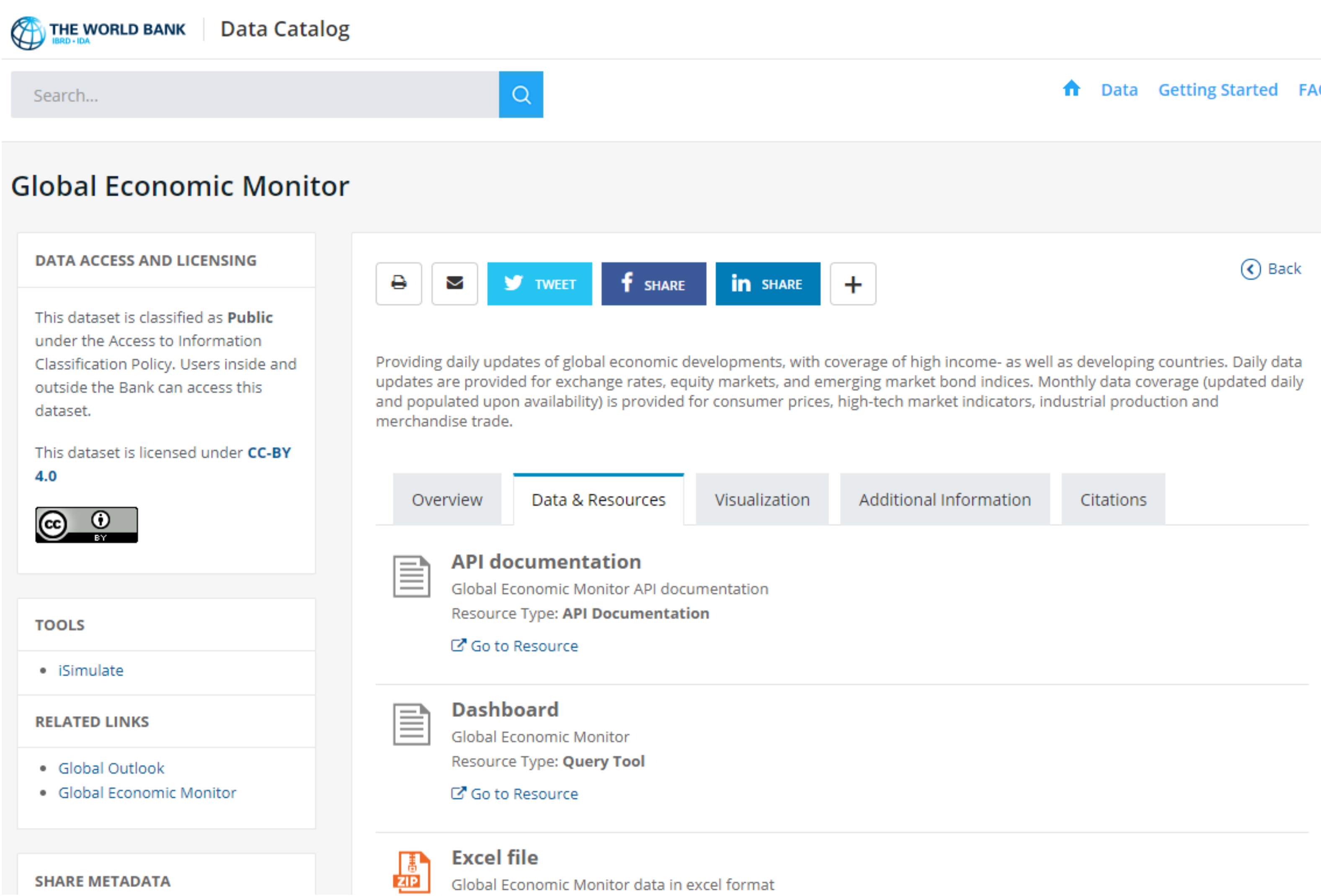
data.world

- https://data.world/worldbank/global-economic-monitor



data.world

- https://datacatalog.worldbank.org/dataset/global-economic-monitor



Elsevier

- Elsevier is a leading scientific publisher
  - Its products are organized mainly along traditional lines: subscriptions to **specific** journals
    - Customers of Elsevier can take subscriptions to specific online content
    - Online availability of these journals has until now not really changed the organization of the product line
  - Traditional journals are ***vertical products***: services **specific** to an industry, a trade, a profession, or a group of customers with **specialized** needs
    - Division into separate sciences covered by distinct journals is no longer satisfactory
    - Customers of Elsevier are interested in covering certain topic areas that spread across the traditional disciplines/journals
    - The demand is for ***horizontal products***

Elsevier

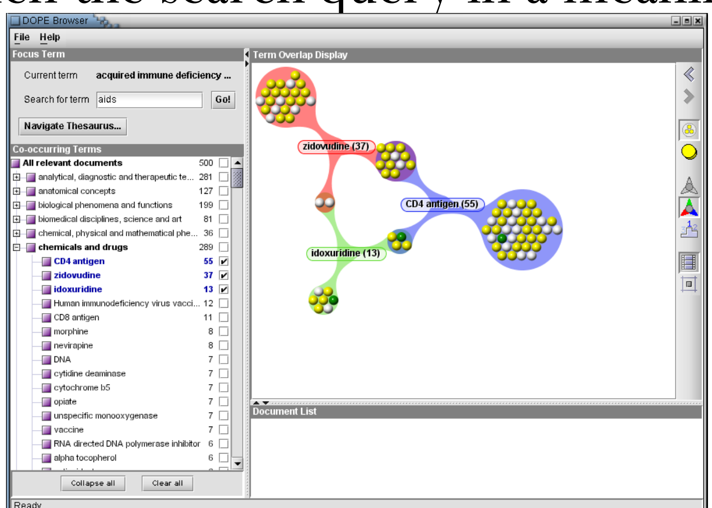
- It is difficult for large publishers to offer horizontal products
  - need a way to search the journals on a **<u>coherent</u>** set of concepts against which all of these journals are indexed
- Ontologies and thesauri (very lightweight ontologies) have proved to be a key technology for effective information access
  - They help to overcome some of the problems of free-text search
  - They relate and group relevant terms in a specific domain
  - They provide a controlled vocabulary for indexing information
- A number of thesauri have been developed in different domains of expertise
  - Medical information MeSH https://www.nlm.nih.gov/mesh
  - Elsevier's life science thesaurus EMTREE https://www.elsevier.com/solutions/embase-biomedical-research

Elsevier

- RDF is used as an interoperability format between heterogeneous data sources
- EMTREE is itself represented in RDF
- Each of the separate data sources is mapped onto this unifying ontology
  - The ontology is then used as the single point of entry for all of these data sources

Elsevier

- Elsevier has sponsored the [DOPE project (Drug Ontology Project for Elsevier)]
  - The EMTREE thesaurus was used to index millions of medical abstracts and full text articles
  - In the interface used, the EMTREE ontology was used to:
    - disambiguate the original free-text user query
    - categorize the results
    - produce a visual clustering of the search results
    - narrow or widen the search query in a meaningful way



Audi – The Problem and Solution

- Data integration is also a huge problem internal to companies
  - It is the highest cost factor in the information technology budget of large companies
- Traditional middleware improves and simplifies the integration process
  - But it misses the sharing of information based on the semantics of the data
- Audi operates thousands of databases
  - Ontologies can rationalize disparate data sources into one body of information
    - Without disturbing existing applications, by:
      - creating common general ontologies for data and content sources
      - adding generic domain information

Audi – Example

```
<SLR rdf:ID="OM-10">
     <viewFinder>twin mirror</viewFinder>
     <optics>
            <Lens>
                   <focal-length>75-300mm zoom</focal-length>
                   <f-stop>4.0-4.5</f-stop>
            </Lens>
     </optics>
     <shutter-speed>1/2000 sec. to 10 sec.</shutter-speed>
</SLR>

<Camera rdf:ID="OM-10">
     <viewFinder>twin mirror</viewFinder>
     <optics>
            <Lens>
                   <size>300mm zoom</size>
                   <aperture>4.5</aperture>
            </Lens>
     </optics>
     <shutter-speed>1/2000 sec. to 10 sec.</shutter-speed>
</Camera>
```

Audi – Example

- Human readers can see that these two different formats talk about the same object
  - We know that SLR is a kind of camera, and that f-stop is a synonym for aperture
- Ad hoc integration of these data sources by translator is possible
  - Would only solve this specific integration problem
- We would have to do the same again when we encountered the next data format for cameras

Audi – Camera Ontology in OWL

```
<owl:Class rdf:ID="SLR">
     <rdfs:subClassOf rdf:resource="#Camera"/>
</owl:Class>

<owl:DatatypeProperty rdf:ID="f-stop">
     <rdfs:domain rdf:resource="#Lens"/>
</owl:DatatypeProperty>

<owl:DatatypeProperty rdf:ID="aperture">
     <owl:equivalentProperty rdf:resource="#f-stop"/>
</owl:DatatypeProperty>

<owl:DatatypeProperty rdf:ID="focal-length">
     <rdfs:domain rdf:resource="#Lens"/>
</owl:DatatypeProperty>

<owl:DatatypeProperty rdf:ID="size">
     <owl:equivalentProperty rdf:resource="#focal-length"/>
</owl:DatatypeProperty>
```

Audi – Using the Ontology

- Suppose that an application A
  - is using the second encoding, and
  - is receiving data from an application B using the first encoding
- Suppose it encounters SLR
  - The ontology returns "SLR is a type of Camera"
  - A relation between something it doesn't know (SLR) to something it does know (Camera)
- Suppose A encounters f-stop
  - The Ontology returns: "f-stop is synonymous with aperture"
- Bridges the terminology gap between something A doesn't know to something A does know
- Syntactic divergence is no longer a hindrance

Swiss Life – The Setting

- Swiss Life is one of Europe's leading life insurers
  - 11,000 employees, $14 billion of written premiums
  - Active in about 50 different countries
- The most important resources of any company for solving knowledge intensive tasks are:
  - The tacit knowledge, personal competencies, and skills of its employees
- One of the major building blocks of enterprise knowledge management is:
  - An electronically accessible repository of people's capabilities, experiences, and key knowledge areas

Swiss Life – The Setting

- A skills repository can be used to:
  - enable a search for people with specific skills
  - expose skill gaps and competency levels
  - direct training as part of career planning
  - document the company's intellectual capital
- Problems
  - How to list the large number of different skills?
  - How to organize them so that they can be retrieved across geographical and cultural boundaries?
  - How to ensure that the repository is updated frequently?

Swiss Life – The Solution

- Hand-built ontology to cover skills in three organizational units
  - Information Technology
  - Private Insurance
  - Human Resources
- Individual employees within Swiss Life were asked to create "home pages" based on form filling driven by the skills-ontology
- The corresponding collection could be queried using a form-based interface that generated RQL queries

Swiss Life – Skills Ontology

```
<owl:Class rdf:ID="Skills">
     <rdfs:subClassOf>
     <owl:Restriction>
            <owl:onProperty rdf:resource="#HasSkillsLevel"/>
            <owl:cardinality
                    rdf:datatype="&xsd;nonNegativeInteger">
                   1
            </owl:cardinality>
            </owl:Restriction>
     </rdfs:subClassOf>
</owl:Class>

<owl:ObjectProperty rdf:ID="HasSkills">
     <rdfs:domain rdf:resource="#Employee"/>
     <rdfs:range rdf:resource="#Skills"/>
</owl:ObjectProperty>
```

Swiss Life – Skills Ontology

```
<owl:ObjectProperty rdf:ID="WorksInProject">
    <rdfs:domain rdf:resource="#Employee"/>
    <rdfs:range rdf:resource="#Project"/>
    <owl:inverseOf rdf:resource="#ProjectMembers"/>
</owl:ObjectProperty>

<owl:Class rdf:ID="Publishing">
    <rdfs:subClassOf rdf:resource="#Skills"/>
</owl:Class>

<owl:Class rdf:ID="DocumentProcessing">
    <rdfs:subClassOf rdf:resource="#Skills"/>
</owl:Class>
```

Swiss Life – Skills Ontology

```
<owl:ObjectProperty rdf:ID="ManagementLevel">
    <rdfs:domain rdf:resource="#Employee"/>
    <rdfs:range>
        <owl:oneOf rdf:parseType="Collection">
            <owl:Thing rdf:about="#member"/>
            <owl:Thing rdf:about="#HeadOfGroup"/>
            <owl:Thing rdf:about="#HeadOfDept"/>
            <owl:Thing rdf:about="#CEO"/>
        </owl:oneOf>
    </rdfs:range>
</owl:ObjectProperty>
```

EnerSearch – The Setting

- An industrial research consortium focused on information technology in energy
- EnerSearch has a structure very different from a traditional research company
  - Research projects are carried out by a varied and changing group of researchers spread over different countries
  - Many of them are not employees of EnerSearch
- EnerSearch is organized as a virtual organization
  - Owned by a number of firms in the industry sector that have an express interest in the research being carried out
  - Because of this wide geographical spread, EnerSearch also has the character of a virtual organization from a knowledge distribution point of view

EnerSearch

- Dissemination of knowledge key function
  - Does load management lead to cost-saving?
  - If so, what are the required upfront investments?
  - Can powerline communication be technically competitive to ADSL or cable modems?
- It is possible to define a domain ontology that is sufficiently stable and of good quality
  - This lightweight ontology consisted only of a taxonomical hierarchy
  - Needed only RDF Schema expressivity
- Use of Ontology:
  - Used in a number of different ways to drive navigation tools on the EnerSearch web site

EnerSearch - QuizRDF

- QuizRDF combined:
  - an entirely ontology based display
  - keyword based search without any semantic grounding
    - The user can type in general keywords
  - It displays those concepts in the hierarchy which describe the complex projects
- All the disclosure mechanisms (textual and graphic, searching or browsing) based on a single underlying lightweight ontology

E-Learning – The Setting

- Traditionally learning has been characterized by the following properties:
  - Educator-driven
  - Linear access
  - Time- and locality-dependent
  - Learning has not been personalized but rather aimed at mass participation
- The changes are already visible in higher education
  - Virtual universities
  - Flexibility and new educational means
  - Students can increasingly make choices about pace of learning, content, evaluation methods

E-Learning – The Setting

- Even greater promise: life long learning activities
  - Improvement of the skills of employees is critical to companies
  - Organizations require learning processes that are just-in-time, tailored to their specific needs
  - These requirements are not compatible with traditional learning, but e-learning shows great promise for addressing these concerns
- E-learning is not driven by the instructor
- Learners can:
  - Access material in an order that is not predefined
  - Compose individual courses by selecting educational material
- Learning material must be equipped with additional information (metadata) to support effective indexing and retrieval

E-Learning – The Setting

- Standards (IEEE LOM) have emerged
  - educational and pedagogical properties, access rights and conditions of use, and relations to other educational resources
- Standards suffer from lack of semantics
  - This is common to all solutions based solely on metadata (XML-like approaches)
  - Combining of materials by different authors may be difficult
  - Retrieval may not be optimally supported
  - Retrieval and organization of learning resources must be made manually
  - Could be done by a personalized automated agent instead!

E-Learning – Contribution of Semantic Web Technology

- Establish a promising approach for satisfying the elearning requirements
  - E.g. ontology and machine-processable metadata
- Learner-centric
  - Learning materials, possibly by different authors, can be linked to commonly agreed ontologies
  - Personalized courses can be designed through semantic querying
  - Learning materials can be retrieved in the context of actual problems, as decided by the learner

E-Learning – Contribution of Semantic Web Technology

- Flexible access
  - Knowledge can be accessed in any order the learner wishes
  - Appropriate semantic annotation will still define prerequisites
  - Nonlinear access will be supported
- Integration
  - A uniform platform for the business processes of organizations
  - Learning activities can be integrated in these processes

Ontologies for E-Learning

- Some mechanism for establishing a shared understanding is needed: ontologies
- In e-learning we distinguish between three types of knowledge (ontologies):
  - Content
  - Pedagogy
  - Structure
- Content Ontologies:
  - Basic concepts of the domain in which learning takes place
  - Include the relations between concepts, and basic properties
    - E.g., the study of Classical Athens is part of the history of Ancient Greece, which in turn is part of Ancient History
    - The ontology should include the relation "is part of" and the fact that it is transitive (e.g., expressed in OWL)

Ontologies for E-Learning

- Pedagogy Ontologies
  - Pedagogical issues can be addressed in a pedagogy ontology (PO)
  - E.g. material can be classified as lecture, tutorial, example, walk-through, exercise, solution, etc.
- Structure Ontologies
  - Define the logical structure of the learning materials
  - Typical knowledge of this kind includes hierarchical and navigational relations like previous, next, hasPart, isPartOf, requires, and isBasedOn
  - Relationships between these relations can also be defined
    - E.g., hasPart and isPartOf are inverse relations
  - Inferences drawn from learning ontologies cannot be very deep

Web Services

- Web sites that do not merely provide static information, but involve interaction with users and often allow users to effect some action
- Simple Web services involve a single Web-accessible program, sensor, device
- ***Complex Web services*** are <u>composed</u> of simpler services
  - Often they require ongoing interaction with the user
  - The user can make choices or provide information conditionally

Web Services

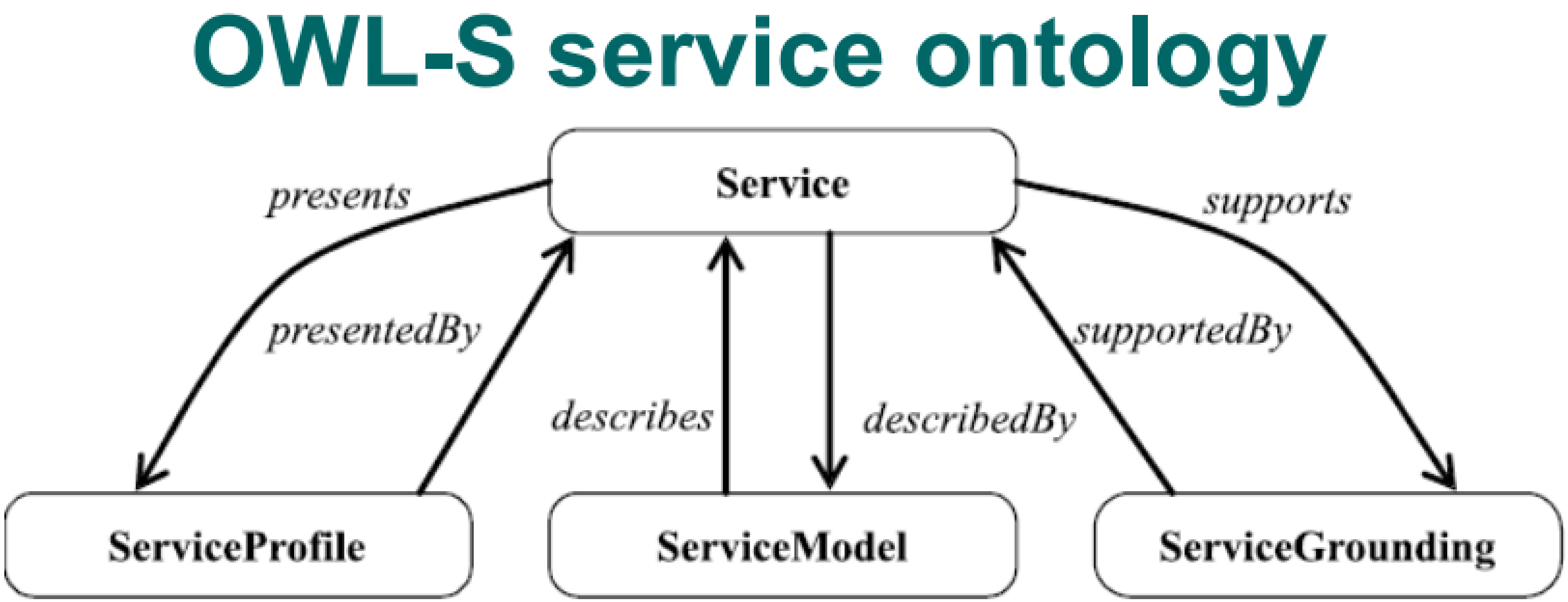
- A Complex Web Service: User interaction with medical supliers
  - searching for medical suppliers by various criteria (location, category, insurance taken)
  - reading reviews
  - providing insurance, address and credit card details
- SOAP, WSDL, UDDI and BPEL4WS are the standard technology combination to build a Web service application
- They fail to achieve the goals of automation and interoperability because the require humans in the loop
- WSDL specifies the functionality of a service only at a syntactic level but does not describe the meaning of the Web service functionality

Web Services

- The Semantic Web community addressed the limitations of current Web service technology by augmenting the service descriptions with a semantic layer in order to achieve
  - Automatic discovery, composition, monitoring, and execution
- The automation of these tasks is highly desirable
- The example task is specializing the more generic task of finding the closest medical provides
- A strategy for performing this task is
  - Retrieve the details of all medical providers
  - Select the closest by computing the distance between the location of the provider and a reference location

Web Services

- Semantic Web service technology aims to automate performing such tasks based on the semantic description of Web services
- https://www.w3.org/Submission/OWL-S/



Generic Web Service Ontologies OWL-S

- A common characteristic of all emerging frameworks for semantic Web service descriptions is the they combine two kinds of ontologies to obtain a service description
  - A generic Web service ontology
  - A domain ontology

Generic Web Service Ontologies OWL-S

- OWL-S ontology is conceptually divided into four subontologies for specifying
  - What the service does (Profile)
    - This information is primary meant for human reading, and includes the service name and description, limitations on applicability and quality of service, publisher and contact information
  - How the service works (Process)
    - The process model describes how a client can interact with the service
    - This description includes the sets of inputs, outputs, pre-conditions and results of the service execution
  - How the service is implemented (Grounding)
    - The service grounding specifies the details that a client needs to interact with the service, as communication protocols, message formats, port numbers, etc.
  - A fourth ontology (Service) contains the Service concept, which links together the ServiceProfile, ServiceModel and ServiceGrounding

The Profile Ontology

- Profile specifies:
  - The functionality offered by the service
  - The semantic type of the inputs and outputs
  - The details of the service provider
  - Several service parameters, such as quality rating or geographic radius
- Profile is a subclass of ServiceProfile
- For each Profile instance we associate
  - the process it describes
  - its functional characteristics together with their type

The Profile Ontology Example

```
Service MedicareSupplier:
  *Profile : FindMedicareSupplierByZip (hasProc P1)
      (I (ZipCode), O (SupplierDetails))
  *Profile : FindMedicareSupplierByCity (hasProc P2)
      (I (City), O (SupplierDetails))
  *Profile : FindMedicareSupplierBySupply (hasProc P3)
      (I (SupplyType), O (SupplierDetails))
  *ProcessModel : …
  *WSDLGrounding : …
```

The Process Ontology

- Many complex services consist of smaller executed in a certain **order**
- For example, buying a book at Amazon.com involves using a **browsing service** and a **paying service**
- OWL-S allows describing such internal process models
- These are useful for several purposes:
  - One can **check** that the business process of the offering service is appropriate
  - One can **monitor** the execution stage of a service
  - These process models can be used to automatically **compose** Web services

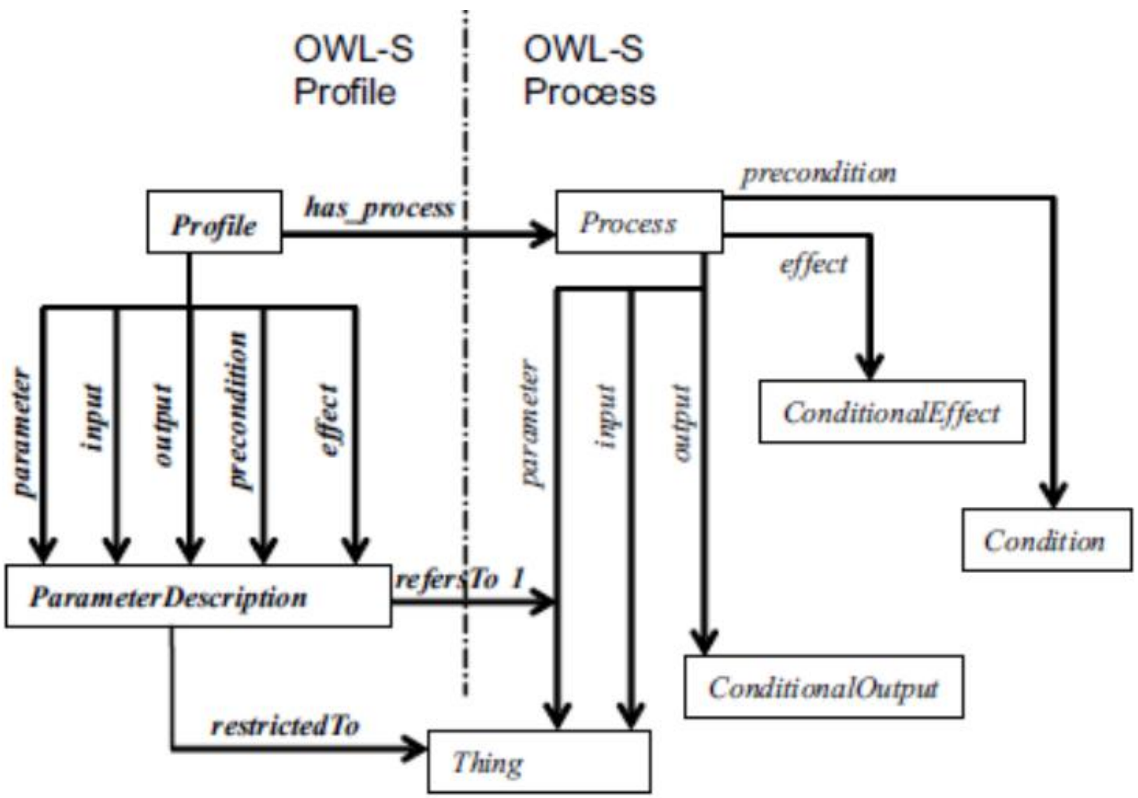
The Process Ontology Example

```
Service MedicareSupplier :
  *Profile : …
  *ProcessModel : …
  CompositeProcess : MedicareProcess : Choice
    AtomicProcess : P1 (I (ZipCode), O
      (SupplierDetails))
    AtomicProcess : P2 (I (City), O (SupplierDetails))
    AtomicProcess : P3 (I (SupplyType), O
      (SupplierDetails))
  *WSDLGrounding : …
```

One can search for medical suppliers using the zip code OR city OR SupplyType

Profile to Process Bridge

- A Profile contains several links to a Process
- Profile states the Process it describes through the unique property ***has_process***
- The ***input, output, precondition, and effects*** (***IOPEs***) of the Profiles correspond to the IOPEs of the Process



The Grounding ontology

- The grounding to a WSDL description is performed according to three rules:
  - Each AtomicProcess corresponds to one WSDL operation
  - Each input of an AtomicProcess is mapped to a corresponding message-part in the input message of the WSDL operation
    - Similarly for outputs
  - The type of each WSDL message part can be specified in terms of a OWL-S parameter
- Example:

```
Service MedicareSupplier :
   *Profile : …
   *ProcessModel : …
   *WSDLGrounding:
     WsdlAtomicProcessGrounding : Gr1 (P1>op:GetSupplierByZipCode)
     WsdlAtomicProcessGrounding : Gr2 (P1->op:GetSupplierByCity)
     WsdlAtomicProcessGrounding : Gr3 (P1->op:GetSupplierBySupplyType)
```

Design Principles of OWL-S

- Semantic versus Syntactic descriptions
  - OWL-S distinguishes between the semantic and syntactic aspects of the described entity
  - The Profile and Process ontologies allow for a semantic description of the Web service, and the WSDL description encodes its syntactic aspects
  - The Grounding ontology provides a mapping between the semantic and the syntactic parts of a description facilitating flexible association between them

Design Principles of OWL-S

- Generic versus domain knowledge
  - OWL-S offers a core set of primitives to specify the type of Web service
  - These descriptions can be enriched with domain knowledge specified in a separate domain ontology
  - This modeling choice allows using the core set of primitives across several domains

Design Principles of OWL-S

- ***Modularity***:
  - A feature of OWL-S is the partitioning of the description over several concepts
  - There are several advantages of this modular modeling:
    - It is easy to reuse certain parts
    - Service specification becomes flexible because if is possible to specify only the part that is relevant for the service
    - Any OWL-S description is easy to extend by specializing the OWL-S concepts

Design Principles of OWL-S

- ***Externally defined knowledge***
  - OWL-S offers a generic framework to describe a service, but to make it truly useful, domain knowledge is required
  - All Web services in a domain should use concepts from the same domain ontology in their descriptions
  - The right services for the task can be selected automatically from a collection of services
  - Semantic metadata allow a flexible selection that can retrieve services that partially match a request

Design Principles of OWL-S

- A service that finds details of medical suppliers will be considered a match for a request for services that retrieve details of Medicare suppliers, if the Web service domain ontology specifies that a MedicareSupplier is a type of MedicalSupplier
- This matchmaking is superior to the keyword-based search offered by UDDI (Universal Description, Discovery, and Integration, an XML-based standard for describing, publishing, and finding web services)

Design Principles of OWL-S

- The ***composition*** of several services into a more complex service can also be automated
  - After being discovered and composed based on their semantic descriptions, the services can be invoked to solve the task at hand

Multimedia Collection Indexing at Scotland Yard

- Investigation of theft of art and antique objects requires an **international databases of stolen art objects**
  - Different parties are likely to offer different descriptions
  - It is difficult to locate specific objects in these databases
  - Human experts are needed to match objects to database entries
- Developed controlled vocabularies such as the Art and Architecture Thesaurus (AAT) from the Getty Trust, or Iconclass thesaurus
  - Scotland Yard extended them into full-blown ontologies
    - Developed automatic classifiers using ontological background knowledge
    - Dealed with the ontology-mapping problem

Online Procurement at Daimler

- Static, long-term agreements with a fixed set of suppliers can be replaced by dynamic, short-term agreements in a competitive open marketplace
- Whenever a supplier is offering a better deal, Daimler-Chrysler wanted to be able to switch
- Major driver behind B2B e-commerce
- RosettaNet is an organization dedicated to such standardization efforts
  - XML-based, no semantics
- Daimler-Chrysler used RDF Schema and OWL instead
  - Product descriptions would "carry their semantics"
    - Much more liberal online B2B procurement processes would exist than currently possible

Device Interoperability at Nokia

- Explosive proliferation of digital devices:
  - Smartphones, digital cameras, laptops, wireless access in public places, GPS-enabled cars
- **Interoperability among these devices?**
  - The pervasiveness and the wireless nature of these devices require network architectures to support automatic, ad hoc configuration
  - A key technology of true ad hoc networks is service discovery
- Service discovery and capability description require a priori identification of what to communicate or discuss

Device Interoperability at Nokia

- A more attractive approach would be "serendipitous interoperability"
  - Interoperability under "unchoreographed" conditions
  - Devices necessarily designed to work together
- These devices should be able to:
  - Discover each others' functionality
  - Take advantage of it
- Devices must be able to "understand" other devices and reason about their functionality
- Ontologies are required to make such "unchoreographed" understanding of functionalities possible

Publication Management

- Information about scientific publications is often maintained by individual researchers
  - Reference management software such as EndNote and BibTeX helps researchers to maintain personal collections of bibliographic references
  - Most researchers have to maintain a Web page about publications for interested peers from other institutes
  - Often personal reference management and the maintenance of Web pages are isolated efforts
  - The author of a new publication adds the reference to his own collection and updates his Web page

Publication Management

- Maintaining personal references and Web pages about publications should not require redundant efforts
  - One can achieve this by directly using individual bibliographical records: generate personal Web pages and create joined publication lists for Web pages at the group or institutional level
- The Open Access movement has changed the world of publishing
  - More and more academic societies are now turning to open source technologies to manage digital publishing
  - Traditional publishing houses have a hard time keeping up
- Several problems need to be solved:
  - Duplicate information should be detected and merged
  - It should be possible to query for specific selections of the bibliographic entries and represent them in customized layouts