http://www.slideshare.net/nimonika/semantic-web-in-cultural-heritage-and-archaeology

Semantic Web in Cultural Heritage and Archaeology

Monika Solanki

Department of Computer Science
University of Leicester

November 13, 2009

Outline

Philosophy in discussion with a Philosopher

- Introduction and Preliminaries
- Semantic Web meets Cultural Heritage
- Semantic Web meets Archaeology
- Questions and Discussion

http://www.kb.nl/manuscripts/

Part I

Setting the Scene

Introduction and Preliminaries

Semantic Web: The Vision

- An extension of the current web.
- A Web of machine - readable and linked data.
- Allows software agents to carry out complex tasks on behalf of the human users, e.g. Search.
- Through systems that can support trusted interactions over the network.
- The key: semantic interoperability of Web resources.

`http://www.w3.org/standards/semanticweb/`

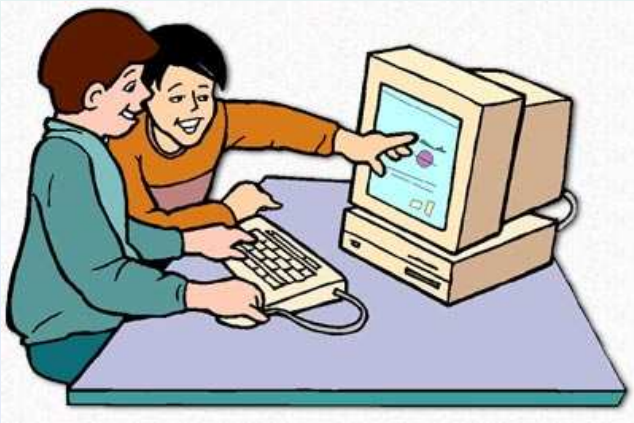
Semantic Web: A“Heritage” Perspective

- Heritage communities: new ways to provide domain experts (historians, curators) and non expert users (school children) with access to their collections.
- Online digital libraries of collections and exhibitions.
- Navigating, searching and retrieving 2D images, 3D models and textual metadata
- Virtual Museum
- Book your next Museum trip!!

Two different kinds of users

"Which museum holds artifacts and records of study of the breakdown and fossilization of bone from mummies found in the bronze age"

"For our history project, we want to visit a museum to study how life was in ancient Egypt"



Semantic Web: A “Heritage” Perspective

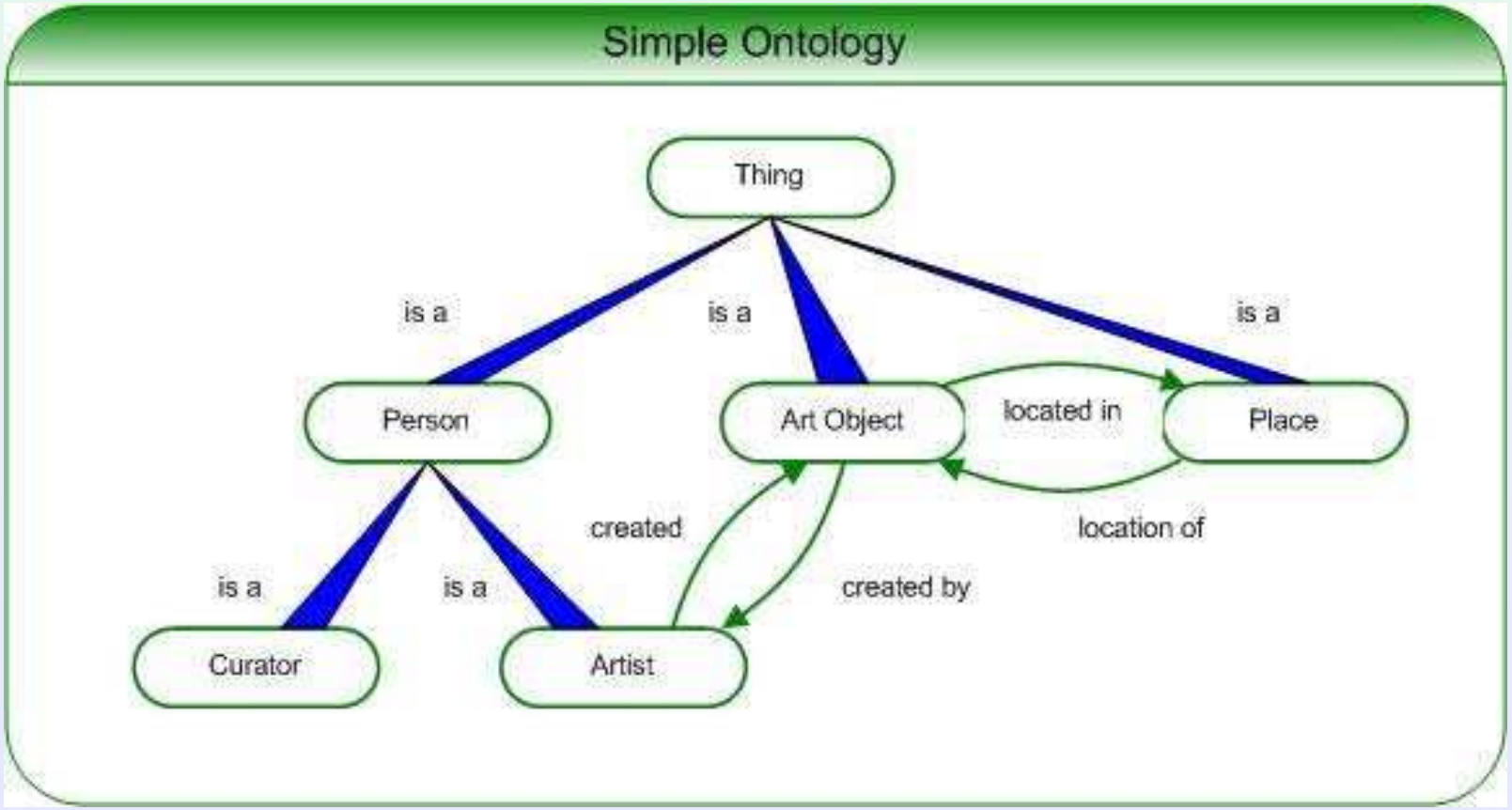
Benefits:

- Curators:
  - Along with simple description - metadata, relationships between the objects in the collection will be available.
  - Planning of their exhibition: identification of the pieces, production of catalogue and publicity material - handled by “exhibition agents”.
- Non experts:
  - Enable “meaningful” searches over online digital libraries of collections and exhibitions.
  - Planning your museum trip by relating opening times against public transport schedules.

Ontologies/Vocabularies/Thesauri

Ontologies

- An ontology is an explicit specification of a domain in terms of entities and relationship between these entities.

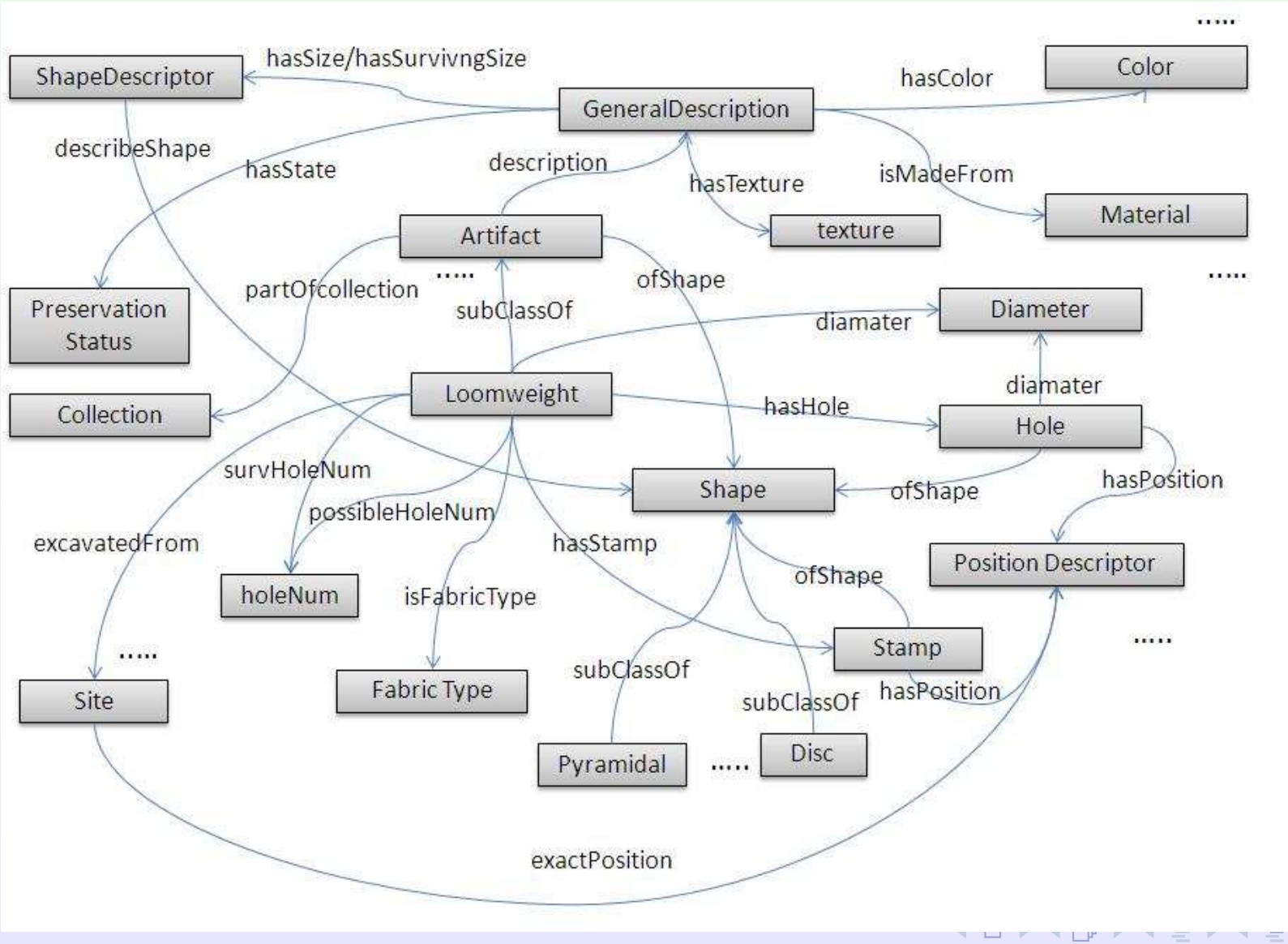


Ontologies/Vocabularies/Thesauri

Ontologies

- The relationship between entities provide the semantics or meaning of the data.
- This information is normally hidden in a dataset, e.g., databases, but it is made more explicit in an ontology.
- This explicit specification facilitates querying, complex reasoning and extension of the knowledge base that can be derived from the dataset.
- Technologies for building ontologies: RDF, SPARQL, OWL, and SKOS.

A Loomweight Ontology



CIDOC CRM Ontology

ICOM-CIDOC: The International Committee for Documentation of the International Council of Museums

- provides definitions and a formal structure for describing the implicit and explicit concepts and relationships used in cultural heritage documentation.
- enables information exchange and interchange between heterogeneous sources of cultural heritage information by providing a common and extensible semantic framework
- promote a shared understanding of cultural heritage information by transforming disparate, localised information sources into a coherent global resource.

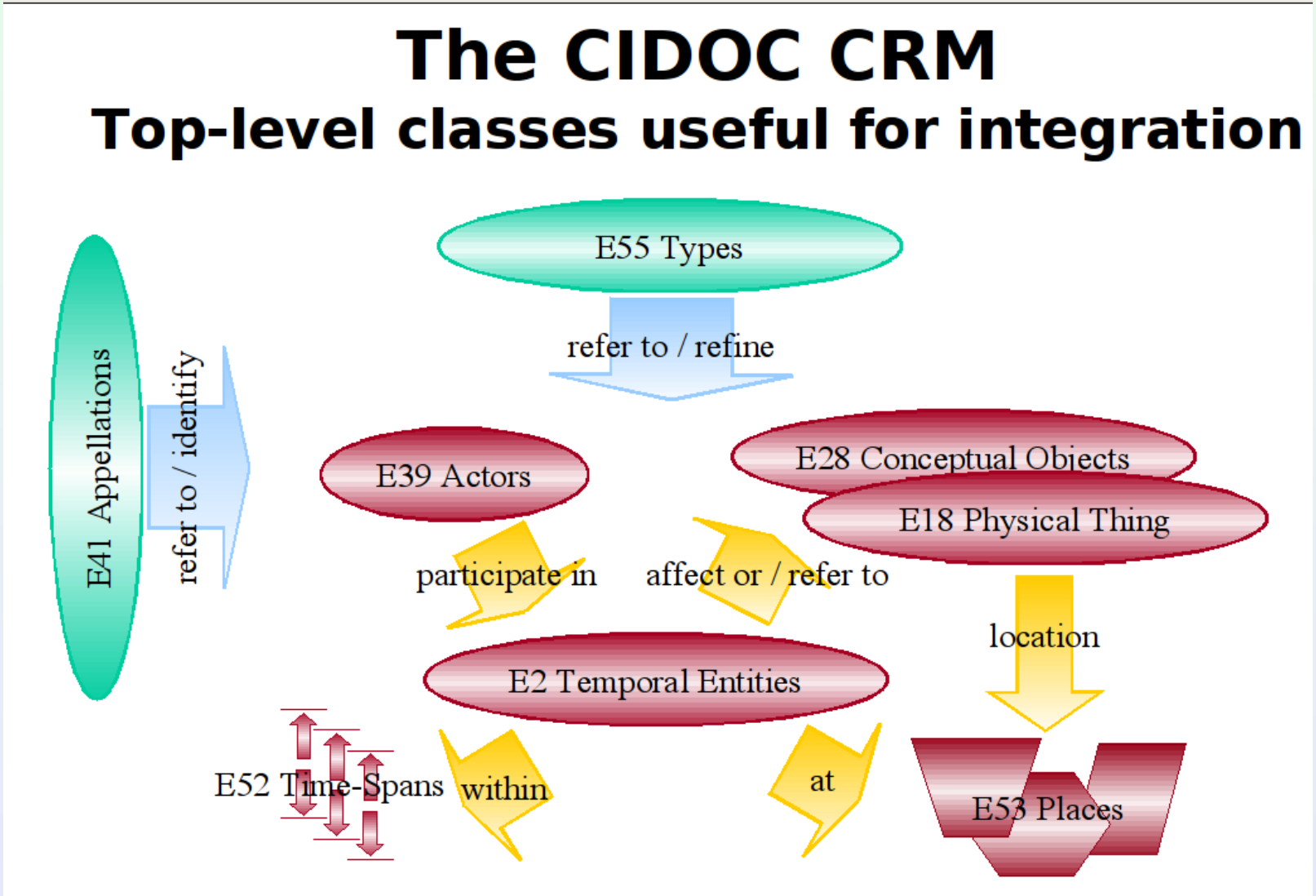
`http://cidoc.ics.forth.gr/index.html`

CIDOC CRM ontology

- A collaboration with the International Council of Museums.
- This is the most dominant ontology in cultural heritage.
- It is intended to cover the full spectrum of cultural heritage knowledge - from Archaeology to Art history, literary and musical entities.
- An ontology of 86 classes and 137 properties for culture and more.
- With the capacity to explain hundreds of (meta)data formats.
- Accepted by ISO TC46 in September 2000.
- International standard since 2006 - ISO 21127:2006.
- The ontology has been encoded in OWL2.0, OWLDL and RDFS.

CIDOC CRM Ontology

Top level classes



Some specific Concepts

The CIDOC CRM

Temporal Entity- Main Properties

- **E2 Temporal Entity**
  - Properties: P4 has time-span (is time-span of): E52 Time-Span
- **E4 Period**
  - Properties: P7 took place at (witnessed): E53 Place
    P9 consists of (forms part of): E4 Period
    P10 falls within (contains): E4 Period
- **E5 Event**
  - Properties: P11 had participant (participated in): E39 Actor
    P12 occurred in the presence of (was present at): E77 Persistent Item
- **E7 Activity**
  - Properties: P14 carried out by (performed): E39 Actor
    P20 had specific purpose (was purpose of): E5 Event
    P21 had general purpose (was purpose of): E55 Type

SKOS: Simple Knowledge Organization System

- SKOS is an area of work developing specifications and standards to support the use of knowledge organization systems (KOS) such as thesauri, classification schemes, subject heading systems and taxonomies within the framework of the Semantic Web.
- SKOS provides a standard way to represent knowledge organization systems RDF.
- SKOS is a data model
  - ..for data that can be published on the web...

`http://www.w3.org/2004/02/skos/intro`

SKOS: Simple Knowledge Organization System

ENGLISH HERITAGE
NATIONAL MONUMENTS RECORD

NMR Archaeological Sciences Thesaurus

View by Letter: A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

View by Class: 12/11/2009

CHARCOAL

SN	Wood that has been burnt and largely reduced to carbon as a result of burning in a reducing atmosphere below 500 degress C (Celsius).
CL	MATERIAL TYPE
BT	WOOD
NT	MICRO-CHARCOAL ROUNDWOOD TWIG
RT	CHARRED

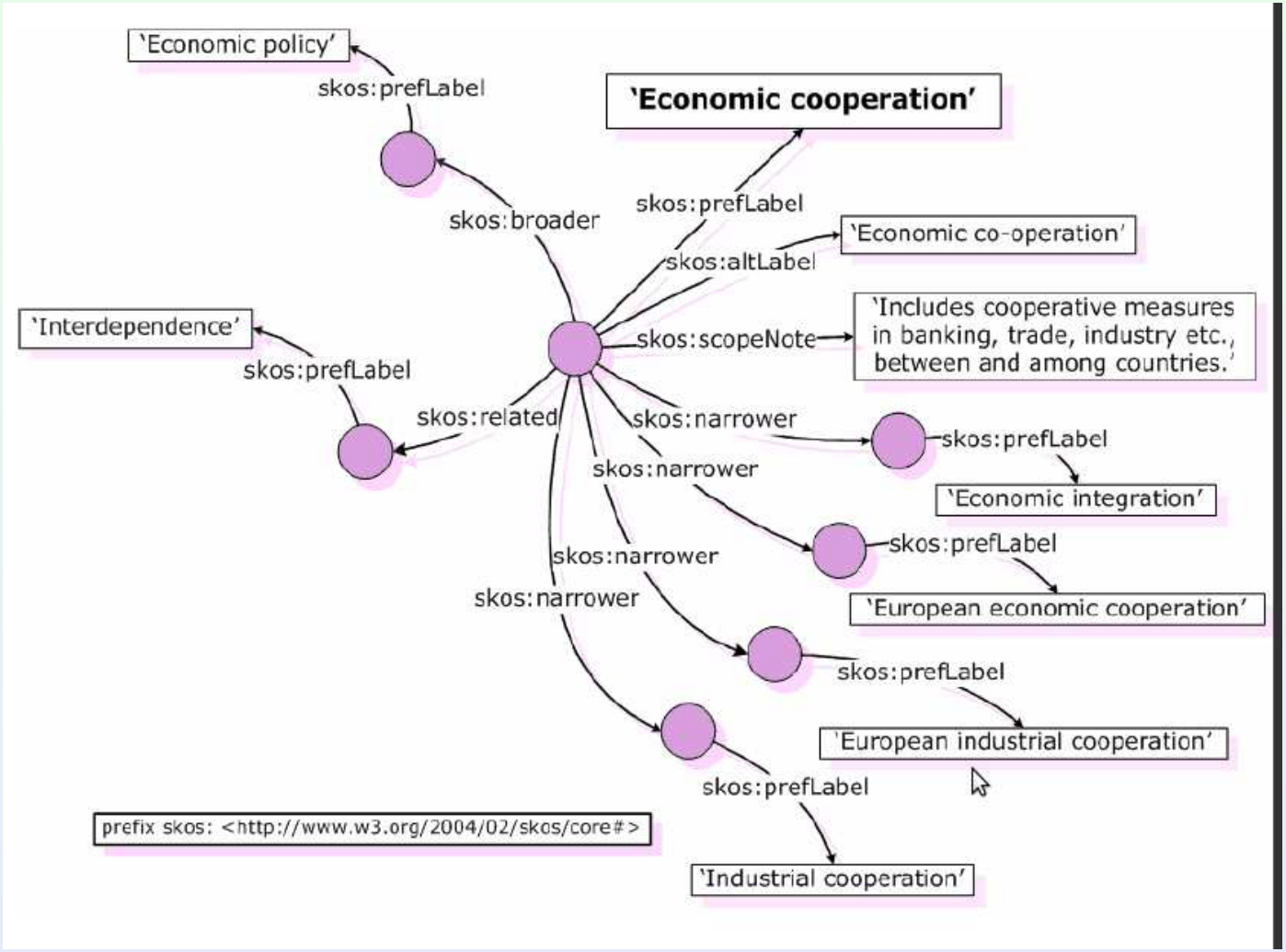
Class Listing (MATERIAL TYPE)

- ANTLER
- BEACH DEPOSIT
- BIOGENIC CARBONATE
- BONE
- BRICK
- BURNT FLINT
- COPROLITE
- EGG SHELL
- ESTUARINE DEPOSIT
- FEATHER
- FIBRE
- GEOLOGICAL SEDIMENT
  - FELDSPAR
  - POLYMINERAL
  - QUARTZ
  - ZIRCON
- HAIR
- LEATHER
- METAL
  - FERROUS METAL
  - NON-FERROUS METAL
    - COPPER ALLOY
    - GOLD
    - SILVER

- ..linked with other data...

http://thesaurus.english-heritage.org.uk/

SKOS: Simple Knowledge Organization System

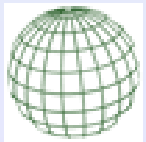


- ..data which is semantically interoperable...

Geonames

- A cool resource that provides a web service to query for information on world wide locations (city, town, place...) and returns the results in RDF.
- It provides a nice google map interface as well.
- The GeoNames Ontology makes it possible to add geospatial semantic information to the Word Wide Web. All over 6.2 million geonames toponyms now have a unique URL with a corresponding RDF web service.
- This has been used to add rich geographical information (e.g. latitude and longitude) to cultural archives that have. ambiguous text entries for place information.

`http://www.geonames.org/`



Geonames

- The service distinguishes “Concept” from “Document” about it.
- For the town Embrun in France we have these two URIs : `http://sws.geonames.org/3020251/` `http://sws.geonames.org/3020251/about.rdf`
- The first URI stands for the town in France. You use this URI if you want to refer to the town.
- The second URI is the document with the information geonames has about Embrun. This is called the “feature” document for the town.
- The geonames web server is configured to redirect requests for [1] to [2].

Geonames: An example “feature” RDF document

```
<Feature rdf:about="http://sws.geonames.org/3020251/">
 <name xml:lang="fr">Embrun</name>
 <alternateName xml:lang="fr">Embrun, Hautes-Alpes</alternateName>
 <featureClass rdf:resource="http://www.geonames.org/ontology#P"/>
 <featureCode rdf:resource="http://www.geonames.org/ontology#P.PPL"/>
 <inCountry rdf:resource="http://www.geonames.org/countries/#FR"/>
 <population>7069</population>
 <postalCode>05200</postalCode>
 <wgs84_pos:alt>900</wgs84_pos:alt>
 <wgs84_pos:lat>44.5667</wgs84_pos:lat>
 <wgs84_pos:long>6.5000</wgs84_pos:long>
 <parentFeature rdf:resource="http://sws.geonames.org/3013738/"/>
 <nearbyFeatures rdf:resource="http://sws.geonames.org/3020251/nearby.rdf"/>
 <locationMap>http://www.geonames.org/3020251/embrun.html</locationMap>
 <wikipediaArticle rdf:resource="http://nl.wikipedia.org/wiki/Embrun"/>
 <owl:sameAs rdf:resource="http://rdf.insee.fr/geo/COM_05046"/>
</Feature>
```

The Getty Vocabularies

- The Getty vocabulary databases are produced and maintained by the Getty Vocabulary Program.
- They are compliant with ISO and NISO standards for thesaurus construction.
- They contain terms, names, and other information about people, places, things, and concepts relating to art, architecture, and material culture.

`http://www.getty.edu/research/conducting_research/vocabularies/`

The Getty Vocabularies

The Getty vocabularies can be used in three ways:

- at the data entry stage, by catalogers or indexers who are describing works of art, architecture, material culture, archival materials, visual surrogates, or bibliographic materials;
- as knowledge bases, providing information for researchers;
- and as search assistants to enhance end-user access to online resources.

The Getty Vocabularies

- The Art & Architecture Thesaurus (AAT)
  - A structured vocabulary of around 34,000 concepts, including 131,000 terms, descriptions, bibliographic citations, and other information relating to fine art, architecture, decorative arts, archival materials, and material culture.
- The Union List of Artist Names (ULAN)
  - A structured vocabulary containing around 127,000 records, including 375,000 names and biographical and bibliographic information about artists and architects, including a wealth of variant names, pseudonyms, and language variants.

The Getty Vocabularies

- The Getty Thesaurus of Geographic Names (TGN)
  - A structured vocabulary containing around 895,000 records, including around 1,115,000 names, place types, coordinates, and descriptive notes, focusing on places important for the study of art and architecture.
- The Cultural Objects Name Authority (CONA)
  - A new Getty vocabulary currently under development (2011). It will include authority records for cultural works, including architecture and movable works such as paintings, sculpture, prints, drawings, manuscripts, photographs, ceramics, textiles, furniture, and other visual media such as frescoes and architectural sculpture, performance art, archaeological artifacts, and various functional objects that are from the realm of material culture and of the type collected by museums.

Part II

Semantic Web in Cultural Heritage

Museums, Digital Libraries and Archives community

Overview

- Online access to even the most important aspects of our past and cultural heritage is still limited and highly fragmented, despite large amounts of public funding being devoted to their digitisation.
- Digital libraries raise the quality, availability and diversity of information.
- The museum community are especially interested in the items and artefacts in their collections and their representation as multimedia.

Overview

- There is a strong demand for the innovation of enhanced facilities for multimedia information organisation, storage and retrieval.
- A constraint on the uptake of multimedia digital libraries is the limited amount of structured metadata available in such systems.

The DigiCULT project (2002-2004)

- Digital Culture (DigiCULT) - an IST Support Measure to establish a regular technology watch for cultural and scientific heritage.
- A rigorous publication agenda of seven Thematic issues, three in-depth Technology watch reports and a DigiCULT info e-journal.
- The third thematic issue addressed the question "What will the SW do for heritage institutions".
- The DigiCULT Expert 13 (historians, language and information technology scientists, academics and publishers) concluded that the cultural heritage sector needed the SW.

http://www.digicult.info/

Semantic web Think Tank (2006-2007)

- AHRC funded Thinktank to identify the key challenges and opportunities for museums and to define how the culture sector should respond to them.
- First Meeting in 2006 as part of the UK Museums and the Web conference led by Ross Parry (School of Museum studies @ UOL).
- 23 participants: technologists, journalists, publishers, culture-sector professionals, theorists, managers, project leaders and strategicians.

`http://culturalsemanticweb.wordpress.com/workshop-reports/`

Museum Finland

- The first major deployment of semantic web technologies in the Museum sector.
- A semantic portal: Finnish museums on the Semantic web.
- Provides tools and services so that Finnish museums could present their collections online through a common semantic web interface.
- Second prize in the 2004 Semantic we challenge.
- Follow up project in 2008: CultureSampo - Finnish Culture on the Semantic Web 2.0.

---

http://www.museosuomi.fi///

http://www.kulttuurisampo.fi/index.shtml

SCULPTEUR (2002)

- SCULPTEUR: Semantic and content-based multimedia exploitation for European benefit.
- Uses ontologies to structure and navigate nuseum collections
- “Concept Browser” allows users to access the information from the museum partners collaborating in the project structured through a common ontology, the CIDOC CRM.
- A web-based demonstrator for navigating, searching and retrieving 2D images, 3D models and textual metadata from the Victoria and Albert Museum - one of their partners
- and also includes a graphical ontology browser for users unfamiliar to the museum collection.

---

http://www.sculpteurweb.org/

eCHASE

- eCHASE: electronic Cultural Heritage made Accessible for Sustainable Exploitation
- A follow up project from SCULPTEUR.
- A demonstrator that allows users to search, collect and share multimedia entries from semantically integrated repositories across Europe.
- also includes a “lightbox” feature to support collaboration between eCHASE users.
- The eCHASE system can be accessed by external software applications as a Web Service.

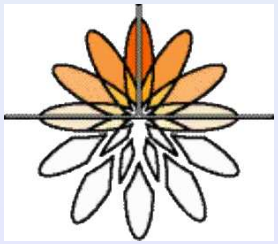
---

http://www.echase.org/

MultimediaN N9C Eculture project

- Development of a set of e-culture demonstrators providing multimedia access to distributed collections of cultural heritage objects.
- The architecture is fully based on open Web standards, in particular XML, SVG, RDF/OWL and SPARQL.
- Winner of the Semantic Web challenge in 2006.
- The demonstrator uses the Getty vocabularies (converted to RDF/OWL) as well as the lexical resource WordNet, version 2.0.

`http://e-culture.multimedian.nl/`



MultimediaN N9C Eculture project

- Collected descriptions of 200,000 objects from six collections.
  - Rijksmusuem Amsterdam.
  - National museum of ethnology.
  - The Royal Tropical institute.
  - The Netherlands institute for Art history.
  - The Royal Library.
  - Web collection archive.
- Collections are annotated with a range of thesauri and several proprietary controlled keyword lists.
- The semantic search is done over these collections.

CHIP @ Rijksmuseum Amsterdam

- The CHIP (Cultural Heritage Information Presentation) project is funded by Dutch Science Foundation.
- Three different components/tools in the CHIP demonstrator
  - Artwork Recommender. a Web-based rating dialog for artworks/topics to build a user profile
  - Tour Wizard. a Web-based tool using the user profile to generate (semi)automatically personalization virtual museum tours for each user.
  - Mobile Tour. a PDA-based tool to map virtual tours into the physical museum space with constraints; to give guidance to users and to synchronize the user profile on the web and in the PDA.

---

http://www.chip-project.org/

NWO TU/e RIJKSMUSEUM amsterdam Telematica Instituut

STITCH@CATCH

Semantic Interoperability To access Cultural Heritage

- To develop theory, methods and tools for allowing metadata interoperability through semantic links between the vocabularies.
- Cultural-heritage collections are typically indexed with metadata derived from a range of different vocabularies,.
- In general, it is unrealistic to assume unification of vocabularies.
- Research challenge is similar to what is called the “ontology mapping” problem in Semantic Web research.

`http://www.cs.vu.nl/STITCH/index.html`

CONTEXTA/SR

- CONTEXTA/SR is a platform whose goal is to semantically integrate cultural heritage heterogeneous information.
- Modelling: ontologies to describe artifacts, ontologies to describe circumstances, and social ontologies.
- It does not build on CIDOC/CRM, rather it builds on ISAD(G) (General International Standard Archival Description) standard.
- ISAD(G) is a very detailed ontology specialized for the heritage domain.
- A very important aspect of their work is their use of resolvable URI in order to support the linked data initiative.

www.archimuse.com/mw2008/papers/astudillo/astudillo.html

Part III

Semantic Web meets Archaeology

Overview

- In contrast to the cultural heritage sector aka museums, the Semantic Web has seen less uptake in archaeology.
- This could be because archaeologists tend to focus on analysis and recording of the data rather than dissemination.
- Experiences are mostly limited to spreadsheets, relational databases and/or spatial data management.
- Many academic archaeologists remain protective of their data especially when it has not been published in traditional media.
- The complexity of combining siloed resources may be overwhelming.

VBI-ERAT-LVPA (2004))

- An early example of the use of structured knowledge bases in the archaeological sphere.
- The VBI-ERAT-LVPA index is a knowledge base that integrates complementary archaeological information sources.
- The source data comprised complementary scientific databases and corpora describing finds with inscriptions and iconography of the Roman era.
- Each source schema was interpreted in terms of the CIDOC CRM model and stored as an RDF knowledge base.

Doerr M., Schaller K. and Theodoridou M. 2004. Integration of complementary archaeological sources. In Computer Applications and Quantitative Methods in Archaeology Conference (CAA 2004). Prato (Italy), 13-17 April 2004.

VBI-ERAT-LVPA (2004))

- A suite of transformation/mapping tools: databases, spreadsheet, text files $\rightarrow$ XML (compatible with CIDOC CRM) $\rightarrow$ RDF.
- RDF Suite which allows for effective and efficient management of large volumes of RDF descriptions.
- Semantic web portal generator in order to provide a Web-based easy to use by archaeologists user interface. The user interface allows the formulation of three types of queries.

STAR (2007 - 2010)

- STAR: Semantic Technologies for Archaeological Resources
- A 3 year AHRC funded project, in collaboration with English Heritage and the Royal School of Library and Information Science Denmark.

Aim

To investigate the potential of semantic terminology tools for widening and improving access to digital archaeology resources, including disparate data sets and associated grey literature.

`http://hypermedia.research.glam.ac.uk/kos/star/`

STAR (2007 - 2010)

Outputs:

- In collaboration with English Heritage, a set of extensions to the CIDOC CRM core ontology (CIDOC CRM-EH Ontology) have been produced as RDF files.
- The project has developed an initial set of semantic web services, based upon the SKOS thesaurus representations.

archeoserver

- Led by Dipartimento di Informatica, Sistemistica e Comunicazione, University of Milano Bicocca
- Developed an ontology and navigation tools for an e-library on pre and proto - history in Italy.
- Looking into integration of excavation datasets in the Po valley.
- It does not use CIDOC/CRM, but uses its own ontology developed using the NavEditOW ontology editor.

`http://www.archeoserver.it/space/start`

Virtual Environments for Research in Archaeology

VERA

- The project aims to produce a fully-fledged virtual research environment for the archaeological community.
- Develop services to increase the uptake of IADS@York.
- Future intentions: incorporate RDF for a broader archaeological data integration system.

`http://vera.rdg.ac.uk/`

Roman Ports Project (2007 -)

- a project undertaken by the University of Southampton and British School at Rome, has brought together a multitude of partners in order to share data about amphora and marble distribution along the Mediterranean littoral.
- A very good application of all round semantic web technologies.
- developed a guided process by which relational data columns can be mapped to concepts within an ontology,
- Natural Language Processing used to facilitate the mapping of local terms to concepts within different concept schemes.
- Spatial toponyms are also extracted and integreated using the Geonames webservices.

http://www.bsr.ac.uk/BSR/sub_arch/BSR_Arch_05Roman.htm

The Open Archaeology Software Suite

- Open Archaeology will host a number of sub-projects that while functional at a component level are intended to interoperate creating a complete AIS (Archaeological Information System), including the components necessary to manage the organisation carrying out the archaeology.
- Software project started during 2006 by Oxford Archaeology.
- Use of RDF is already evident. The project information is available to be downloaded as RDF metadata.

`https://launchpad.net/openarchaeology`

Summary: The Heritage perspective

- Archaeological data sources are fragmentary by nature.
- Theoretical approaches used by practitioners are diverse.
- “Data is sacred” - expressing one’s knowledge base in terms of another’s ontology, may not always be “acceptable”.
- Adoption of the Semantic Web by the heritage sector depends upon the syntactical and semantic mark-up of content.
- The sector should coordinate their efforts to ensure that the fundamental building blocks that can enable their success on the semantic web are in place.
- Try not to “reinvent the wheel” in terms of metadata - use existing annotation schemes.

Summary: The Semantic Web perspective

- Provide facilities for harvesting, enriching and aligning collection metadata and vocabularies.
- Provide facilities for semantic search through the resulting graph, including various presentation mechanisms for the search results.
- User involvement is a key thing in cultural heritage: provide facilities for non experts to contribute to the metadata and annotations through folksonomies aka social tagging and faceted classification.

Questions

&

Discussions

http://www.kb.nl/manuscripts/