Aalto University
School of Science

SeCo

Semantic Web

CS-E4410 Semantic Web, 9.1.2019

Eero Hyvönen
Aalto University, Semantic Computing Research Group (SeCo) *http://seco.cs.aalto.fi*
University of Helsinki, HELDIG *http://heldig.fi*

eero.hyvonen@aalto.fi

Outline

The idea of Semantic web

Core of Semantic web

- Metadata, ontologies, reasoning
- Review of the technological solutions and standards

Application domains

Fundamental barrier for the development of the web: machine-processability

The web contents are created for human readers

- HTML, PDF, JPEG, ...

Machine mediates and displays, but does not "understand" contents of the web

- E.g., a Finnish text article

A web service ≈ machine helps human

- Requires machine-"understandability" of the contents

➔ **A fundamental contradiction**

How can we build a more intelligent web?

1. **Applications are programmed to be more intelligent**
- The contents stay as they are
- The machines operate more human-like (Artificial Intelligence)
2. **Contents are represented in a more intelligent way**
- The contents are easier to understand
- Machines stay more or less as they are

In practice, both ways are needed

- More intelligent systems process more intelligently represented contents

Approach 1: More intelligent applications

Automatic interpretation of natural language is difficult

- Free form of the documents
- Semantics of the content

Non-textual contents

- Pictures, sound, music, video, software, …
- How to interpret algorithmically?

More than the document itself is needed for interpretation

- Context + common sense needed
- Fundamental problems of Artificial Intelligence, easy for humans!
- Great scientific and technological challenges

Approach 2: Contents represented in a more intelligent way

The foundation of Semantic web

- The information is stored in a way that a machine understands it!
- Human helps the machine
  - *Machine can also help in this (user-friendly tools for semantic content creation)*

The development began in the beginning of the 2000s

- W3C Semantic Web Activity 2001
- W3C Web Services Activity 2002

Web generations

1G WWW:
- WWW pages for human interpretation
- HTML language

2G WWW:
- Structures for human/machine interpretation
- XML language

3G WWW: Semantic Web
- Meanings for human/machine use
- RDF(S) language

4G WWW: Ubiquitous web for humans and machines

⇒ **Semantics = new foundation for intelligent web services**
- Semantic = “understandable” to machines

Limitations of non-semantic web: case MuseumFinland

```
<artifact>
  <id>NBA:H26069:467</id>
  <target>cup and plate</target>
  <material>porcelain</material>
  <creationLocation>Germany</creationLocation>
  <creator>Meissen</creator>
</artifact>
```

- This metadata cannot answer the following questions:
  - Find all vessels?
  - Find all ceramic products?
  - Find artifacts manufactured in Europe?
  - Does the city of Meissen manufacture ceramics?

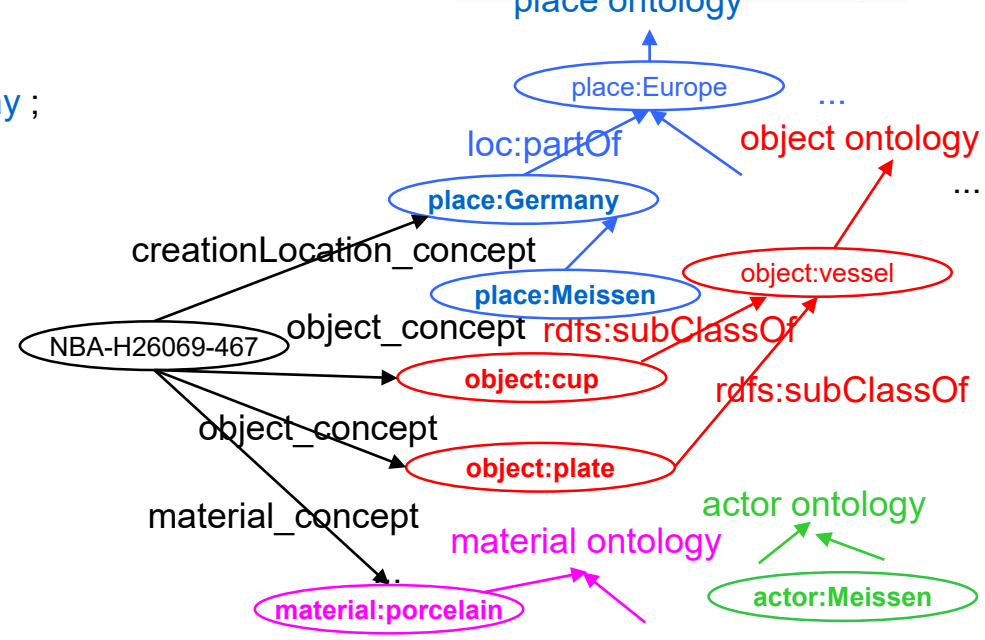
Semantic web solution: ontologies

NBA-H26069-467
:object ”cup and plate” ;
:object_concept object:cup ;
:object_concept object:plate ;

:material ”porcelain” ;
:material_concept object:porcelain ;

:creationPlace ”Germany” ;
:creationPlace_concept place:Germany ;

:creator ”Meissen”
:creator_concept actor:Meissen .



Find all vessels?
Find all ceramic products?
Find artifacts manufactured in Europe?
Does the city of Meissen manufacture ceramics?

Case Rijksmuseum Amsterdam: CHIP Demonstrator

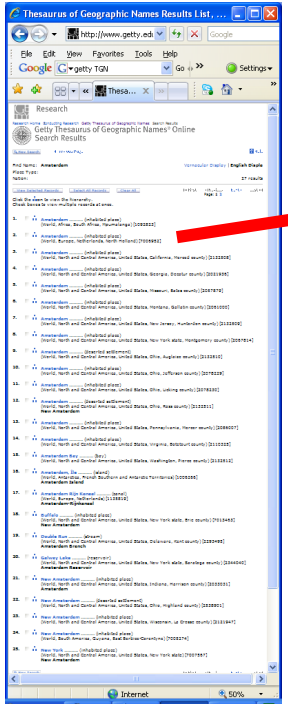
Example in Turtle notation

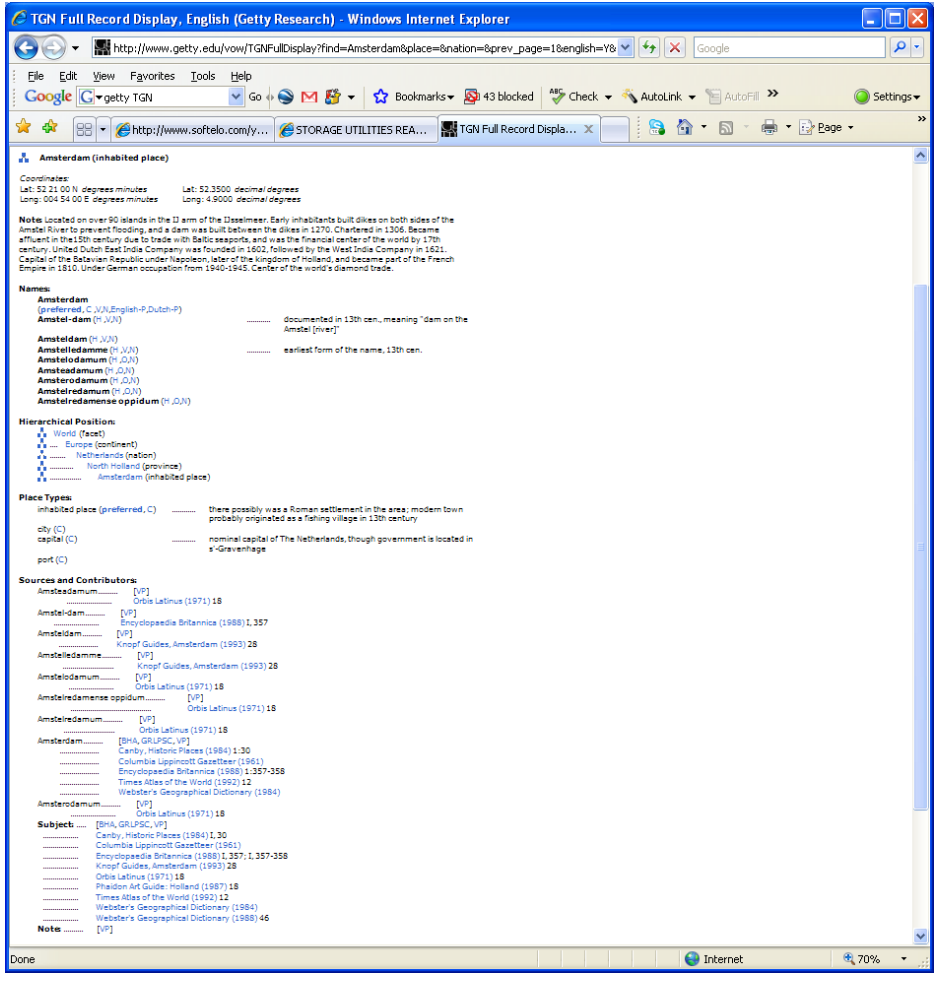
- VRA metadata schema (extension of Dublin Core)
- (Aroyo et al., 2007)

```
rijks:artefactSK-C-K
  vra:type vra:Work ;
  vra:title "The Night Watch" ;
  vra:date "1642" ;
  vra:creator: 500011051 ;          # Rembrandt
  vra:subject iconclass:45F31 ;     # Call to arms
  vra:culture tgn:7006952 ;         # Amsterdam
  vra:material aat:30015050 .       # Oil paint
```

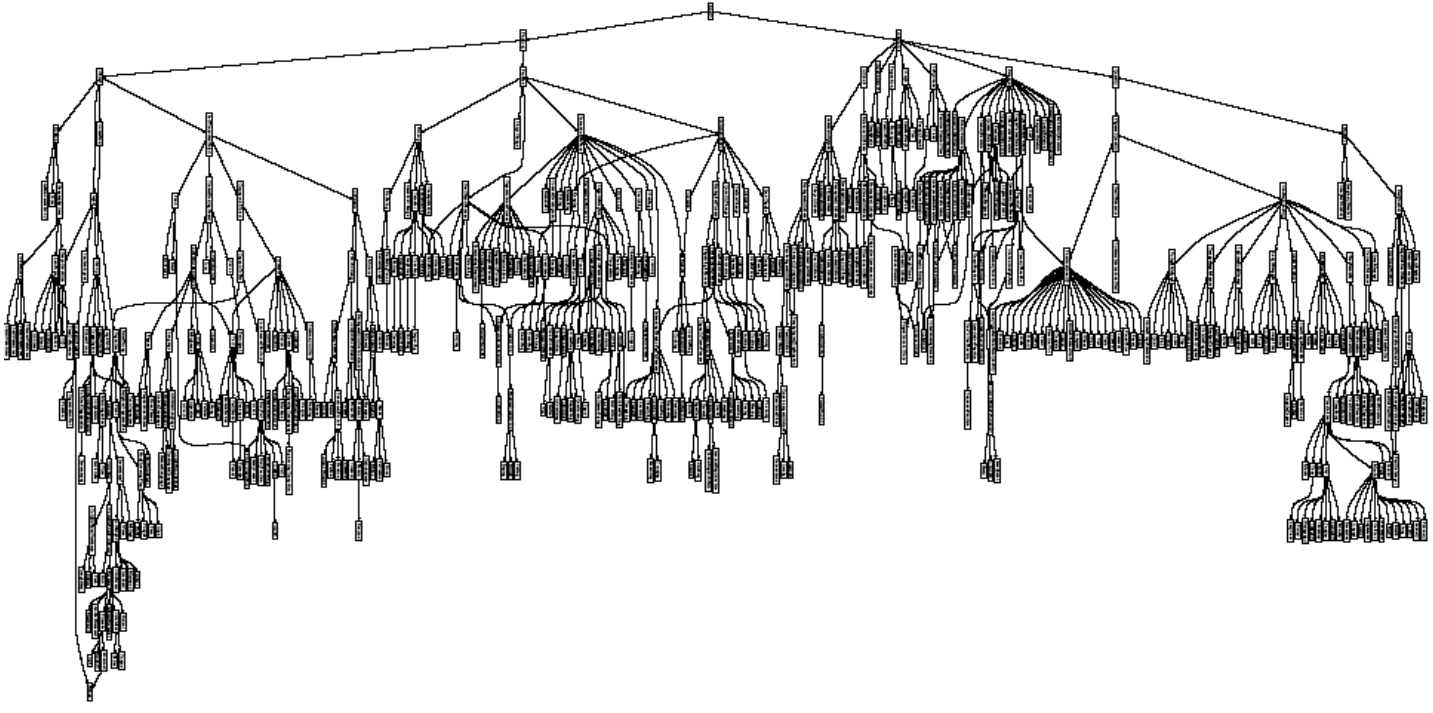
A resource in the TGN ontology / vocabulary

Amsterdam in TGN





An Ontology Concept Hierarchy: Standard Upper Merged Ontology SUMO



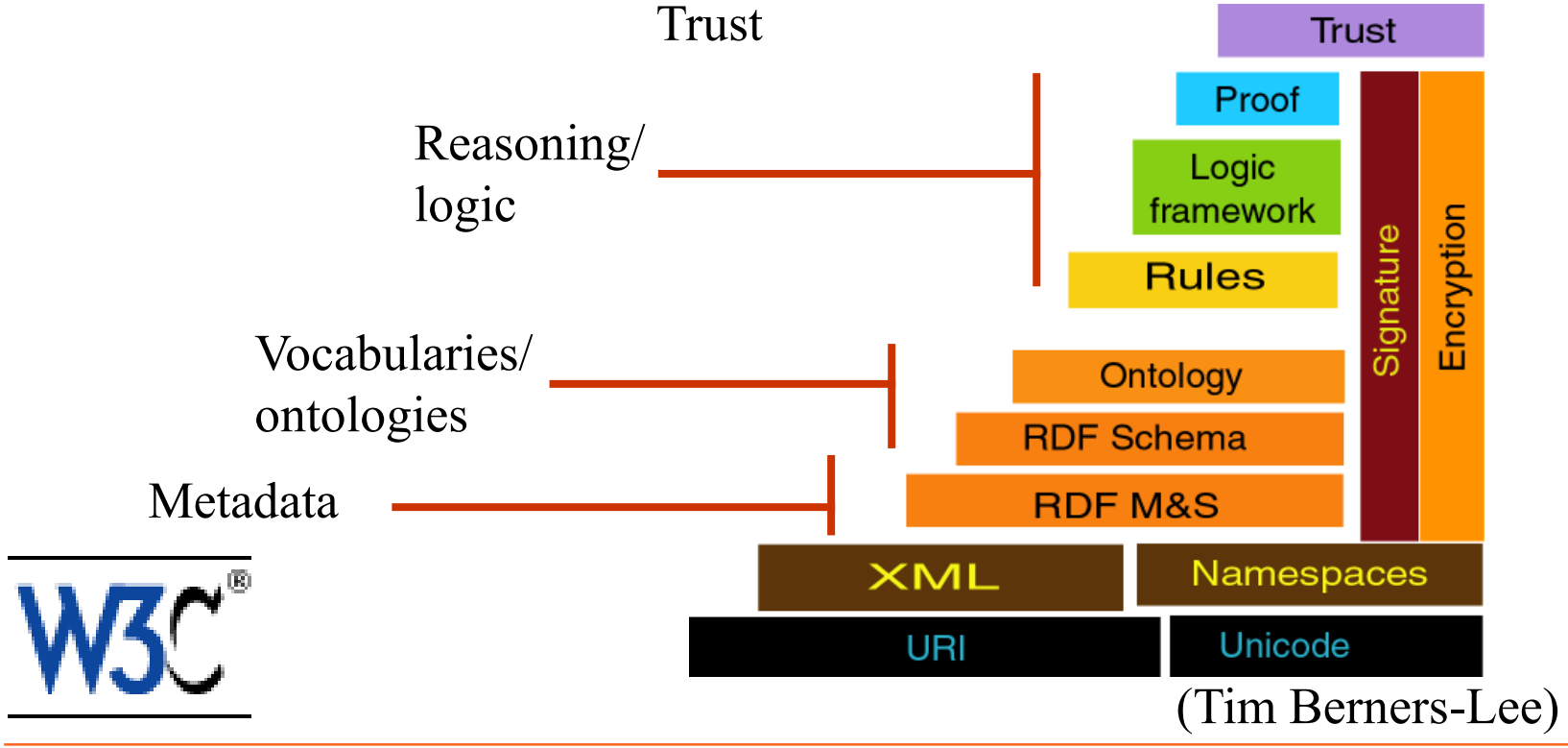
Technological basis of Semantic Web

Aalto University
School of Science

Department of
Computer Science

SeCo

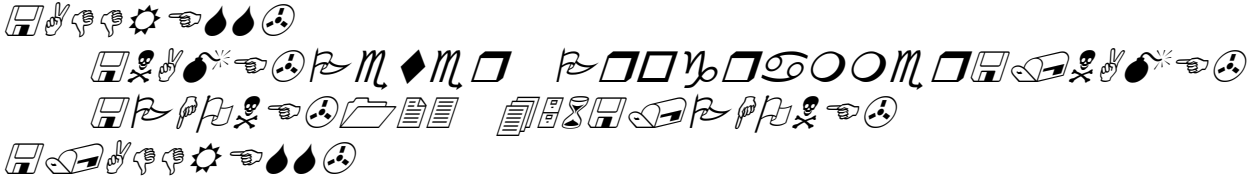
The classical ”layer cake model”



(Tim Berners-Lee)

Metadata level

Why isn't XML alone sufficient for the basis of Semantic web?

- Interpretation of XML languages has to be defined in a domain-specific way
- Combining different XML languages is often difficult
- We need a markup language, whose interpretation is:
  - *Commonly agreed*
  - *Shared across different application domains*
  - *Machine-"understandable"*
- The semantics of XML is only in human brain
  - *<ADDRESS>*
    <NAME>Peter Programmer</NAME>
    <PHONE>123 456</PHONE>
    </ADDRESS>
  - 

The Semantic web solution: RDF Resource Description Framework

- General metadata description language for web resources
- Relational model, *not* a syntax (as opposed to XML)
  - *RDF description = directed graph*
- Semantics is defined based on logic
- Syntax/serialization
  - *XML-based RDF/XML, especially for machines*
  - *Simple triple notations (N3, Turtle, N-triples) for humans*
- Standardized and commonly used
  - *W3C draft 1999*
  - *W3C recommendation RDF 1.0, 10.2.2004*
  - *W3C recommendation RDF 1.1, 25.2.2014*

RDF Example

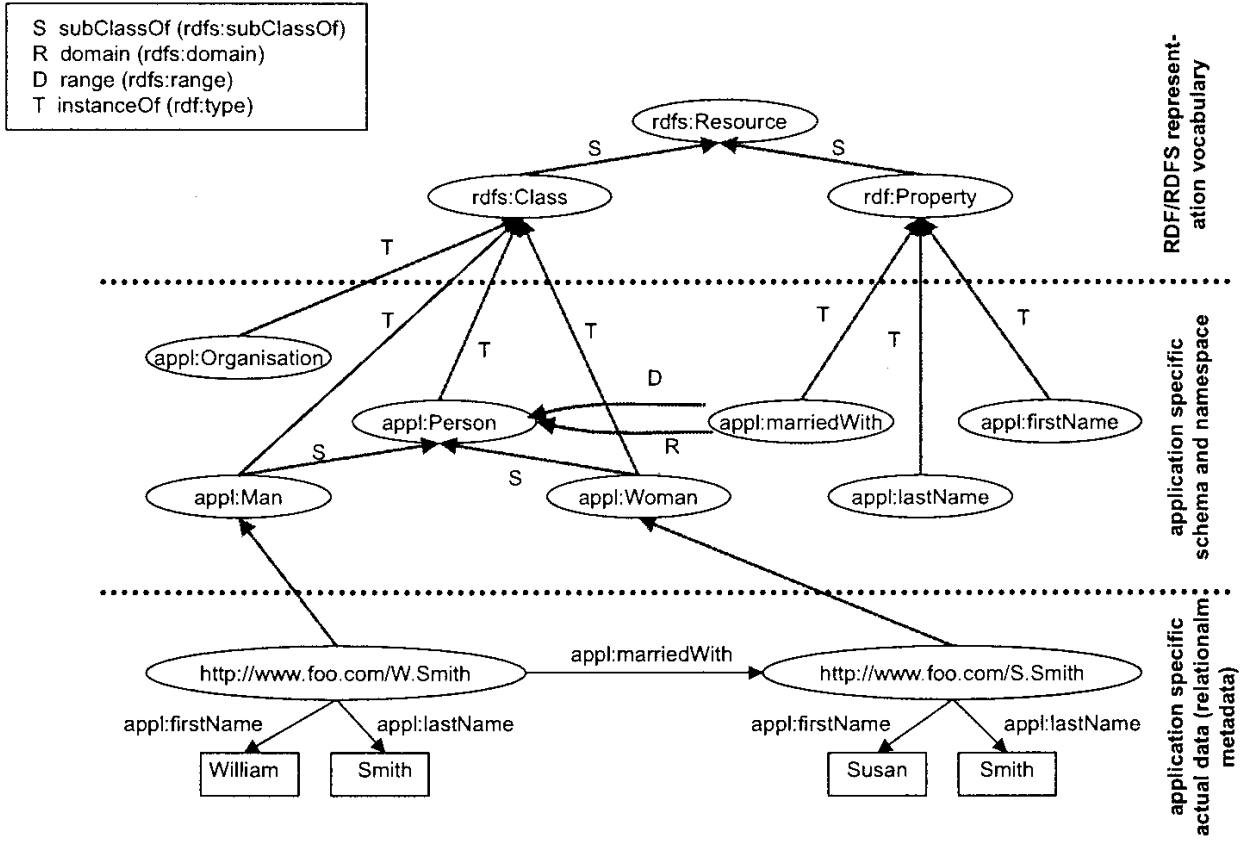


Figure 3.4. An RDF-Schema Example

(Maedche, 2002)

Metadata schemas

Standardized templates for representing (meta)data

- A set of elements describing object types (e.g., books in a library)
- Values for the properties describing individual objects (e.g., “War and Piece”)

Different content types typically require different properties

Example: Dublin Core for web documents

Set of 15 general properties for different content types

- Dublin Core Metadata Element Set (ISO Standard 15836)
  - *Title*
  - *Creator*
  - *Subject*
  - *Description*
  - *Publisher*
  - *Contributor*
  - *Data*
  - *Type*
  - *Format*
  - *Identifier*
  - *Relation*
  - *Source*
  - *Language*
  - *Coverage*
  - *Rights*

Metadata Schema in HealthFinland

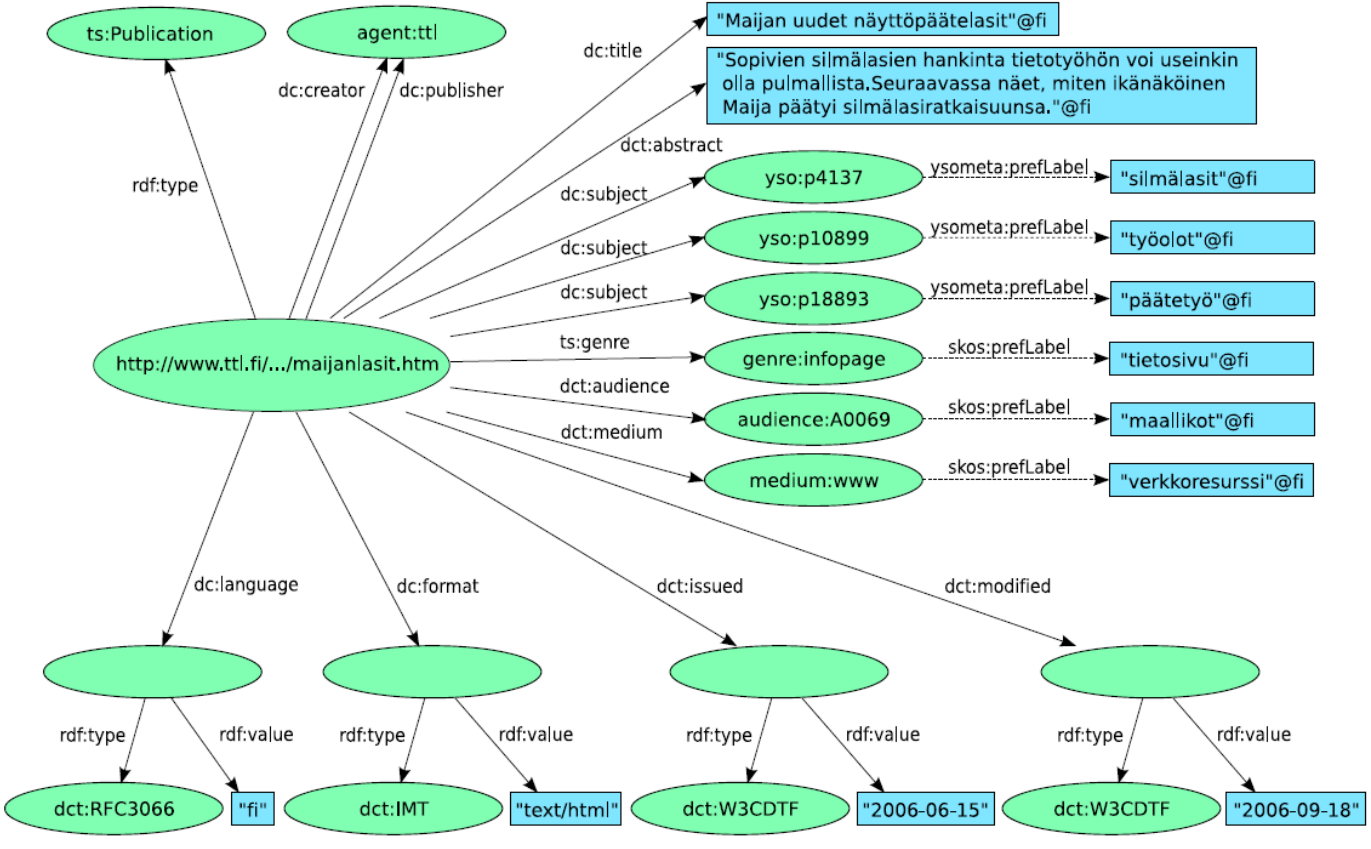
Table 1. HEALTHFINLAND Metadata Schema. Obligatory fields are marked in **bold**. Cardinalities are presented in the column C.

	Name	QName	C	Value type	Value range
General metadata	**Identifier**	**dc:identifier**	1	URI	
	Locator	ts:url	0..1	URL	
	Title	**dc:title**	1[a]	Free text	Non-empty string.
	Abstract	**dcterms:abstract**	1[a]	Free text	Non-empty string.
	Language	**dc:language**	1..*	String	RFC 3066
	Publication time	**dcterms:issued**	1	String	W3CDTF (ISO 8601)
	Acceptance time	dcterms:dateAccepted	0..*	String	W3CDTF (ISO 8601)
	Modification time	dcterms:modified	0..*	String	W3CDTF (ISO 8601)
	Publisher	**dc:publisher**	1..*	Instance	foaf:Organization
	Creator	dc:creator	0..*	Instance	foaf:Organization, foaf:Person or foaf:Group
Content classification	**Subject**	**dc:subject**	1..*	Concept	YSO, MeSH and HPMulti Ontologies
	Audience	**dcterms:audience**	1..*	Concept	Audience Ontology
	Genre	**ts:genre**	1..*	Concept	Genre Ontology
	Presentation type	**dc:type**	1..*	Concept	DCMI Type vocabulary
	Format	**dc:format**	1	String	IANA MIME types
	Medium	**dcterms:medium**	1	Concept	Medium Ontology
	Spatial coverage	dcterms:spatial	0..*	String or concept	DCMI Point, DCMI Box or Location Ontology
	Temporal coverage	dcterms:temporal	0..*	String or concept	W3CDTF, DCMI Period or Time Ontology
Relations	Part of	dcterms:isPartOf	0..*	Document	URI
	Rights	dc:rights	0..*	Free text or document	URI or textual description
	Source	dc:source	0..*	Free text or document	URI (e.g., ISBN) or bibliographical reference
	Reference	dcterms:references	0..*	Free text or document	URI (e.g., ISBN) or bibliographical reference
	Translation of	ts:isTranslationOf	0..*	Document	URI
	Format of	dcterms:isFormatOf	0..*	Document	URI

[a] Multilingual values are allowed, but only one value in each language.

HealthFinland portal:
Maija's eyeglasses – PDF document on the web

Maija’s eyeglasses: metadata in RDF form



Ontology Level

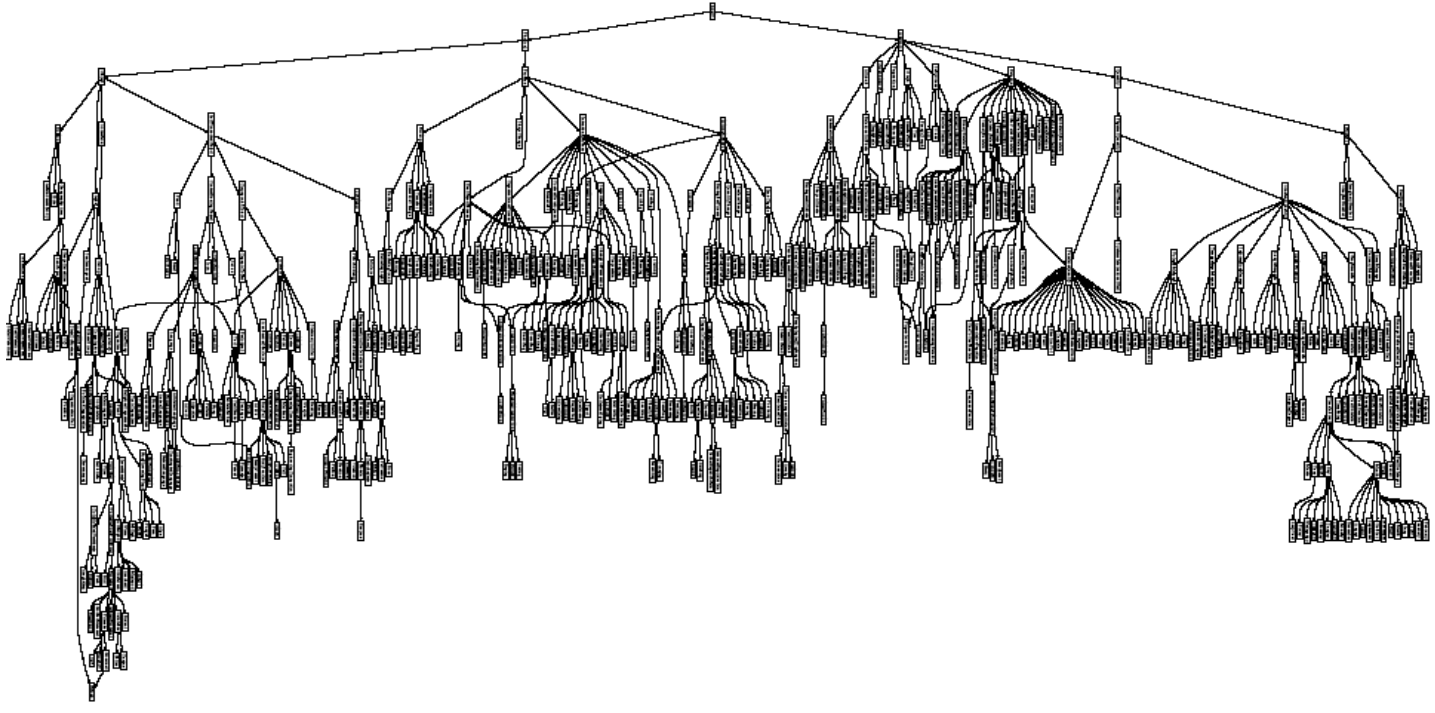
What is an ontology?

"An ontology is an explicit specification of a conceptualization ...definitions need to be couched in some common formalism" (Gruber, 1993)

- *Explicit: machine can understand*
- *Common (shared): communication is possible*
- *Formal: precisely defined*

- **Defines the concepts/objects and their relations in a given domain**
- **A first requirement for the humans and machines to understand each other**

Standard Upper Merged Ontology SUMO



SUMO principal distinctions

SUMO Object:

- Object
  - SelfConnectedObject
    - Substance
      - PureSubstance
        - ElementalSubstance
          - Metal
          - Atom
          - SubatomicParticle
            - AtomicNucleus
            - Electron
            - Proton
            - Neutron
        - CompoundSubstance
          - Water
          - Molecule
      - Mixture
        - Solution
        - Mineral
        - BodySubstance
      - BiologicallyActiveSubstance
        - Nutrient
        - Hormone
    - CorpuscularObject
      - OrganicObject
        - Organism
        - AnatomicalStructure
      - Artifact
      - ContentBearingObject
    - Food
  - Region
    - GeographicArea
    - AstronomicalBody
    - Hole
  - Collection
    - Group
      - GroupOfPeople
      - Organization
  - Agent
    - Organism
    - Group
    - GeopoliticalAgent
    - SentientAgent

Address http://webapps.getty.edu/vow/AATHierarchy?find=&logic=AND&note=&english=N&subjectid=300000000 Go Links

Research Tools / Vocabulary Databases

Art & Architecture Thesaurus On Line

AAT Hierarchy Display

NEW SEARCH

back to previous page

VIEW CHECKED RECORD(S) CLEAR ALL HELP

Click icons ( ) to view the hierarchy. Check boxes to view multiple records at once.

- Top of the AAT hierarchies
- ..... Associated Concepts Facet
- ..... Associated Concepts
- ..... Physical Attributes Facet
- ..... Attributes and Properties
- ..... Conditions and Effects
- ..... Design Elements
- ..... Color
- ..... Styles and Periods Facet
- ..... Styles and Periods
- ..... Agents Facet
- ..... People
- ..... Organizations
- ..... Activities Facet
- ..... Disciplines
- ..... Functions
- ..... Events
- ..... Physical and Mental Activities
- ..... Processes and Techniques
- ..... Materials Facet
- ..... Materials
- ..... Objects Facet
- ..... Object Groupings and Systems
- ..... Object Genres
- ..... Components
- ..... Built Environment
- ..... Furnishings and Equipment
- ..... Visual and Verbal Communication

Back to top

AAT Art & Architecture Thesaurus

- maintained by J. Paul Getty Trust

- 7 main classes, 125 000 concepts

abstract concepts

agents

events

materials

items

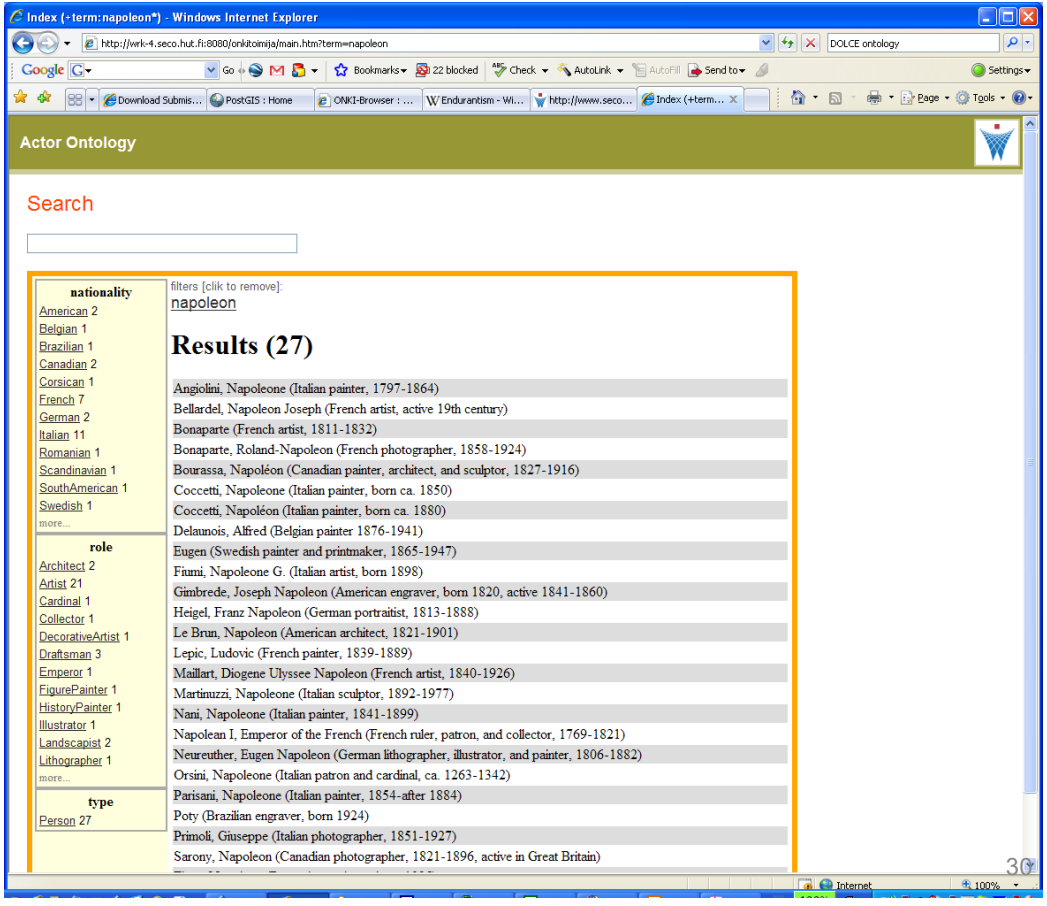
archive and library material

organisms

environments

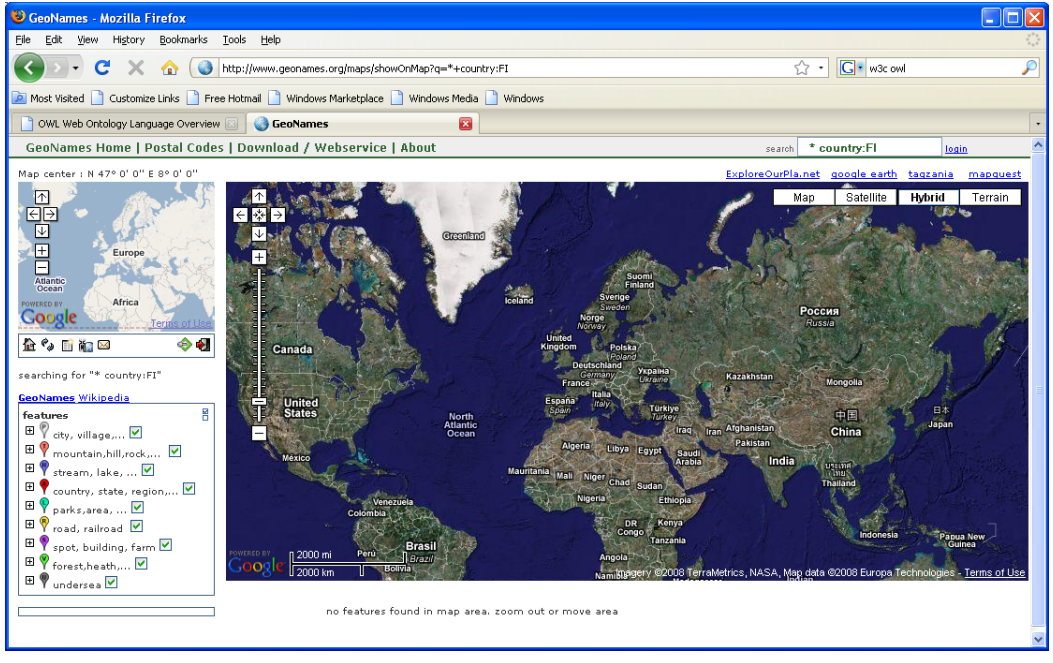
Universal List of Artist Names ULAN

- 120,000 instances
- 293,000 names



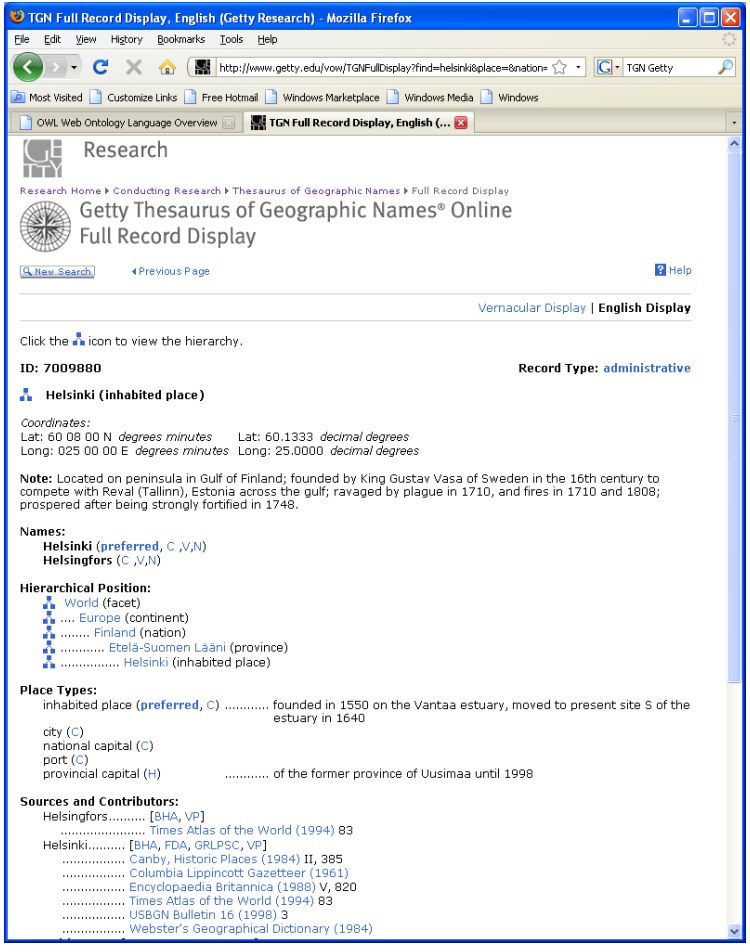
Geonames

- Classes: 9 feature classes, 645 feature codes
- Instances:
  - *8 million geographical names, 6.5 million unique features, 2.2 million populated places, 1.8 million alternate names*
  - *Registries and Wiki used for populating the ontology*

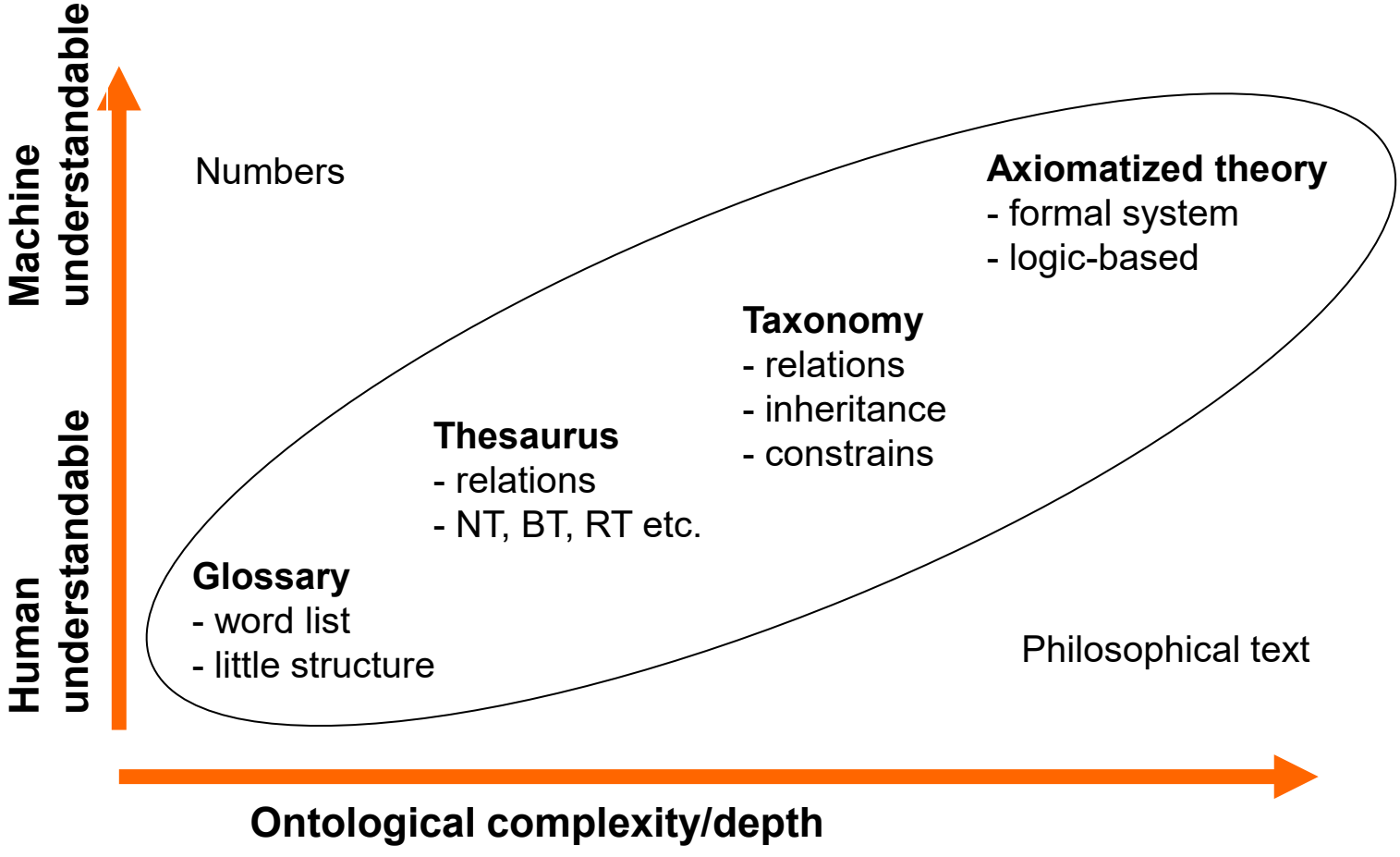


TGN Thesaurus of Geographical Names

- 912,000 records
- 1.1 million names, place types, coordinates, and descriptive notes
- Places important for the study of art and architecture



Ontology types



W3C standards for Semantic web ontologies/vocabularies

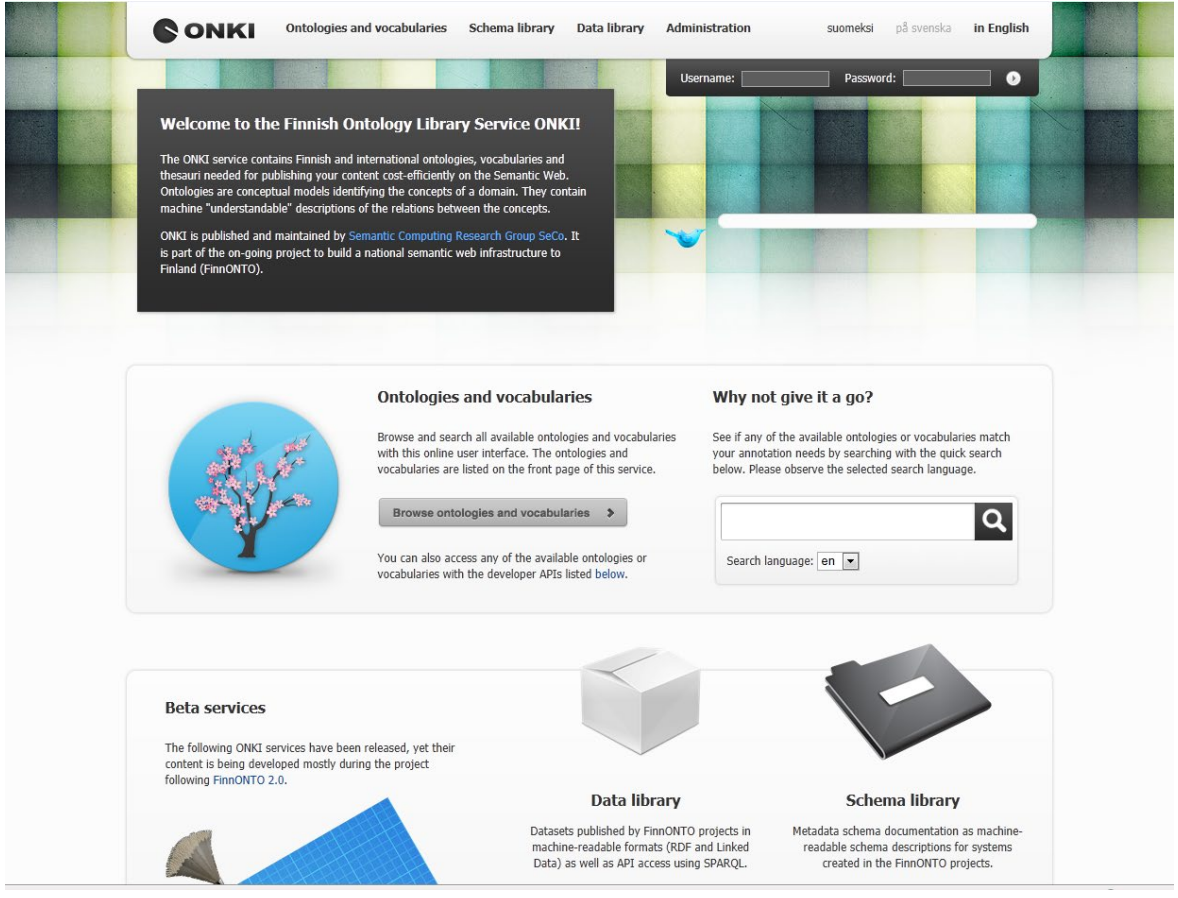
SKOS Simple Knowledge Organization System

- Light-weight semantics
- E.g., for representing existing glossaries, classification schemes, thesauri

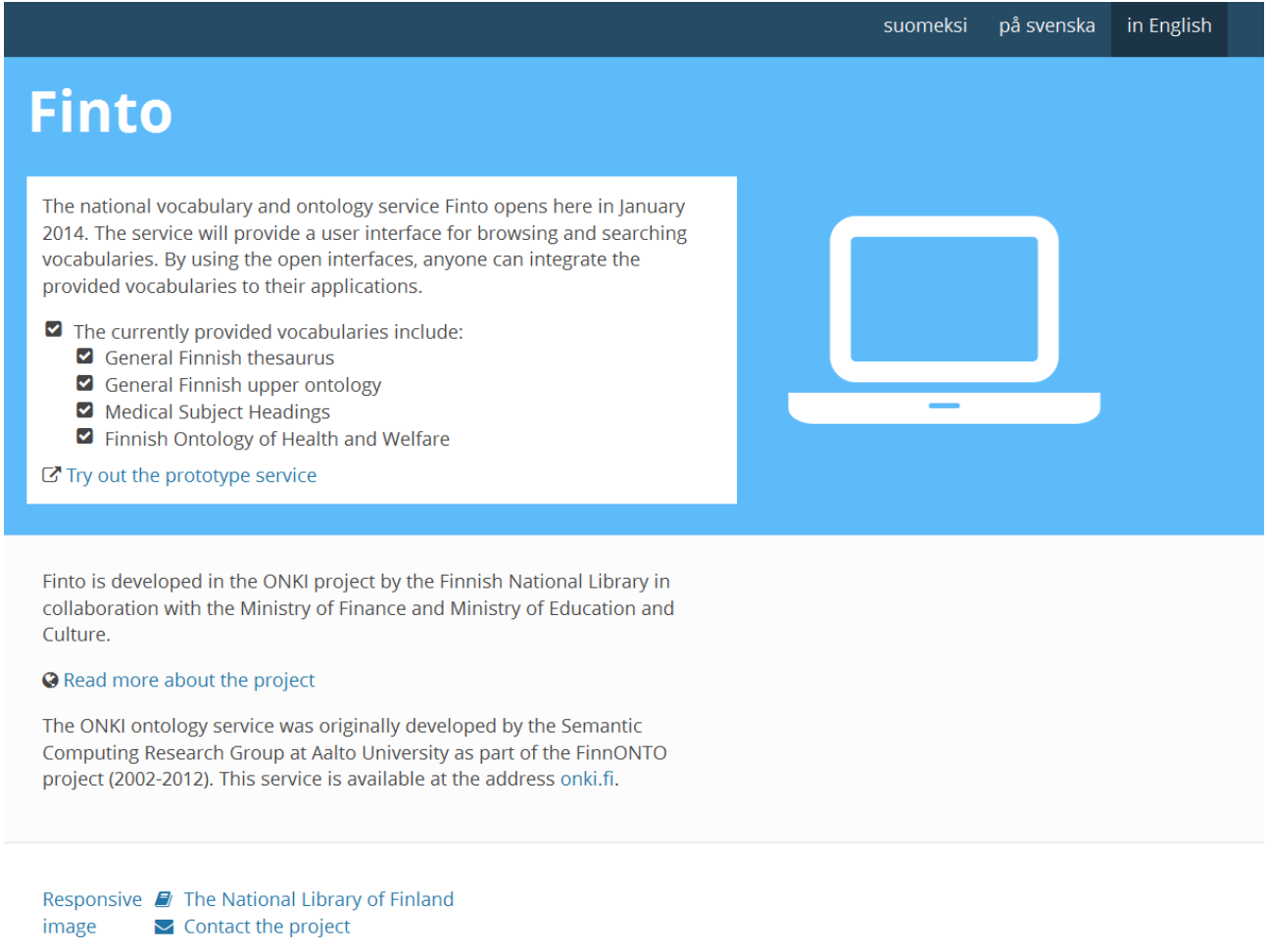
OWL Web Ontology Language

- Rich semantics based on logic
- Supports more reasoning

Finnish Ontologies: ONKI.fi



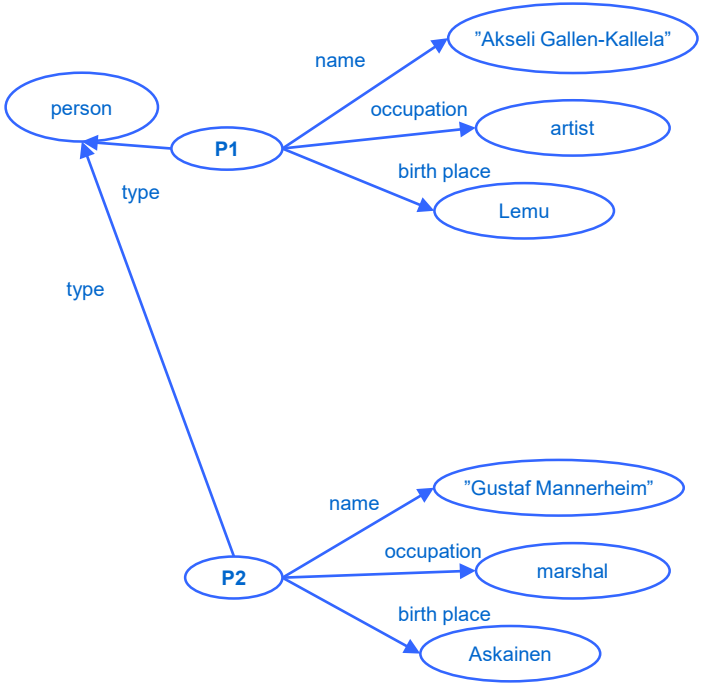
ONKI -> Finto.fi



Metadata + Ontologies = Linked Data (Web of Data)

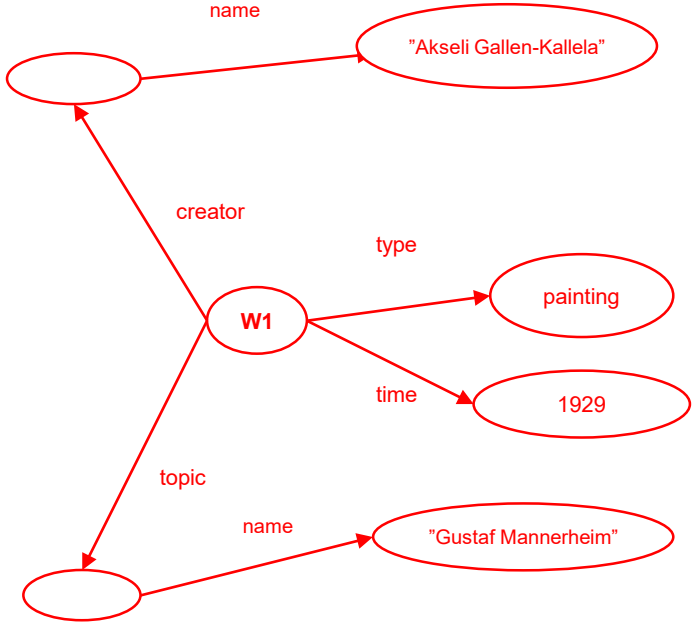
Finnish Biography center and libraries collect historical data of people

person	name	occupation	birth place	…
P1	Akseli Gallen-Kallela	artist	Lemu	
P2	Gustaf Mannerheim	marshal	Askainen	
…				



Museum catalogues paintings

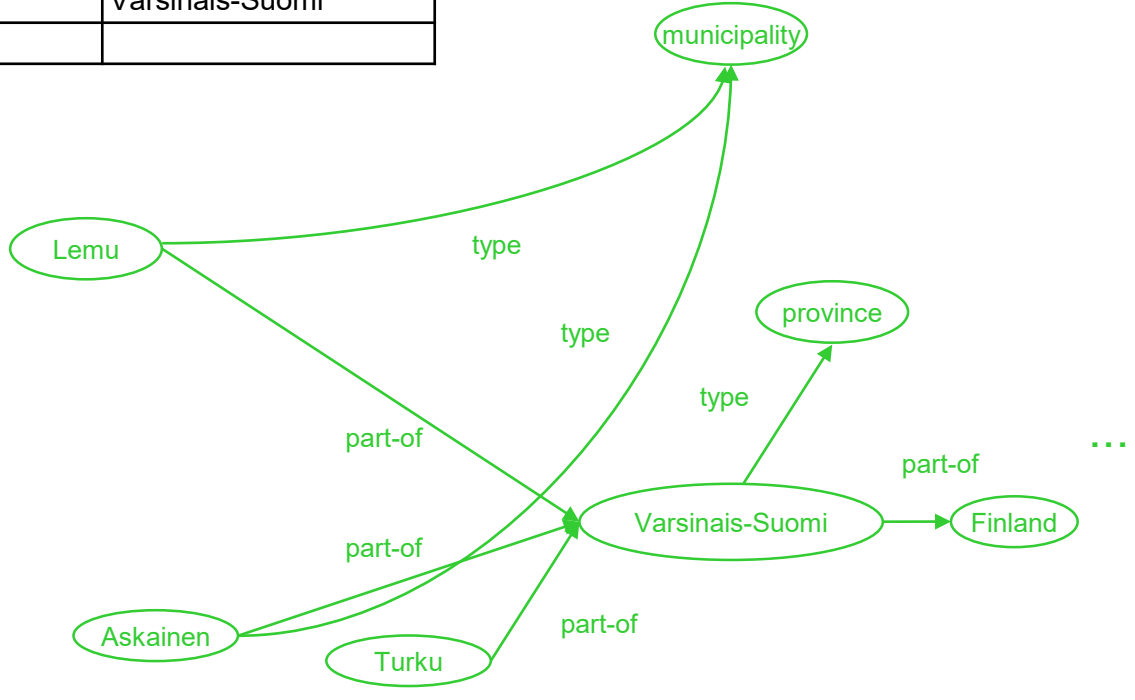
Work	name	creator	time	Topic	…
W1	Portrait of Mannerheim	Akseli Gallen-Kallela	1929	Gustaf Mannerheim	
W2	Aino Triptych	Akseli Gallen-Kallela	1891	Aino, Kalevala	
…					



…

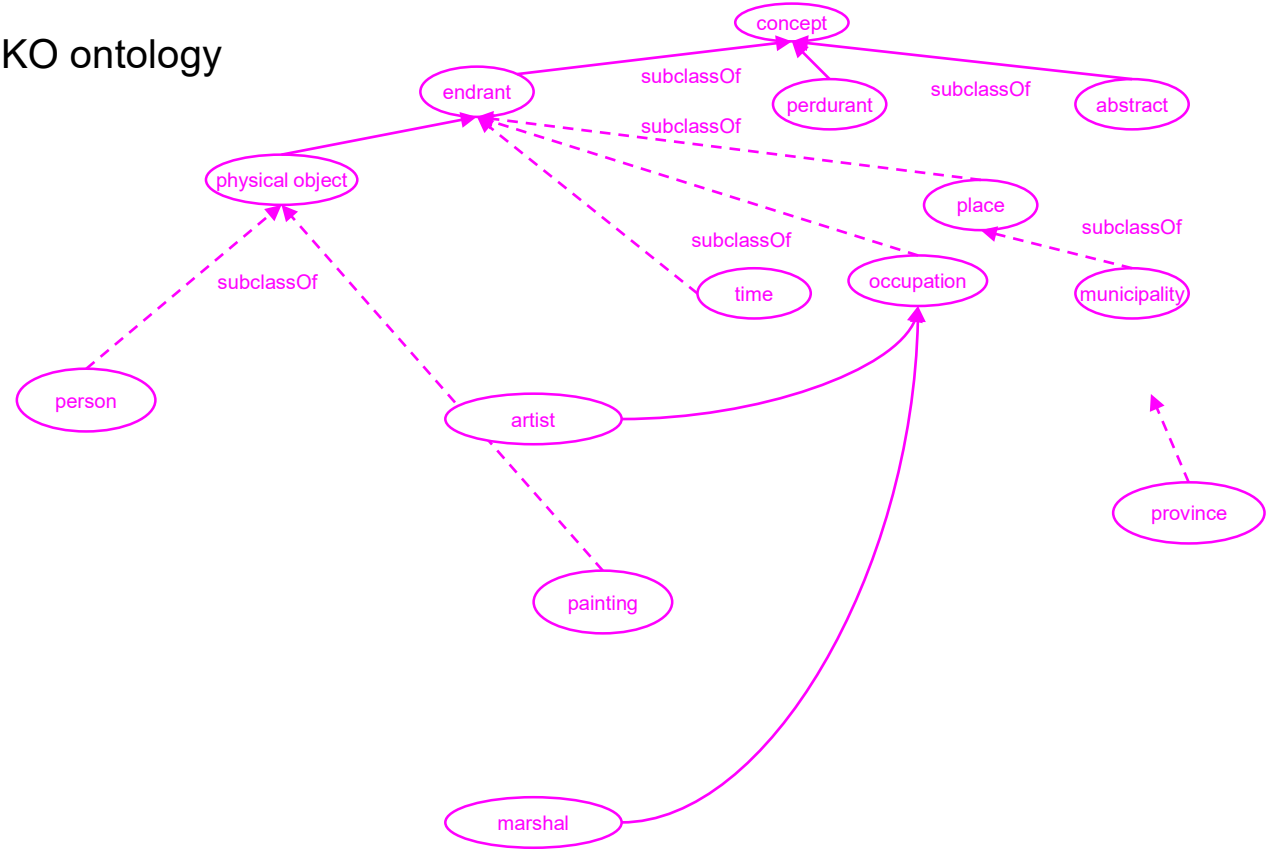
Land survey maintains place registries

municipality	province
Askainen	Varsinais-Suomi
Helsinki	Uusimaa
Lemu	Varsinais-Suomi
Turku	Varsinais-Suomi
…	

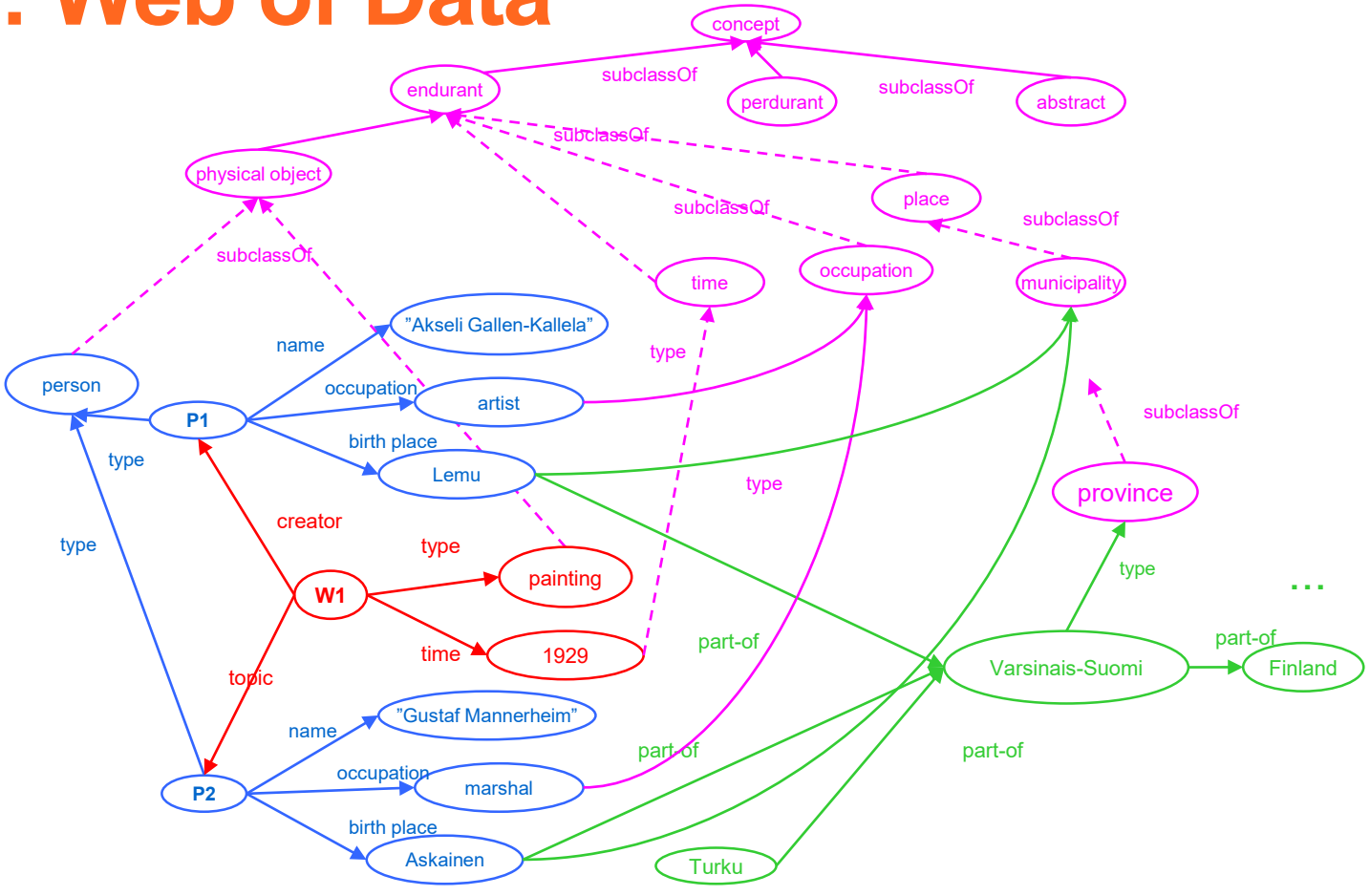


National library builds ontologies

KOKO ontology



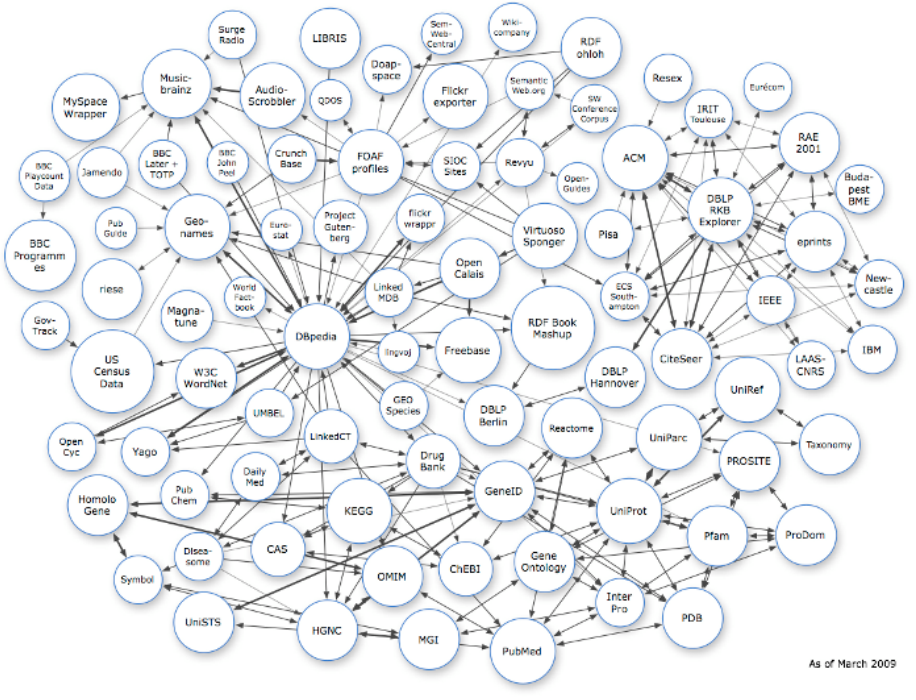
Semantic RDF graph combines them all: Web of Data



Linked Data – Web of Data

- Utilization of distributed work
- Aggregating massive cross-domain contents
- Linked Open Data thinking
- Semantic portals

http://linkeddata.org



Rule level

The idea of rules

- Semantic web semantics is based on **logic**
- Logic: “new” information can be derived from old by reasoning

Rule Markup Language RuleML

Standardized XML notation for rules

$$hasParent(?x1,?x2) \wedge hasBrother(?x2,?x3) \Rightarrow hasUncle(?x1,?x3)$$

```
<ruleml:imp>
 <ruleml:_rlab ruleml:href="#example1"/>
 <ruleml:_body>
   <swrlx:individualPropertyAtom  swrlx:property="hasParent">
     <ruleml:var>x1</ruleml:var>
     <ruleml:var>x2</ruleml:var>
   </swrlx:individualPropertyAtom>
   <swrlx:individualPropertyAtom  swrlx:property="hasBrother">
     <ruleml:var>x2</ruleml:var>
     <ruleml:var>x3</ruleml:var>
   </swrlx:individualPropertyAtom>
 </ruleml:_body>
 <ruleml:_head>
   <swrlx:individualPropertyAtom  swrlx:property="hasUncle">
     <ruleml:var>x1</ruleml:var>
     <ruleml:var>x3</ruleml:var>
   </swrlx:individualPropertyAtom>
 </ruleml:_head>
</ruleml:imp>
```

Application example: MuseumFinland recommends

Inference rules tell machine about the world

- E.g., that ”student’s cap” is related to ”parties”
- E.g., that entities are related to each other if their superclasses are related to each other
- Etc.

Based on the graph of metadata+ontologies, machine can:

- Reason interesting new relations between museum items, and
- Provide them to end users as recommendation links

Application example: MuseumFinland

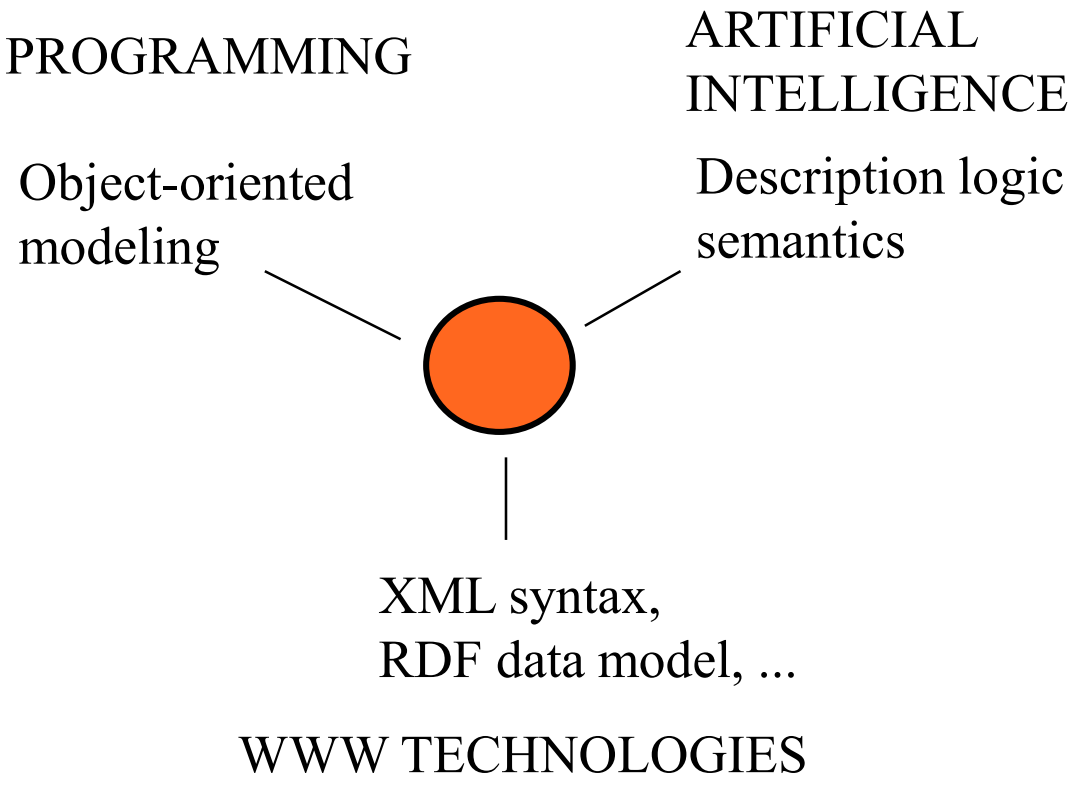
Application domains of Semantic web

- Interoperability
- Information retrieval
- Recommender systems
- Knowledge management
- E-business and web services
- Profiling and customization

…

https://www.w3.org/2001/sw/sweo/public/UseCases/

What is new?


PROGRAMMING
ARTIFICIAL INTELLIGENCE
Object-oriented modeling
Description logic semantics
XML syntax,
RDF data model, ...
WWW TECHNOLOGIES

What is the Semantic web?

Content perspective: A new metadata layer on the web describing its contents in terms of shared vocabularies, i.e., ontologies

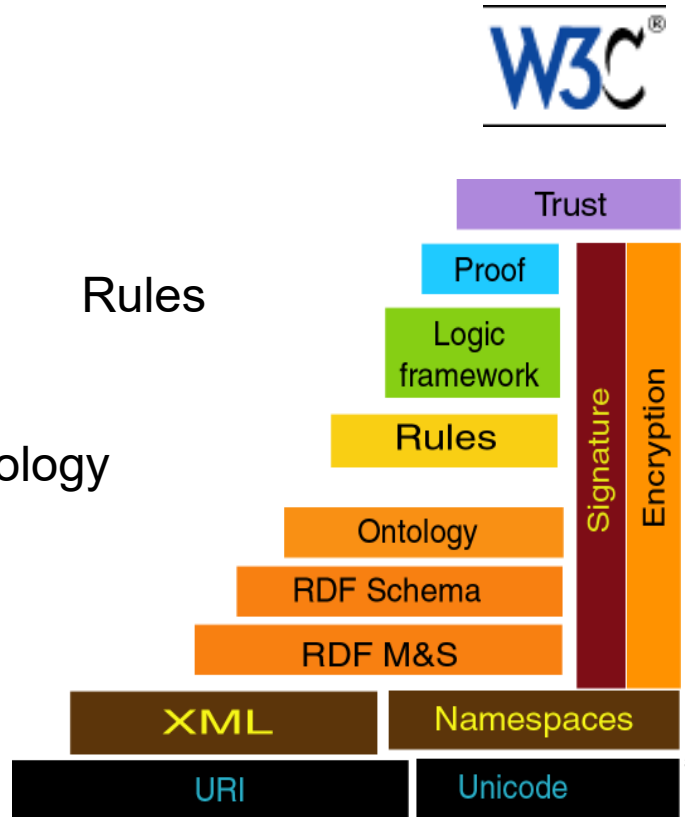
- Web as a global database system
- Web of Pages vs. Web of Data

Application perspective: Machine-understandable web

- The meaning (semantics) of contents accessible to machines
- Enables human usage
  - *Intelligent web services*
  - *Semantic interoperability*

Technological perspective:
Next layers above XML

- W3C standards:
  RDF(S), OWL , SPARQL, etc.



Questions