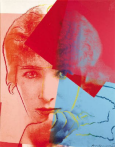
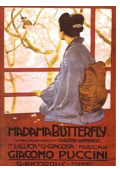




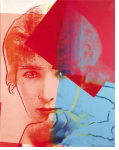





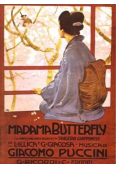
Proposals for

principles of knowledge engineering In the 21st century

Guus Schreiber
VU University Amsterdam



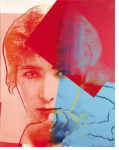


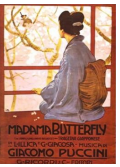


Knowledge engineering in the 20th century

- Closed systems
- Growing importance of knowledge patterns
  - Focus on patterns of problem-solving tasks
- The great divide between knowledge-engineering and knowledge-representation communities
- Protégé is prime descendant of KAW breeding ground of knowledge-engineering research







Knowledge engineering in the 21st century

- Open Web systems
- Rich availability of (new) knowledge sources
- New programming paradigms
- Ontologies have become “en vogue”

Knowledge engineering and the Semantic Web Project

- The Semantic Web is not a research discipline, but an application domain
- Knowledge-engineering research has been and still is a key driver for the Semantic Web Project
- Knowledge engineering flourishes through the multi-disciplinary cooperation within the Semantic Web Project

Hypothesis

- Semantic Web technology is in particular useful in knowledge-rich domains

or formulated differently

- If we cannot show added value in knowledge-rich domains, then it may have no value at all

This talk

Can we formulate principles for knowledge engineering in the 21st century?

Knowledge-engineering case study:

Distributed heritage collections

http://eculture.cs.vu.nl/europeana/session/search

Google

login | help | English

europeana
think culture

This is a research prototype of Europeana's semantic search engine.
Enter a search term, for example: Egypt, Rembrandt, window.

search

Collections | Thesauri

Rijksmuseum
46,038 artworks

RKD
82,781 artworks

Louvre
11,327 artworks

home disclaimer datacloud acknowledgments

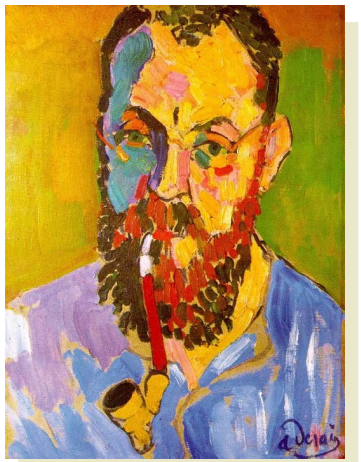
W3C RDF

powered by ClioPatria 1.0 beta 2 (18/12/2008)
SWI Prolog 5.7.10-28-gaf37c11
eculture.cs.vu.nl

The Web: resources and links

URL

Web link



URL

The Semantic Web: typed resources and links

Painting
"Woman with hat
SFMOMA

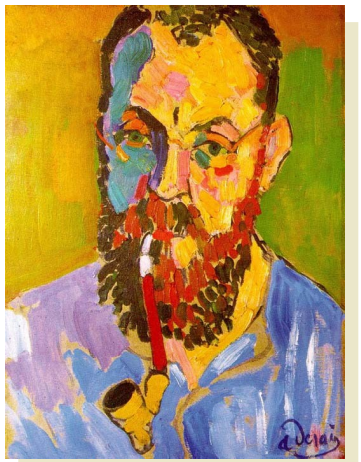
Dublin Core

creator

ULAN

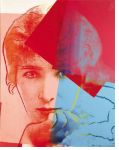
Henri Matisse

Web link

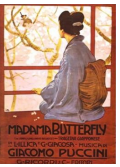


URL

URL







http://www.getty.edu/vow/ULANFullDisplay?find=matisse&role=&nation=&prev_page=1&subjectid=5000173

Research

Research Home ▸ Conducting Research ▸ Union List of Artist Names ▸ Full Record Display

Union List of Artist Names® Online
Full Record Displa

New Search Previous Page

Click the icon to view th

ID: 500017300

Matisse, Henri (Frenc

Names
Matis
Henr
Matis
Matis

Nationa
Frenc

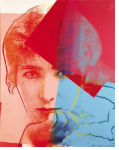
Roles:
artist
painte
printr
sculpt
desig
writer

Related People or Corporate Bodies:
apprentice was .... Jolin, Einar 1911-1913
..... (Swedish painter, 1890-1990) [500014093]
parent of .... Duthuit, Marguerite Matisse
..... (French painter, born ca. 1900) [500075813]
patron was .... Barnes, Dr. Albert C.
..... (American collector, 1872-1951) [500057478]
student of .... Cormon, Fernand
..... (French painter and teacher, 1845-1924) [500115385]
student of .... Moreau, Gustave
..... (French painter, 1826-1898) [500115776]

Roles:
artist (**preferred**)
painter
printmaker
sculptor
designer
writer

Gender: male

Birth and Death Places:
Born: Le Cateau-Cambrésis (Nord, Nord-Pas-de-Calais, France) (inhabited place)
Died: Nice (Alpes-Maritimes, Provence-Alpes-Côte d'Azur, France) (inhabited place)








http://www.getty.edu/vow/TGNHierarchy?find=netherlands&place=&nation=&prev_page=1&english=Y&subj

Research

Research Home ▸ Conducting Research ▸ Thesaurus of Geographic Names ▸ Hierarchy Display

Getty Thesaurus of Geographic Names® Online Hierarchy Display

New Search Previous Page Help

Vernacular Display | **English Display**

View Selected Records Clear All

Click the icon to view the hierarchy.
Check the boxes to view multiple records at once.

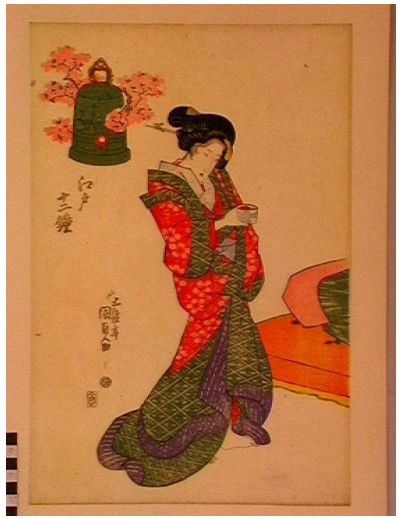
- Top of the TGN hierarchy (hierarchy root)
- .... World (facet)
- ..... Europe (continent)
- ..... Netherlands (nation)
- ..... ***[view physical features]***
- ..... Aarkanal (canal)
- ..... Afsluitdijk (dam)
- ..... Alblasserwaard (general region)
- ..... Altena, Land van (general region)
- ..... Amstelland (general region)
- ..... Amsterdam Rijn Kanaal (canal)
- ..... Aruba (dependent state) [N]
- ..... Bernisse Molen (mill center)
- ..... Biesbos (general region)
- ..... Brouwersdam (dam)
- ..... Calandkanaal (canal)
- ..... Delfland (general region)

The myth of a unified vocabulary

- In large virtual collections there are always multiple vocabularies
  - In multiple languages
- Every vocabulary has its own perspective
  - You can't just merge them
- But you can use vocabularies jointly by defining a limited set of links
  - "Vocabulary alignment"
- It is surprising what you can do with just a few links

Power of (simple and partial) vocabulary alignments

"Tokugawa"



AAT style/period
Edo (Japanese period)
Tokugawa

AAT is Getty's Art & Architecture Thesaurus

SVCN period
Edo

SVCN is local in-house ethnology thesaurus

Knowledge engineering activities for distributed heritage collections

Vocabulary interoperability

Vocabulary aligment

Metadata schema interoperability

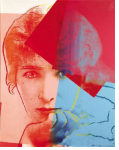
Metadata enrichment

Semantic search

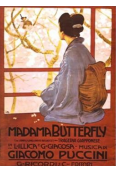
Semantic annotation

Levels of interoperability

- Syntactic interoperability
  - using data formats that you can share
  - XML family is the preferred option
- Semantic interoperability
  - How to share meaning / concepts
  - Technology for finding and representing semantic links

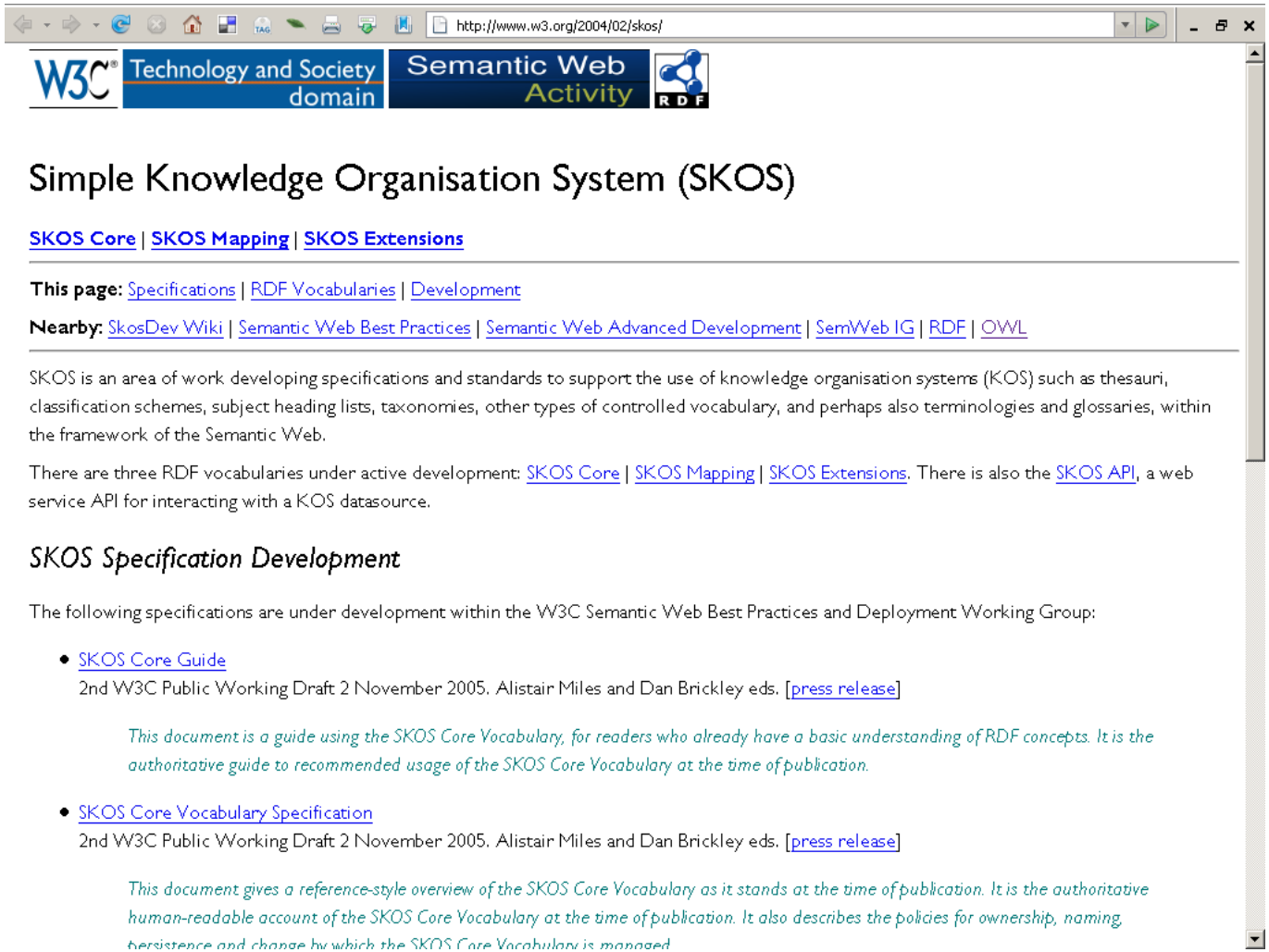




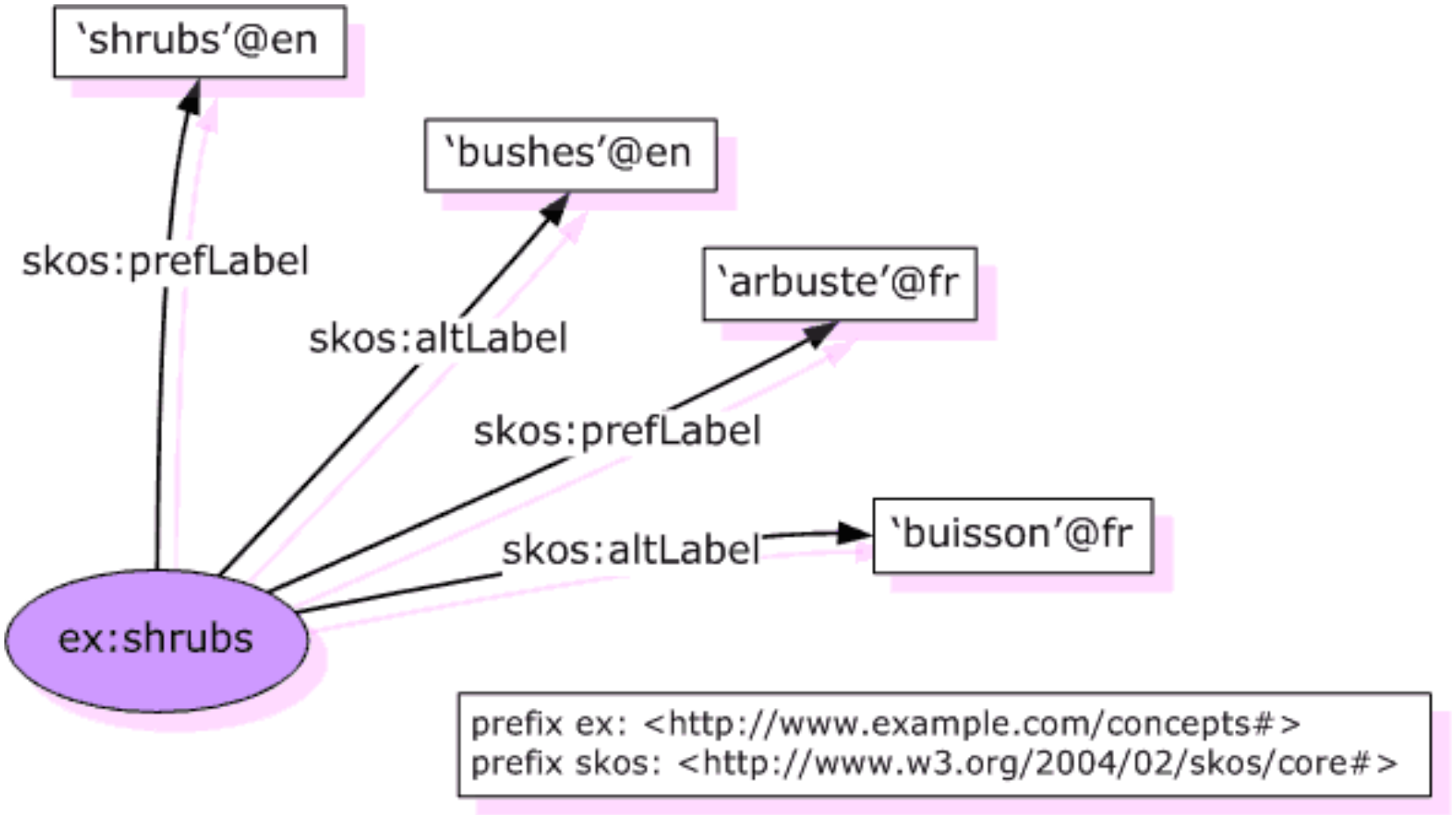





Vocabulary interoperability: an ad for SKOS

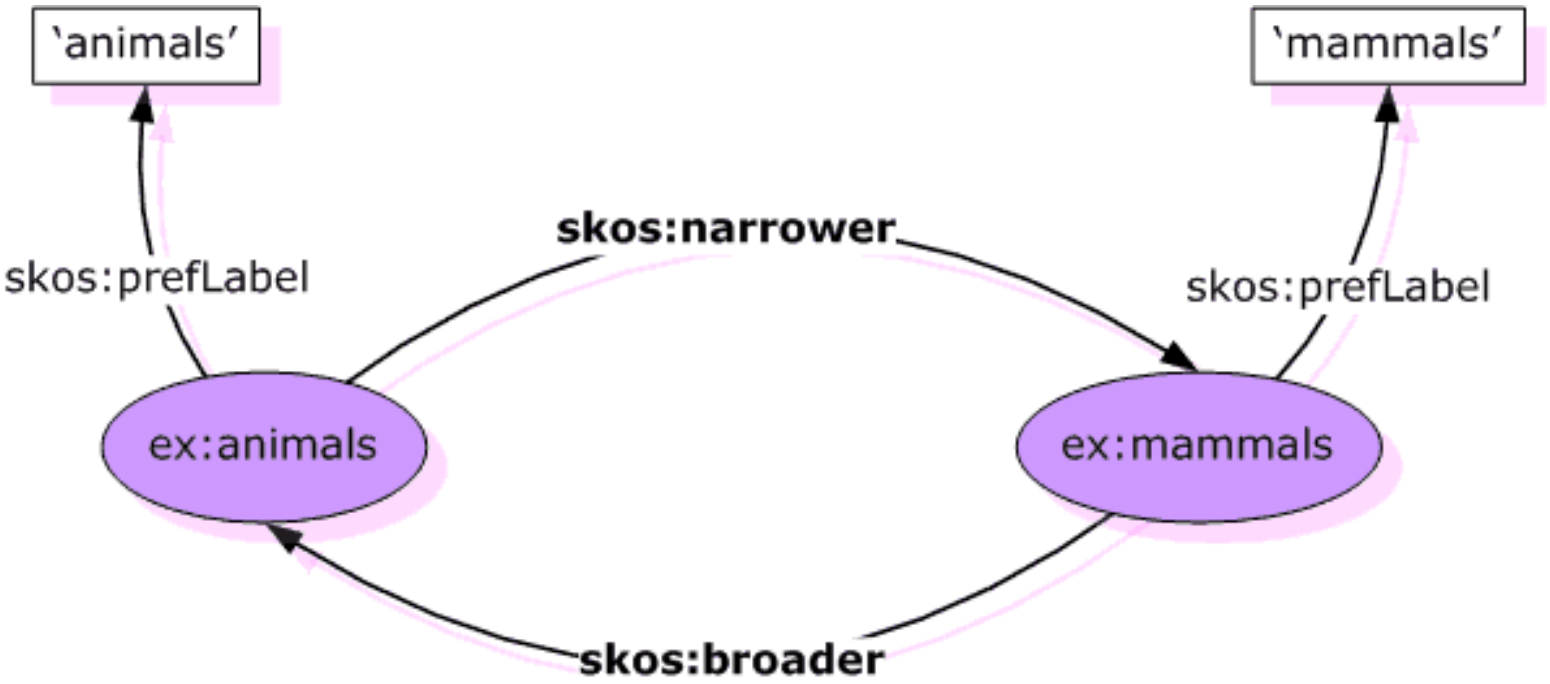


Multi-lingual labels for concepts



Semantic relation: broader and narrower

- No subclass semantics assumed!




```
prefix ex: <http://www.example.com/concepts#>
prefix skos: <http://www.w3.org/2004/02/skos/core#>
```

Issues in specification of SKOS semantics

- SKOS should cover a large range of “vocabularies”, “thesauri”, “terminologies”, “classification schemes”, etc.
- Therefore: objective was to define the minimal semantics
- Leave hooks for specializations
- See SKOS Primer for examples

core (http://www.w3.org/2004/02/skos/core) – [http://www.w3.org/2004/02/skos/core#]

core

Classes | Object Properties | Data Properties | Individuals | OWLViz | DL Query

Object properties: exactM

- label
- hasTopConcept
- inScheme
- member
- memberList
- note
- semanticRelation
  - broaderTransitive
    - broader
      - broadMatch
  - mappingRelation
    - broadMatch
    - closeMatch
      - exactMatch
    - narrowMatch
    - relatedMatch
  - narrowerTransitive
  - related

Annotations | Object Property Usage

Annotations: exactMatch

Annotations

comment

"skos:exactMatch is disjoint with each of the properties skos:broadMatch and

Characteristics: exact

- [ ] Functional
- [ ] Inverse functional
- [x] Transitive
- [x] Symmetric
- [ ] Asymmetric
- [ ] Reflexive
- [ ] Irreflexive

Description: exactMatch

Domains (intersection)

Ranges (intersection)

Equivalent object properties

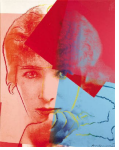
Super properties

closeMatch

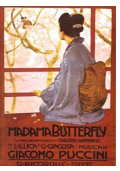
Inverse properties

Example requirement

- Being able to define relations between labels
  - “WHO” is an acronym of “World Health Orgnization” (in English)
  - “WGO” is an acronym of “Wereldgezonheidsorganisatie” (in Dutch)
- Treat llexical labels as resources with URI?
  - But many simple vocabularies don't needs this
  - Would be burden







skos-xl (http://www.w3.org/2008/05/skos-xl) – [http://www.w3.org/2008/05/skos-xl#]

skos-xl (http://www.w3.org/2008/05/skos-xl)

Active Ontology | Entities | Classes | Object Properties | Data Properties | Individuals | OWLViz | DL Query

Asserted class hierarchy

Asserted class hierarchy: Label

- Thing
  - Collection
  - Concept
  - ConceptScheme
  - Label

Object property hierarchy

Object properties: prefLabel

- altLabel
- hiddenLabel
- labelRelation
- prefLabel

Annotations | Object Property Usage

Annotations: prefLabel

Annotations

comment

"If C skosxl:prefLabel L and L skosxl:literalForm V, then X skos:prefLabel V."@en

isDefinedBy

Characteristics: prefLa

- Functional
- Inverse functional
- Transitive
- Symmetric
- Asymmetric
- Reflexive
- Irreflexive

Description: prefLabel

Domains (intersection)

Ranges (intersection)

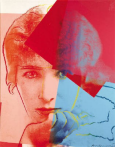
Label

Equivalent object properties

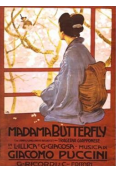
Super properties

Inverse properties

Disjoint properties

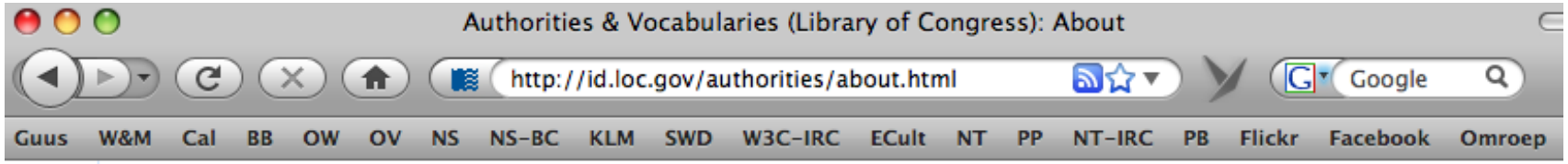






Large organizations have adopted SKOS



How it works

Users and machines simply request the URI of interest over HTTP. For example, to access the data value "World Wide Web" in the Library of Congress Subject Headings, one would request this URI:

- http://id.loc.gov/authorities/sh95000541#concept

When requesting this URI, users have mechanisms for specifying how they want to serialize the data they wish to access. These include common RDF serializations carrying Simple Knowledge Organization System (SKOS) metadata, and Javascript Object Notation (JSON).

See the Technical Center for more details.

Benefits

For users (whether human or machine):

- Access to data at no cost.
- Granular access to individual data values.
- Ability to download entire controlled vocabularies and the values within them in numerous formats.

Metadata schema interoperability

- Cultural heritage has an abundance of metadata format standards
  - Dublin Core, VRA (images), MARC, ...
- Current practice: XSLT transformations (and similar)
- owl:EquivalentProperty and rdfs:subPropertyOf are well suited for defining partial alignments between schemata

Aligning VRA with Dublin Core

- VRA is specialization of Dublin Core for visual resources
- VRA properties “material.medium” and “material.support” are specializations of Dublin Core property “format”

vra:material.medium
rdfs:subPropertyOf dc:fotmat .
vra:material.support
rdfs:subPropertyOf dc:format .

Strong pojnt of OWL

"*For collection X the range of dc:creator is a value from the ULAN thesaurus*"

=> Define an owl:Restriction for resources in X which specifies a corresponding *local* range restriction for the dc:creator value

Built-in overcommitment in OWL DL

Is dc:creator an owl:DatatypeProperty or an owl:ObjectProperty?

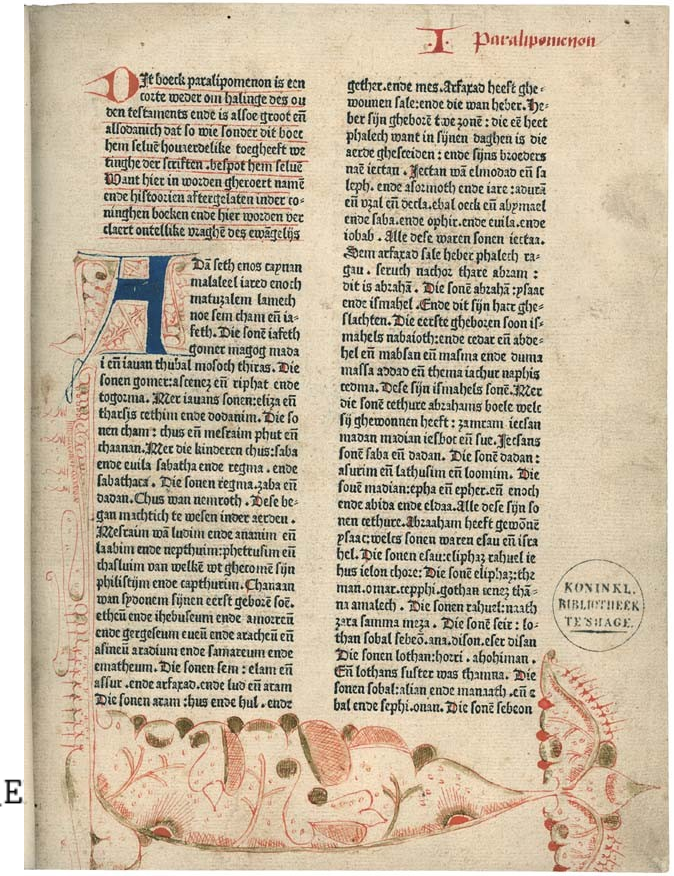
Answer: depends on the context!

The minimal commitment is:

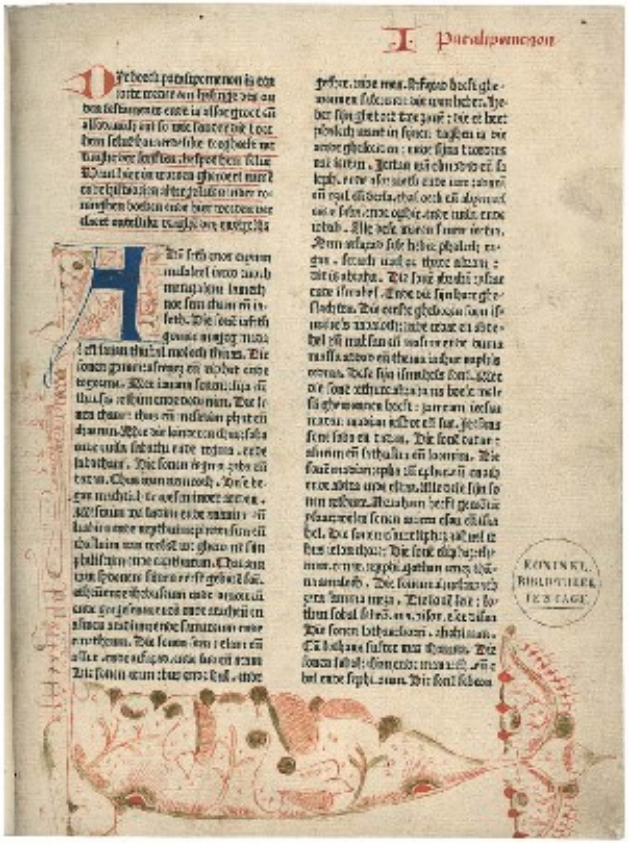
dc:creator rdf:type rdf:Property .

Metadata enrichment

```
<inm:Record>
    <inm:NUMMER>6</inm:NUMMER>
    <inm:TITEL>Delftse Bijbel...</inm:TITEL>
    <inm:TITEL_EN>Delft Bible...</inm:TITEL_EN>
    <inm:MAKER>Yemantszoon, Mauricius : d</inm:MAKER>
    <inm:OBJECT>tekstbladzijde</inm:OBJECT>
    <inm:TECHNIEK>boekdruk</inm:TECHNIEK>
    <inm:DATERING>10 jan. 1477</inm:DATERING>
    <inm:CLASSIFICATIE>D</inm:CLASSIFICATIE>
    <inm:ORIGINEEL>Bijbel. Oude
        Testament...</inm:ORIGINEEL>
    </inm:REPRODUCTIE>
    <inm:TWNAAM/>
    <inm:TWOND>typografische vormgeving</inm:TWOND>
    <inm:TWOND>bijbels</inm:TWOND>
    <inm:TWGEO>Delft</inm:TWGEO>
    <inm:OMSCHRIJVING>Eerste bijbel die in het
        Nederlands verscheen...</inm:OMSCHRIJVING>
    <inm:OMSCHRIJVING_EN>The first Bible to
    appear in the Dutch language...</inm:OMSCHRIJVING_E
    <inm:AFMETINGEN>27 x 20 cm</inm:AFMETINGEN>
    ...
</inm:Record>
```



Replace strings with concepts: quality issues of automatic extraction



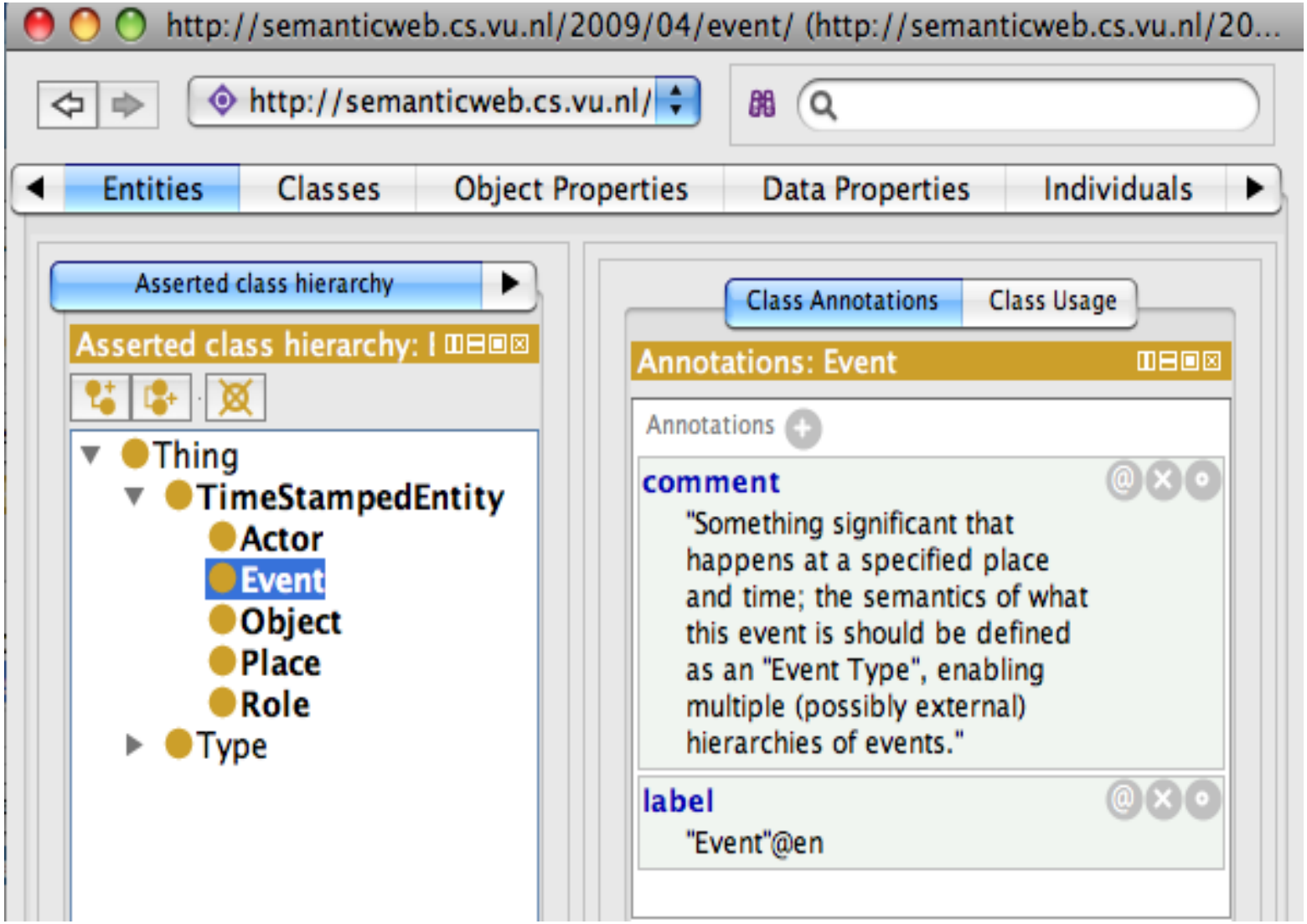
Description:

classificatie	Geschiedenis van de boekdrukkunst;
drukker	Meer, Jacob Jacobszoon van der; Yemantszoon, Mauricius;
origineel	Bijbel. Oude Testament. - Delft: Jacob Jacobszoon van der Meer en Mauricius Yemantszoon, 10 jan. 1477, dl. 2, p. 1;
Date	10 jan. 1477;
Description	The first Bible to appear in the Dutch language, known as the Delft Bible. It consists of the Old Testament only and is an anonymous adaptation of the - again anonymous - History Bible of 1360. It is an example of an incunabulum where the hand-written book still served as an example for lay-out and design. Contrary to many other incunabula, the place of origin, the names of the printers and even the day of its completion are mentioned in the colophon.;
Measurements.Dimensions	27 x 20 cm;
rights.copyright	Den Haag Koninklijke Bibliotheek;
Source	Bibliopolis;
Subject	bibles; incunabula; initials; ornamental borders; rubrications; typographical design;
subject.geographicPlace	Delft;
Technique	letterpress printing;
Title	Delft Bible, printed in Delft by Jacob Jacobszoon van der Meer and Mauricius Yemantszoon, 1477;
Type	tekstbladzijde;
type	Work;

Used as value to describe other resources:

BBB_169E56_1477_P1.JPG;	relation.depicts

Hot issue: event modelling
“what is happening on an image?”

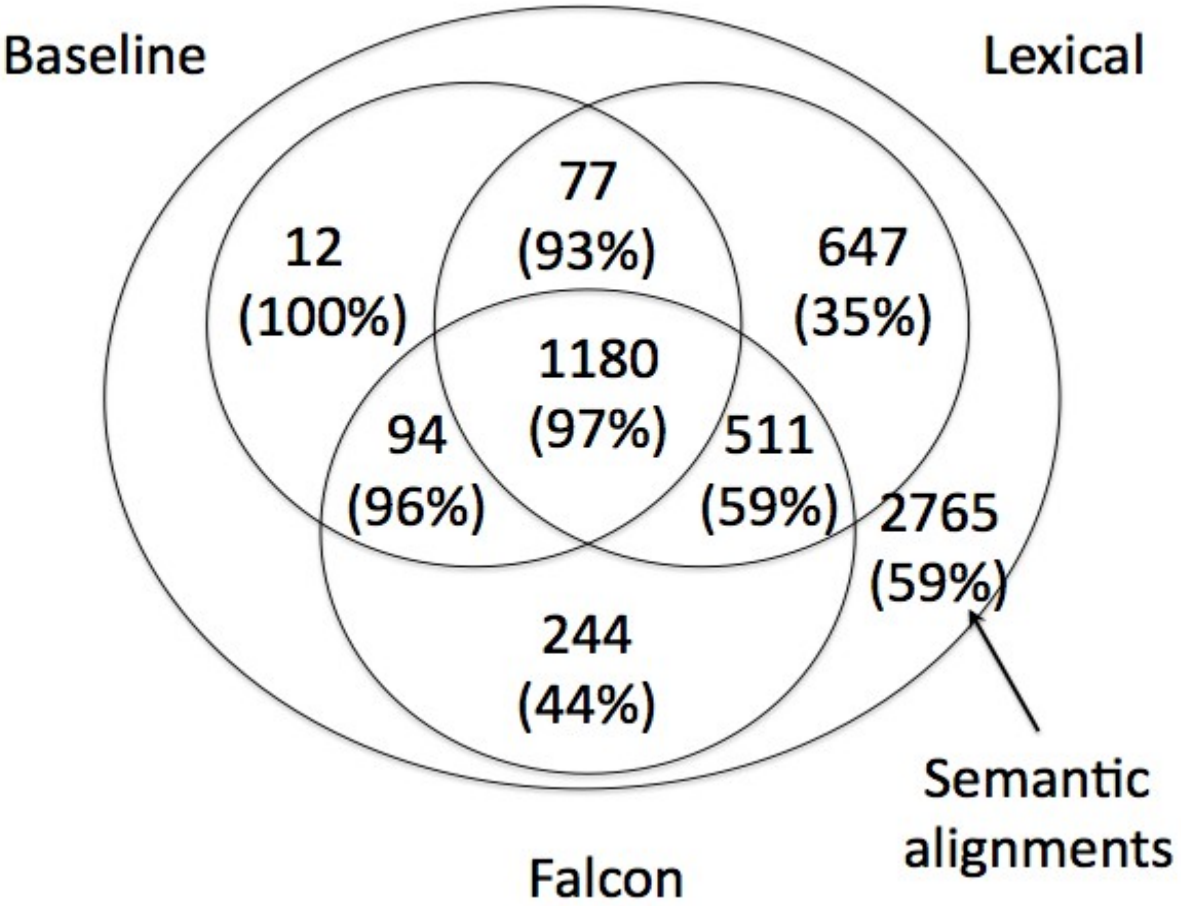


Vocabulary alignment

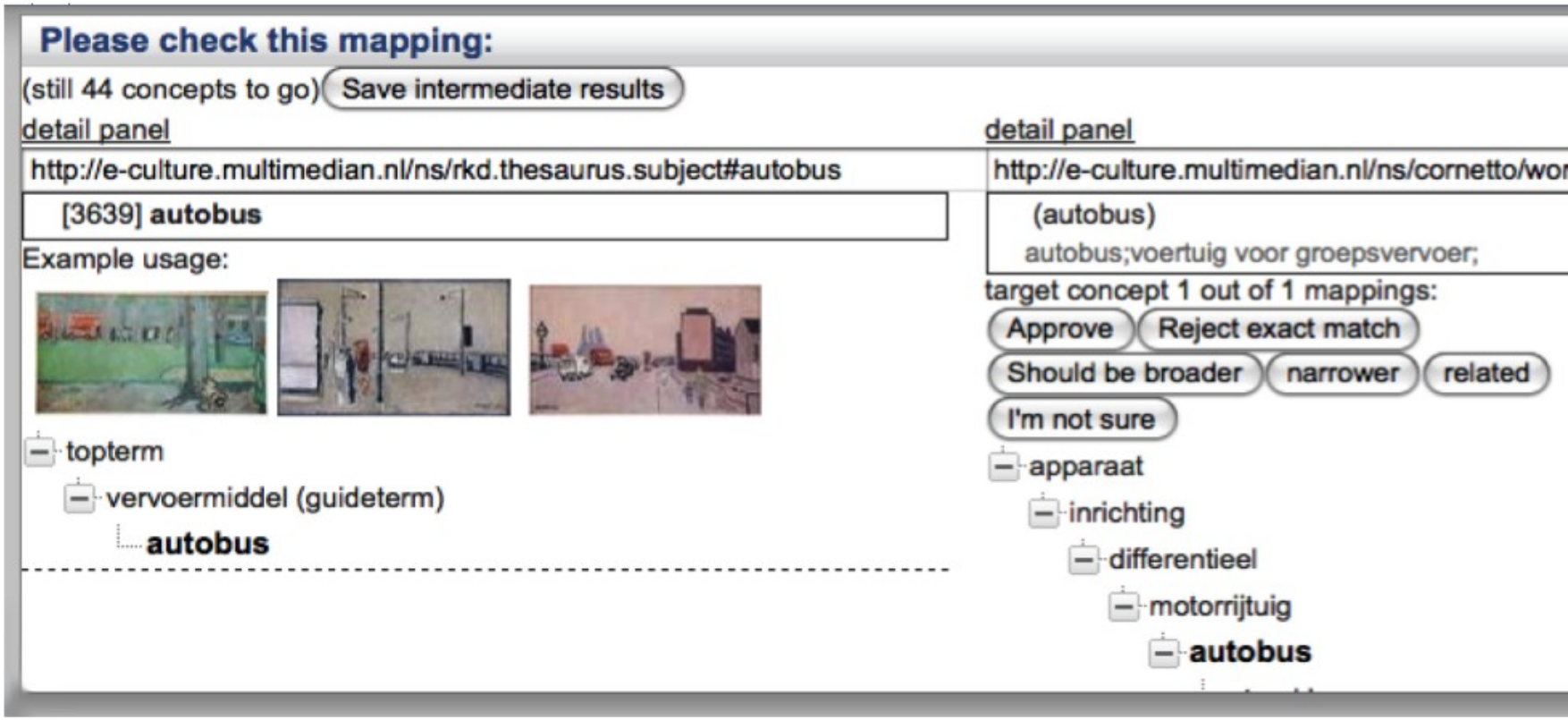
- Learning relations between art styles in AAT and artists in ULAN through NLP of art historic texts
  - "Who are Impressionist painters?"

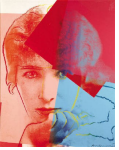
Artist Name	*IS*	*In GS*
edgar degas	0.0699	1
edouard manet	0.0548	1
pierre-auguste renoir	0.0539	1
morisot, berthe	0.0393	1
gogh, vincent van	0.0337	0
cassatt, mary	0.0318	1
cezanne, paul	0.0302	1

Results of automatic alignment vary in quality

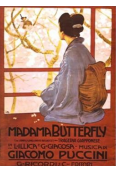


Partial human engineering and/or evaluation is often time/cost effective









Semantic search: clustering and cluster-order principles

cultural heritage on the web :: humans

Pictures and metadata of artworks

Works created by (95)


Three Women
Picasso, Pablo


Glass of Absinthe
Picasso, Pablo


Accordionist
Picasso, Pablo


The Aficionado
Picasso, Pablo

Works by professionally related artist (31)


Fruit Dish, Ace of Clubs
Braque, Georges

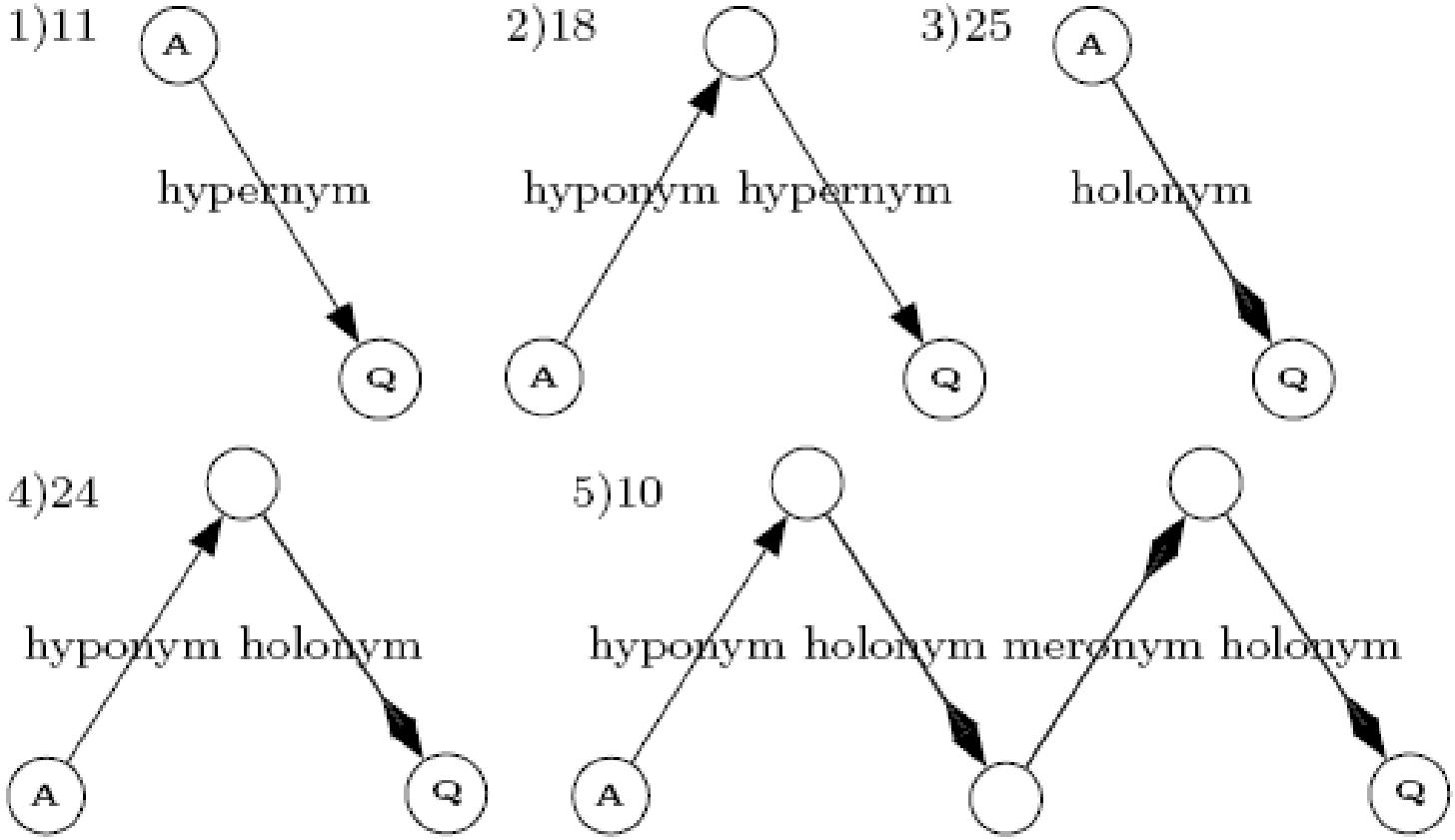

Man with a Violin
Braque, Georges


Bottle, Newspaper, Pipe,
Braque, Georges

Still Life BACH
Braque, Georges

Works created by artists with style/period Surrealist also used by artist (6)

Research topic: semantic patterns which increase recall without sacrificing precision



Semantic annotation: granularilty level

The execution of Johan van Oldenbarnevelt

RP-P-OB-77.320

Print with a scene

update cancel

Who Historical persons

person

Oldenbarnevelt, Johan van x

What Iconclass (en), WordNet (en), events (nl)

(mythological) concept, object or event

beheading x

Where Name of place or region

geographical place

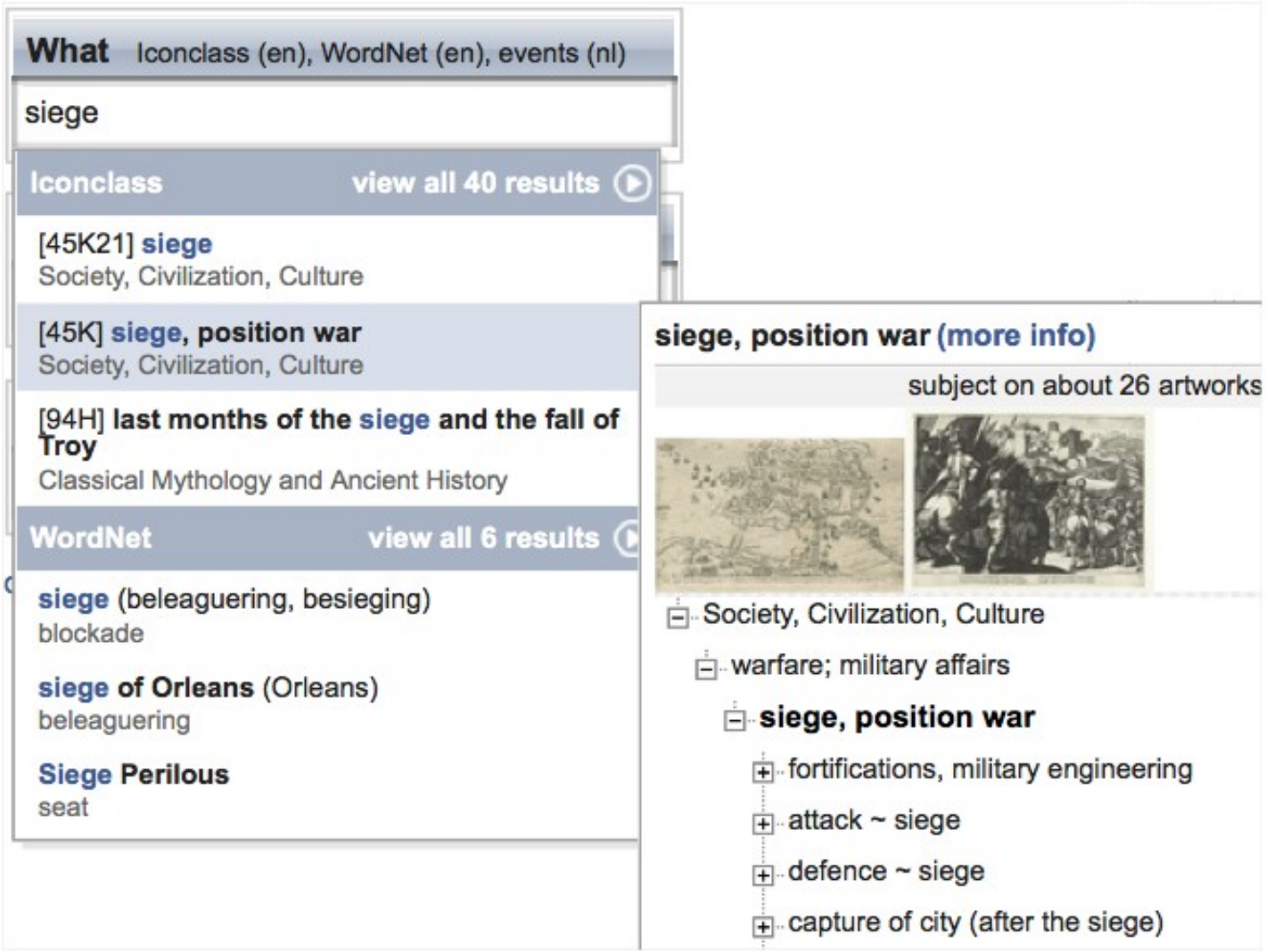
Den Haag x

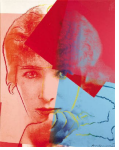
When Date, year or period

enter date

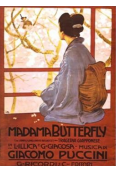
done | cancel

Autocompletion and disambiguation issues









Principles for knowledge engineering on the Web

Principle 1: Be modest!

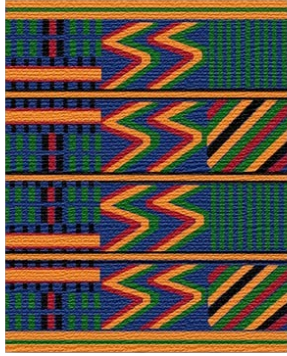
- Ontology engineers should refrain from developing their own idiosyncratic ontologies
- Instead, they should make the available rich vocabularies, thesauri and databases available in an interoperable (web) format
- Initially, only add the originally intended semantics

Principle 2: Think large!

Doug Lenat

"Once you have a truly massive amount of information integrated as knowledge, then the human-software system will be superhuman, in the same sense that mankind with writing is superhuman compared to mankind before writing."

Principle 3: Develop and use patterns!



- Don’t try to be (too) creative
- Ontology engineering should not be an art but a discipline
- Patterns play a key role in methodology for ontology engineering
- See for example patterns developed by the W3C Semantic Web Best Practices group

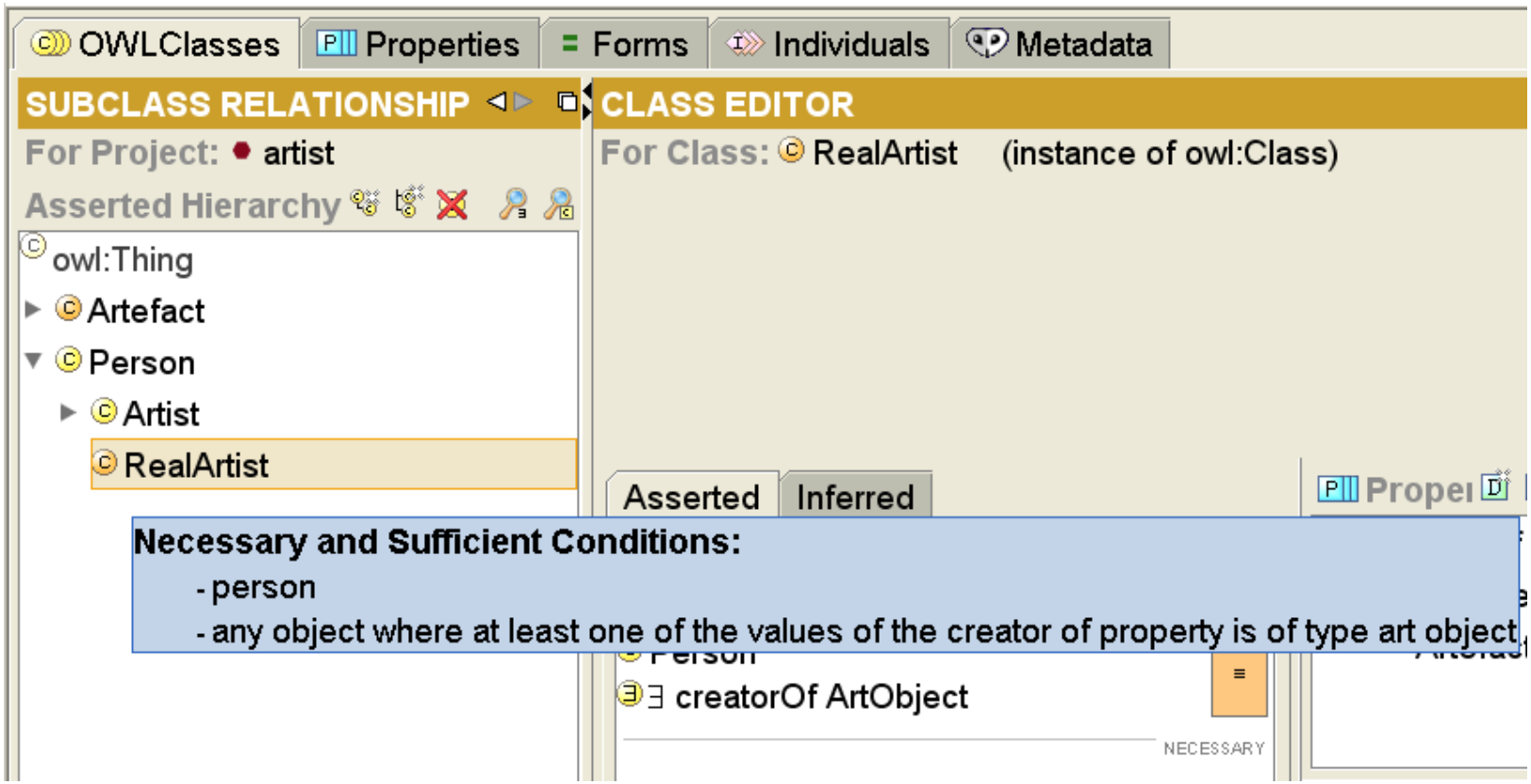
  http://www.w3.org/2001/sw/BestPractices/
- SKOS can also be considered a pattern

Principle 4: Don’t recreate, but enrich and align

- Techniques:
  - Learning ontology relations/mappings
  - Semantic analysis, e.g. OntoClean
  - Processing of scope notes in thesauri

  - Manual evaluation sometimes key

Principle 5: Beware of ontological over-commitment!

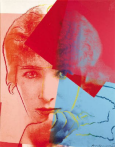


Principle 6: Specifying a data model in OWL does ot make it an ontology!

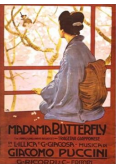
- Papers about your own idiosyncratic “university ontology” should be rejected at conferences
- The quality of an ontology does not depend on the number of OWL constructs used

Principle 7: Required level of formal semantics depends on the domain!

- In our semantic search we use three OWL constructs:
  - owl:sameAs, owl:TransitiveProperty, owl:SymmetricProperty
- But cultural heritage has is very different from medicine and bioinformatics
  - Don't over-generalize on requirements for e.g. OWL







Thank you!

Acknoledgments: slides and ideas from many co-workers within VU, Amsterdam and KE and SW communities, in particular Lora Aroyo, Michiel Hildebrand, Antoine IsaacJacco van Ossenbruggen, Anna Tordai, Jan Wielemaker.