

Publishing vocabularies on the Semantic Web using SKOS - with examples from Cultural Heritage

SemAst 2009

Antoine Isaac

Vrije Universiteit Amsterdam, National Library of the Netherlands
aisaac@few.vu.nl

Preamble: acknowledgements

- Participants of the W3C Semantic Web Deployment Working Group

<http://www.w3.org/2006/07/SWD/>

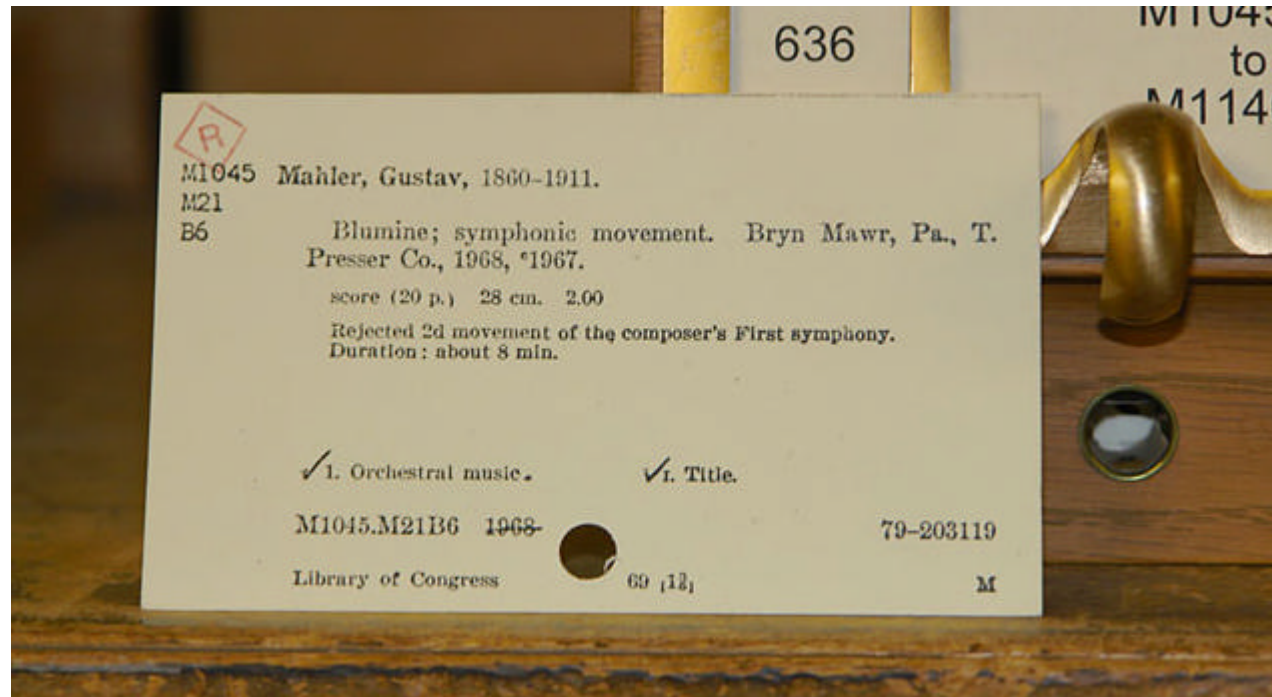
Talk overview

- Problem statement
- SKOS: a guided tour
- Some more details
- And/or demos

Knowledge Organization Systems? (KOS)

- Domain-specific vocabularies
 - Medicine: UMLS, SNOMED, MESH, Galen
 - Art history: AAT, ULAN
 - Geography: TGN
 - Food: Agrovoc
 - Libraries: LCSH, DDC, UDC
- Generic vocabularies
 - Lexical vocabularies: WordNet
 - Country codes, ...

Example: CH Metadata



- Use of controlled **documentary languages**
 - Thesauri, classification systems, subject heading lists

Example: Iconclass

Browse by subject

using the iconclass classification system

0 Abstract, Non-representational Art

1 Religion and Magic

2 Nature

25 earth, world as celestial body

25F animals [show images >25](#)

25F3 birds [show images >25](#)

25F31 groups of birds [show images < 5](#)

25F32 song-birds [show images >25](#)

25F33 predatory birds [show images >25](#)

25F34 owls [show images < 25](#)

25F35 ornamental birds [show images < 25](#)

25F36 water-birds [show images >25](#)

25F37 shore-birds and wading-birds [show images >25](#)

25F38 walker and runner birds [show images < 25](#)

25F39 other birds [show images >25](#)

3 Human Being, Man in General

4 Society, Civilization, Culture

5 Abstract Ideas and Concepts

6 History

7 Bible

8 Literature

9 Classical Mythology and Ancient History

◆ Show only notations used in the manuscript database

Show full Iconclass hierarchy

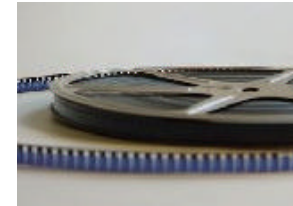
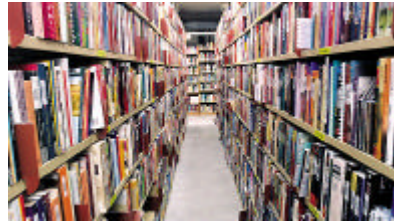
Search by keyword

Search



Why are we interested in that now?

- Museums, libraries, archives



- Crucial technological evolutions
 - Digitization
 - Web
- Trend: (web) access to CH objects from different collections

Europeana.eu case



[My Europeana](#)

[Communities](#)

[Partners](#)

[Timeline \(beta\)](#)

[Thought lab](#)

[Choose a language](#)

























***Providing access to 2 million objects
from 50 institutions***

Search

[Advanced search](#)



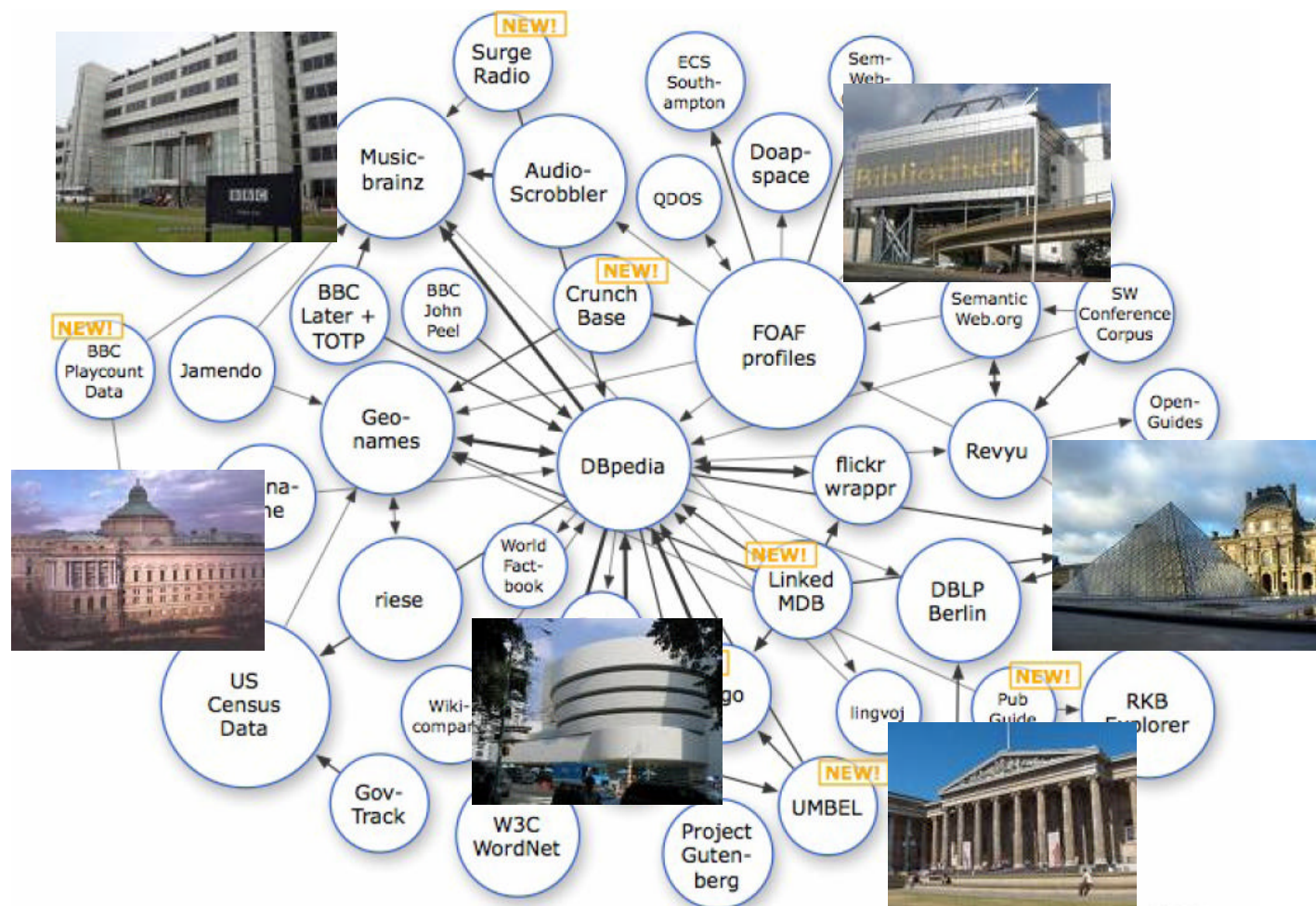
Matches for: amsterdam

All	Texts (3,572)	Images (68,727)	Videos (304)	Sounds (51)
Results 13 - 24 of 72,634 Page: 1 2 3 4 5 6 7 8 9 10 ← →				
				
AMSTERDAM 1940 Het Geheugen van Nederlan... 	Amsterdam Réquillart Bruno (né en 1947) , 1977 Culture.fr/collections 	Amsterdam.. Institut national de l'Au... 	Amsterdam Fotograaf: J. Doeser, Laren 1959 Het Geheugen van Nederlan... 	
				
Amsterdam Fotograaf: J. Doeser, Laren 1959 Het Geheugen van Nederlan... 	Amsterdam Fotograaf: J. Doeser, Laren 1959 Het Geheugen van Nederlan... 	Amsterdam Fotograaf: J. Doeser, Laren 1959 Het Geheugen van Nederlan... 	Montre ronde : La Foi HEMME Cornelis Culture.fr/collections 	
				
Montre ronde : femme alla... GERRIT KNIP & ZON 1750 Culture.fr/collections 	Amsterdam heeft een eigen... 2004 Het Geheugen van Nederlan... 	Kiezen of Deelen! -f zuin... 1940 Het Geheugen van Nederlan... 	Past op! Weest zuinig met... 1940 Het Geheugen van Nederlan... 	

CH metadata and SW

- To implement this, metadata is key
- Problems
 - how to access seamlessly different collections?
 - how to allow other applications to exploit (knowledge about) CH objects?
 - can solutions be *flexible*?
- ***The Semantic Web can be useful for this***

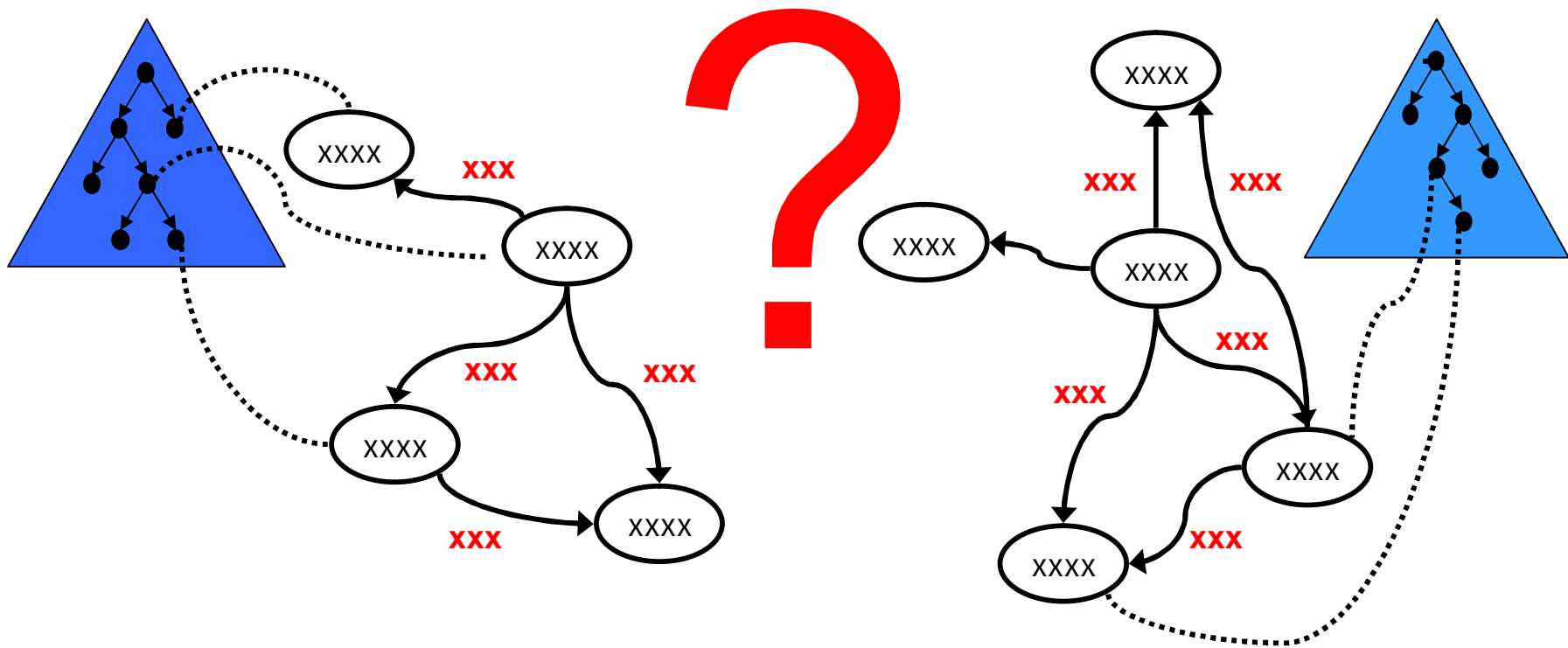
Can we have that for the CH metadata?



<http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>

SKOS Focus:

how to port KOSs to the SW?



SKOS

- Simple Knowledge Organization System

SKOS

- Observation: there are many KOS models/formats:
- But also common features, used by typical classes of applications
- SKOS is a model to represent KOSs on the SW in a *simple* way
 - Ontology for **concepts**

Not owl:Class?

- Ontologizing KOSs is possible, but:
 - they are large
 - and have loose semantics
 - `Car wheel` BroaderTerm `Car`
- Existing information can be ported as such
 - loose semantics can be useful for many applications!
 - Search, annotation

SKOS

- Concepts and Concept Schemes
- Lexical properties
- Semantic relations
- Notes

Thesaurus example

animals

NT (*narrower term*) cats

cats

UF (*used for*) domestic cats

RT (*related term*) wildcats

BT (*broader term*) animals

SN (*scope note*) used only for domestic cats

domestic cats

USE cats

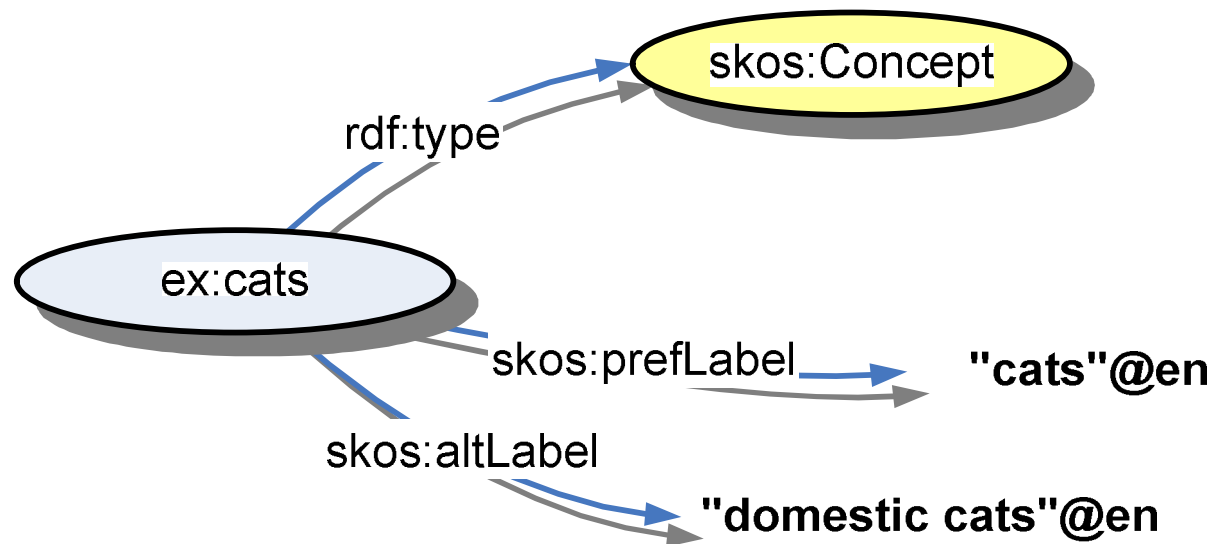
wildcats

ISO 2788 model

SKOS concepts and labels

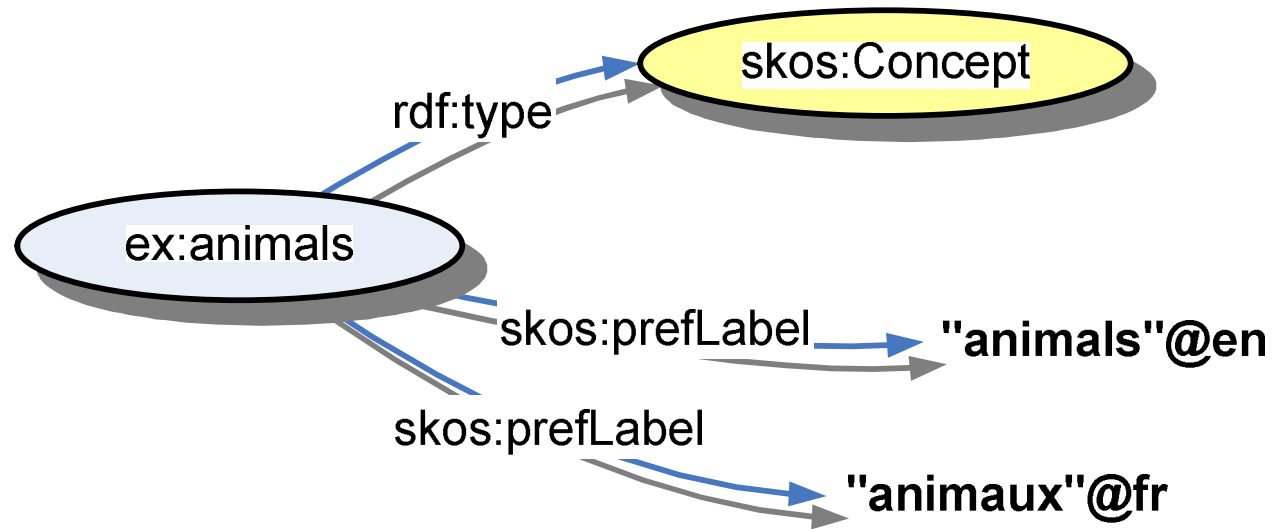
cats

UF (*used for*) domestic cats



- SKOS is concept-oriented
 - Concepts are first-order resources
 - Labels are RDF literals

(Multilingual) labels

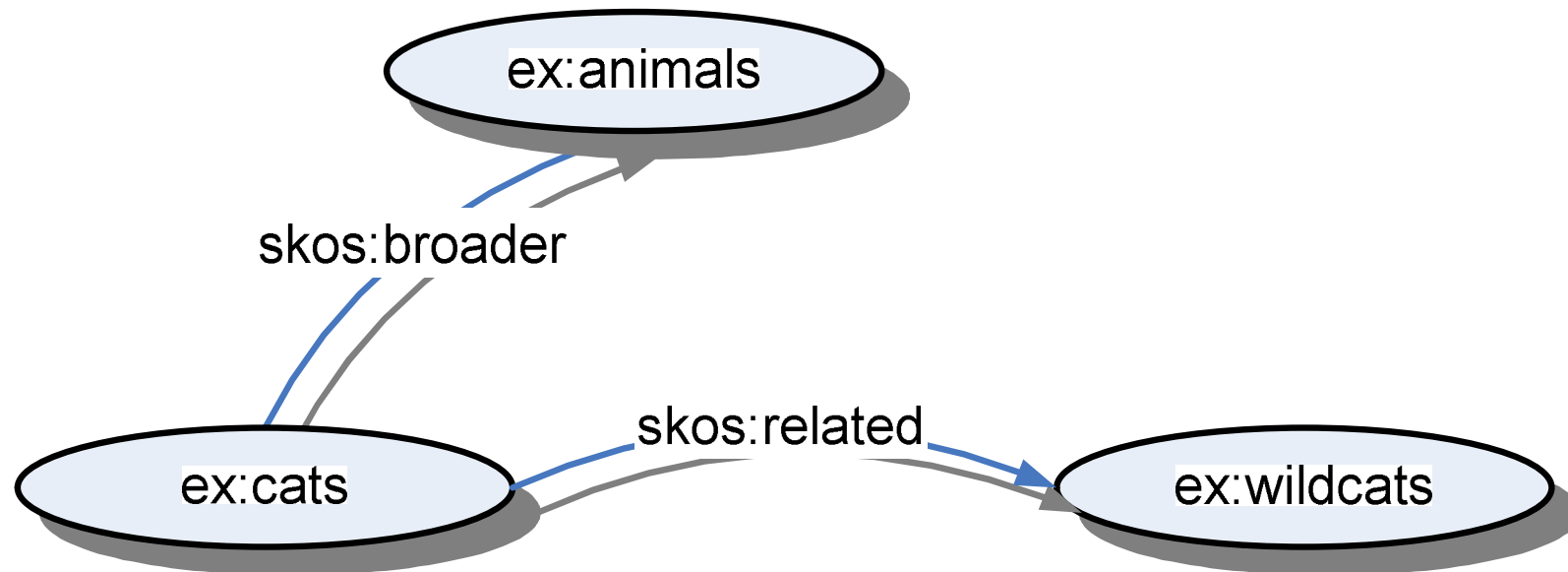


Semantic relations

cats

RT (*related term*) wildcats

BT (*broader term*) animals



Documenting concepts

skos:note

|

+-- skos:definition

|

+-- skos:scopeNote

|

+-- skos:example

|

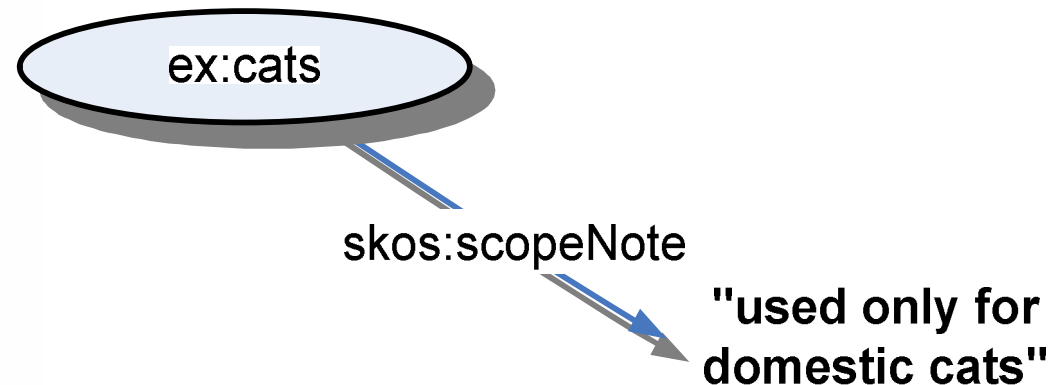
+-- skos:historyNote

|

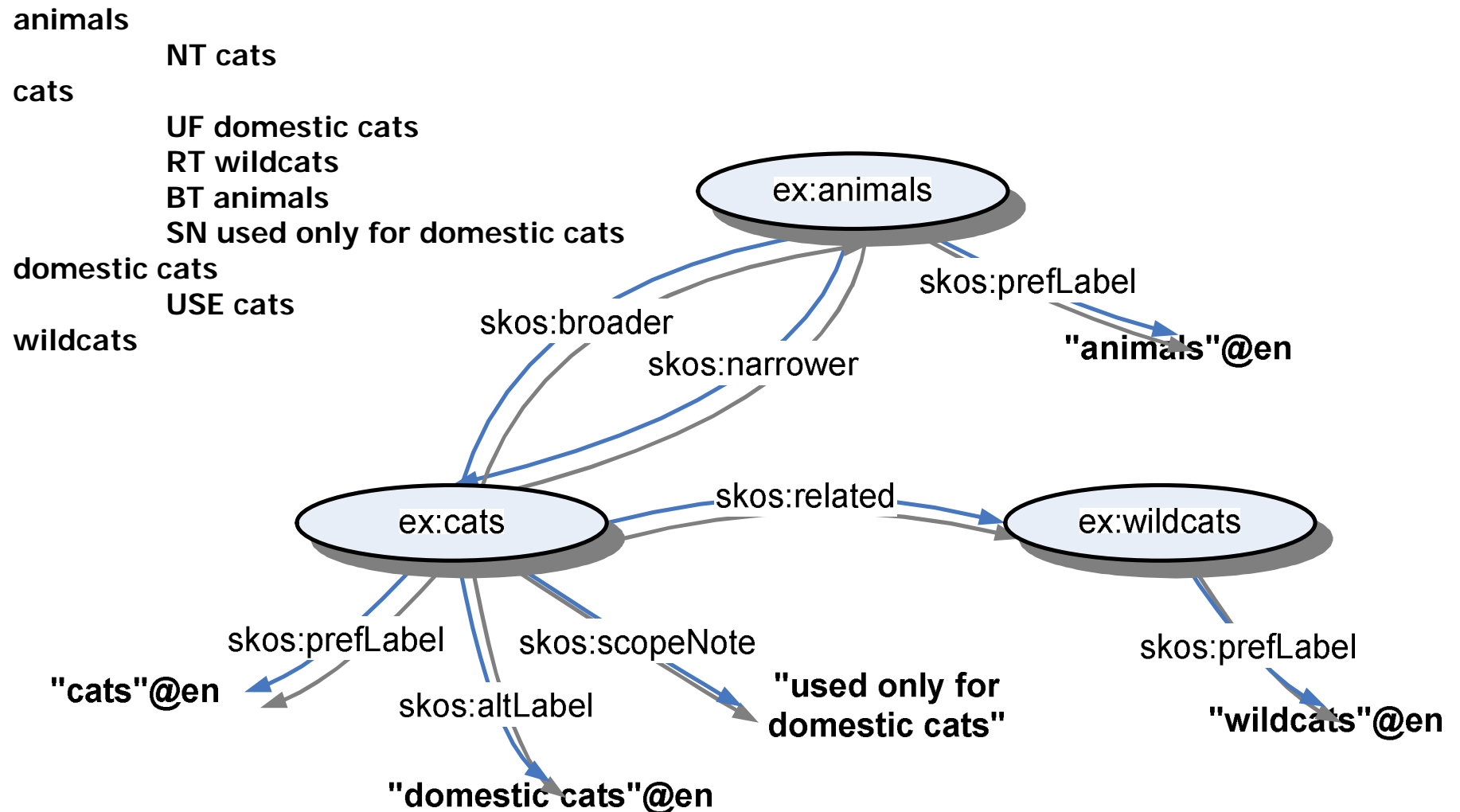
+-- skos:editorialNote

|

+-- skos:changeNote



Example: SKOS graph



Some other features

- Collections of concepts
 - skos:Collection, skos:OrderedCollection, skos:member
- Explicit representation of Concept Schemes
 - skos:ConceptScheme, skos:inScheme, skos:hasTopConcept
- Notations
 - skos:notation
- Mapping relations
 - coming now...

CH case: Libris

- <http://libris.kb.se/>
- Swedish Library as linked data

The screenshot displays the Libris website interface. At the top, the Libris logo is on the left, and navigation links for 'LIBRIS SEARCH FACILITIES', 'HELP', 'PÅ SVENSKA', 'PREFERENCES', and 'CLEAR HISTORY' are on the right. Below this is a green navigation bar with tabs for 'Start', 'Extended search', 'Index A-Ö', 'Boolean', 'Subdatabases', and 'Search history'. A search bar contains the text 'the difference engine' with a green 'Search' button to its right. Below the search bar, the results are shown for 'Search: the difference engine > The difference engi...'. The main content area features a blue header with '1 of 6' and navigation links 'Previous record', 'Next record', and 'To hitlist'. Below this is a tabbed interface with 'Overview' and 'Details' tabs. The 'Overview' tab is active, showing the title 'The difference engine / William Gibson & Bruce Sterling', authors 'Gibson, William, 1948- (author)' and 'Sterling, Bruce, 1954- (author)', ISBN '0-575-04762-3', publisher 'London : Gollancz, 1990', and length 'English 383 s.'. A small orange book icon is next to the word 'Book'. Below this are links for 'Abstract' and 'Subject headings'. A pop-up window titled 'Subject headings' is open, showing 'Steampunk'. To the right of the main content area, there are two green boxes. The first, 'Find similar', lists 'More titles by' (Gibson, William, 194 ...; Sterling, Bruce, 195 ...) and 'More titles about' (Steampunk). The second, 'Search outside LIBRIS', lists 'Extend your search to:' (Google, Google Book Search, Google Scholar, Scirus, LibraryThing, OAIster) and 'Wikipedia about the author:'.

LIBRIS

LIBRIS SEARCH FACILITIES | HELP | PÅ SVENSKA | PREFERENCES | CLEAR HISTORY

Start | Extended search | Index A-Ö | Boolean | Subdatabases | Search history

the difference engine Search

Search: the difference engine > The difference engi...

1 of 6 | Previous record | Next record | To hitlist

Overview | Details

The difference engine / William Gibson & Bruce Sterling

Gibson, William, 1948- (author)
Sterling, Bruce, 1954- (author)

ISBN 0-575-04762-3
London : Gollancz, 1990
English 383 s.

Book

Abstract | Subject headings

Subject headings
Steampunk

Find similar

More titles by
• Gibson, William, 194 ...
• Sterling, Bruce, 195 ...

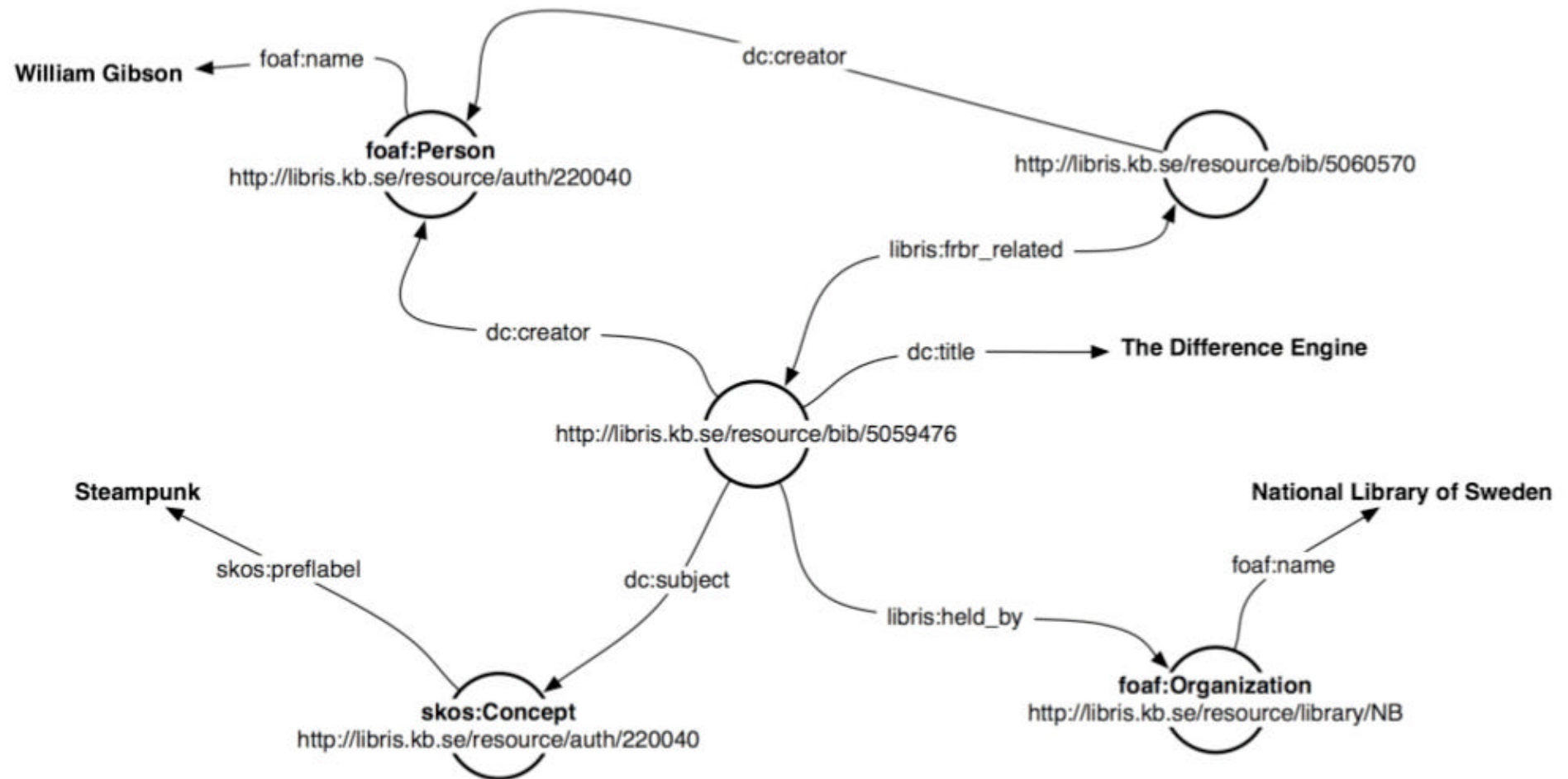
More titles about
• Steampunk

Search outside LIBRIS

Extend your search to:
• Google
• Google Book Search
• Google Scholar
• Scirus
• LibraryThing
• OAIster

Wikipedia about the author:

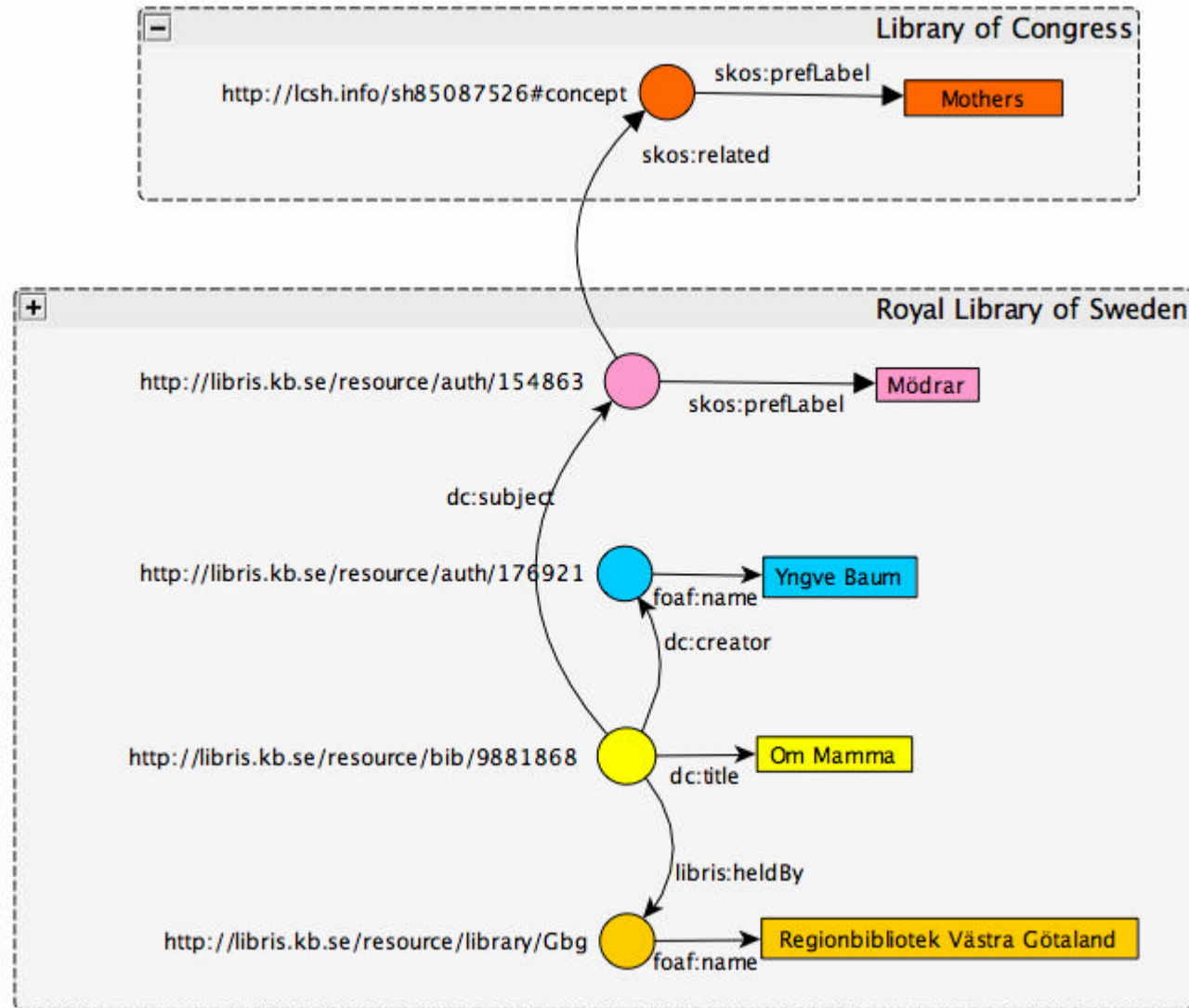
Linked descriptions of resources in Libris



Martin Malmsten, Dublin Core 2008

<http://dc2008.de/wp-content/uploads/2008/09/malmsten.pdf>

External links in Libris: Library of Congress Subject Headings



Ed Summers et. al., Dublin Core 2008

<http://dc2008.de/wp-content/uploads/2008/09/summers-isaac-redding-krech.pdf>

Searching using multiple vocabularies

LIBRIS

Start Extended search Index A-Ö Boolean

Fill in **one** or **multiple** fields

Keywords (title, author, subject, etc.)

Title/title words

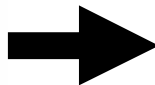
Author/creator

Subject

Classification

ISBN/ISSN etc.

[Clear](#)

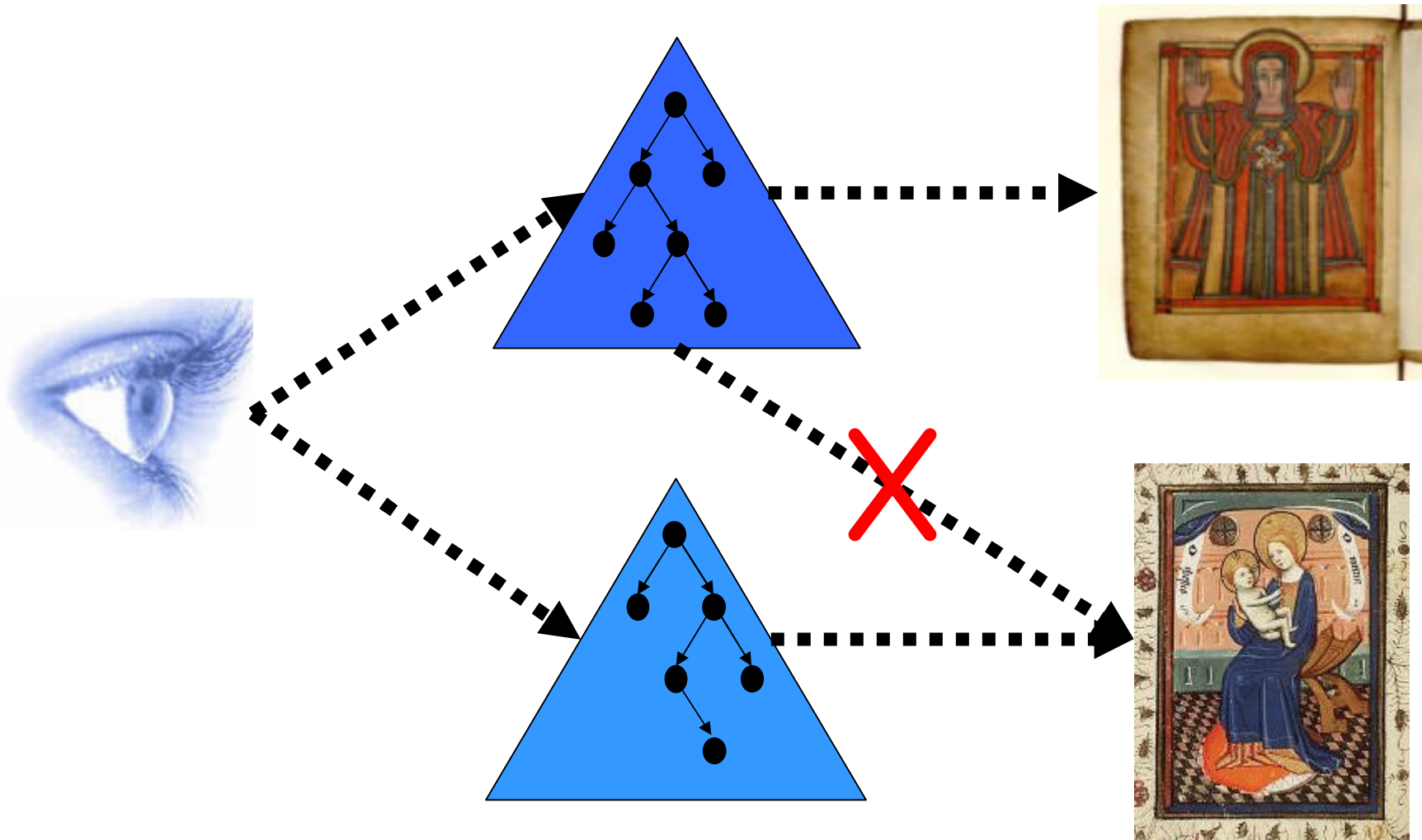


Result 1-10 of 561 [1] 2 3 4 5 6 7 ... 57 ▶

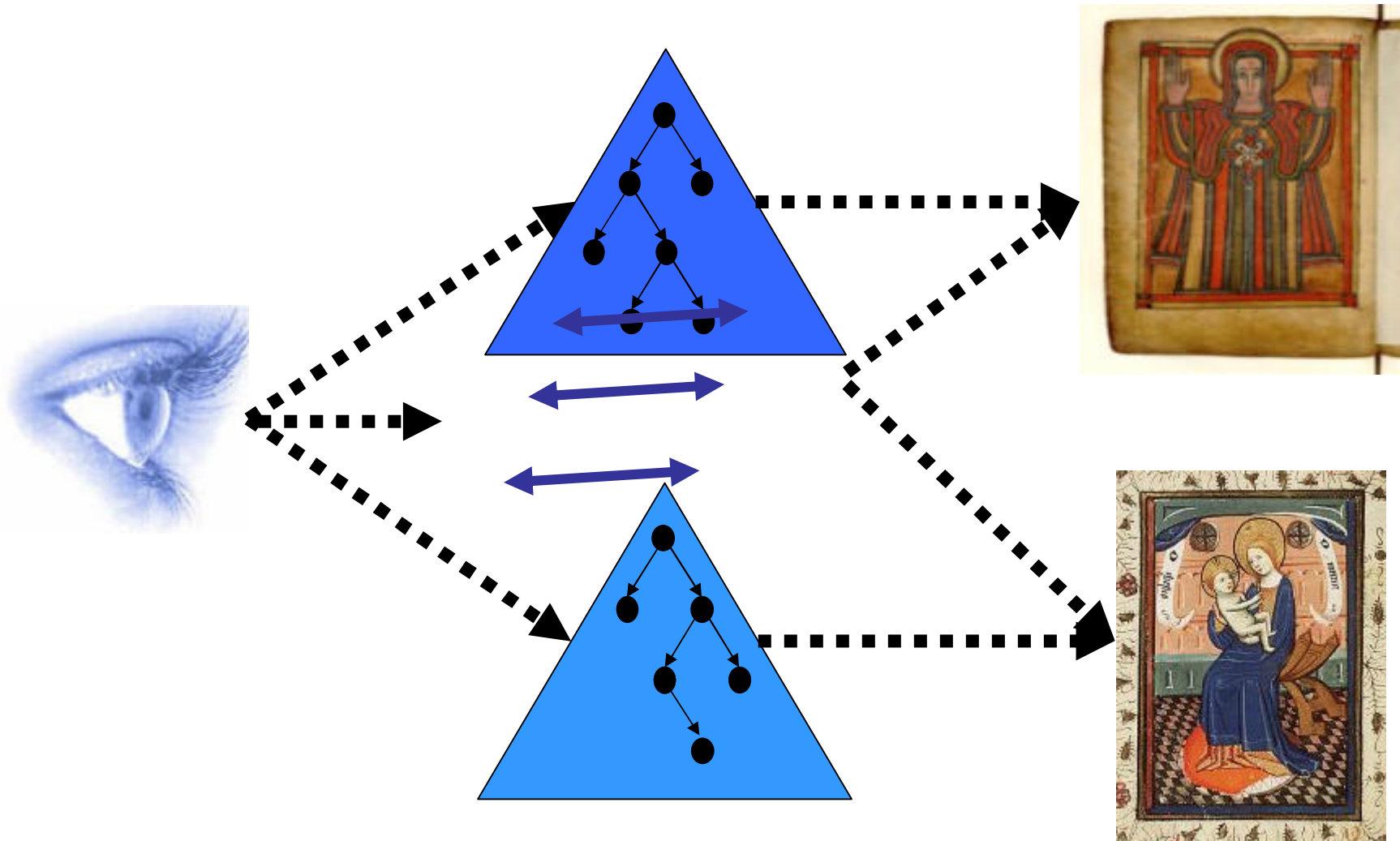
Sort by: relevancy Group by: none Hits per page: 10

- ☐ Björnberg, Ulla, 1943- (author)
[Mödrar på nittiotalets arbetsmarknad : en jämförelse mellan gifta och ensamstående](#)
1997
In: Glastak och glasväggar? / redaktörer: Inga Persson & Eskil Wadensjö. - Stockholm : Fritze, 1997. - (Statens offentliga utredningar, 0375-250X ; 1997:137). - ISBN 91-38-20712-5 ; S. 213-228 : diagr., tab.
 Article/chapter
- ☐ Brancato, Sabrina. (author)
[Mother and motherland in Jamaica Kincaid / Sabrina Brancato.](#)
2005
 Book
- ☐ Garey, Anita Ilta (author)
[Weaving work and motherhood / Anita Ilta Garey](#)
1999
 Book
- ☐ [Mommy wars : stay-at-home and career moms face off on their choices, their lives, their families / edited by Leslie Morgan Steiner.](#)
2006. - 1st ed.
 Book
- ☐ [Social policies, labour markets and motherhood : a comparative analysis of European countries / edited by Daniela del Boca and](#)

Semantic interoperability problem



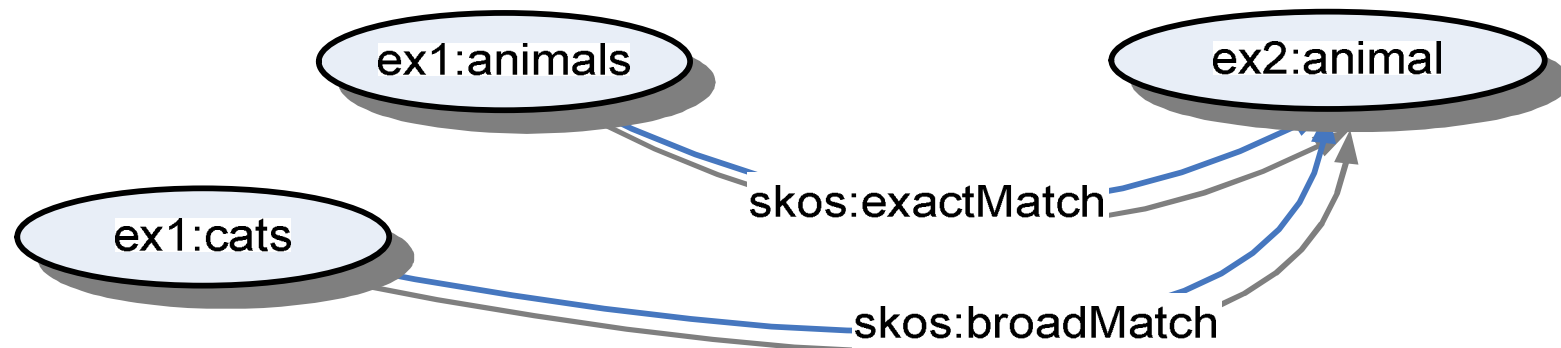
Using semantic alignment



Networking controlled vocabularies in SKOS

animals
cats
wildcats

animal
human
object





Mass adoption!

Vocabularies in the Virtual Observatory Version 1.16

IVOA Proposed Recommendation, 2008 November 4

Editors

Alasdair J G Gray, University of Glasgow, UK
[Norman Gray](#), University of Leicester / University of Glasgow, UK
 Frederic V Hessman, University of Göttingen, Germany
 Andrea Preite Martinez, INAF, Italy

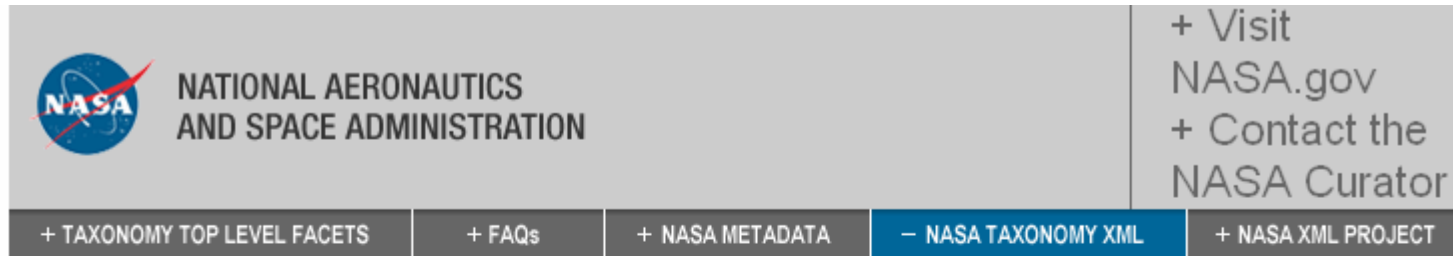
Authors

Sébastien Derrière, Alasdair J G Gray, Norman Gray, Frederic V Hessman, Tony Linde, Andrea Preite Martinez, Rob Seaman and Br

Abstract

As the astronomical information processed within the *Virtual Observatory* becomes more complex, there is an increasing need for a more formal format for vocabularies based on the W3C's *Resource Description Framework* (RDF) and *Simple Knowledge Organization System* (SKOS)

Mass adoption!



NASA Taxonomy - XML DTDs for Use with the NASA Taxonomy

Important Update Regarding the XML format of the NASA Taxonomy - Jan 9, 2007

The next version of the NASA taxonomy will be in the [SKOS format](#).

<http://nasataxonomy.jpl.nasa.gov/>



Taxonomy 2.0

CURRENT VERSION

WHAT IS A TAXONOMY?

A taxonomy is a standards-based classification scheme used to organize electronic content.

HOW ARE TAXONOMIES USED?

Taxonomies can improve many aspects of on line information management.

They can be used to:

- make search more robust
- facilitate data interoperability
- allow business analysis across content in disparate repositories
- support records management for long term archiving
- enable role-based content delivery for portals
- kick off work flows for process based information transactions

HOW IS THE NASA TAXONOMY USED?

The NASA taxonomy provides first steps towards the unification of the NASA information space by documenting a high level set of terms that can be used for mapping together varying data structures. Reconciliation of terms and topics is essential to understanding NASA discoveries in a larger context. [Find out more in our Resources section.](#)

■ [Current version of the NASA Taxonomy](#)

- Taxonomy Facets (or branches)
- Terms, Definitions, Synonyms, Relationships

■ [Maintenance](#)

■ [For Developers](#)

- Current NASA Core Metadata Specification
- DTD Files
- SKOS Files
- Editorial Style Guide

Questions?

- Tricky bits
 - [Demos](#)

About some SKOS modeling choices

- Model constructs
- Formal semantics
- How much **interoperability** does porting to SKOS really allow?
 - Are there different ways to convert similar things?
 - Different interpretations of SKOS constructs?
 - Things impossible to convert?

Preamble: W3C standardization process

- Input: draft specification
- Collect use cases & derive requirements
- Create issues list: requirements that cannot be handled by the draft specification
- Propose resolutions for issues
- Get consensus on spec
- Find two independent implementations for each feature in the spec
- *Continuously*: ask for public feedback/comments

We are dead serious

Guus Schreiber



SKOS Use Cases and Requirements

W3C Working Draft 16 May 2007

This version:

<http://www.w3.org/TR/2007/WD-skos-ucr-20070516/>

Latest version:

<http://www.w3.org/TR/skos-ucr>

Previous version:

This is the first public Working Draft

Editors:

Antoine Isaac, Vrije Universiteit Amsterdam, aisaac@few.vu.nl

Jon Phipps, Cornell University, jhipps@madcreek.com

Daniel Rubin, Stanford Medical Informatics, dlrubin@stanford.edu

Example use case and requirement

- **2.3 Use Case #3 — Semantic search service across mapped multilingual thesauri in the agriculture domain**

“This application coming from the AIMS project [...] includes some more specific links [...] String-to-String relationships ...”

acronym	Food and Agriculture Organization	FAO
spelling_variant	organisation	organization
translation	vache	cow

“Requires: [...] [R-RelationshipsBetweenLabels](#)”

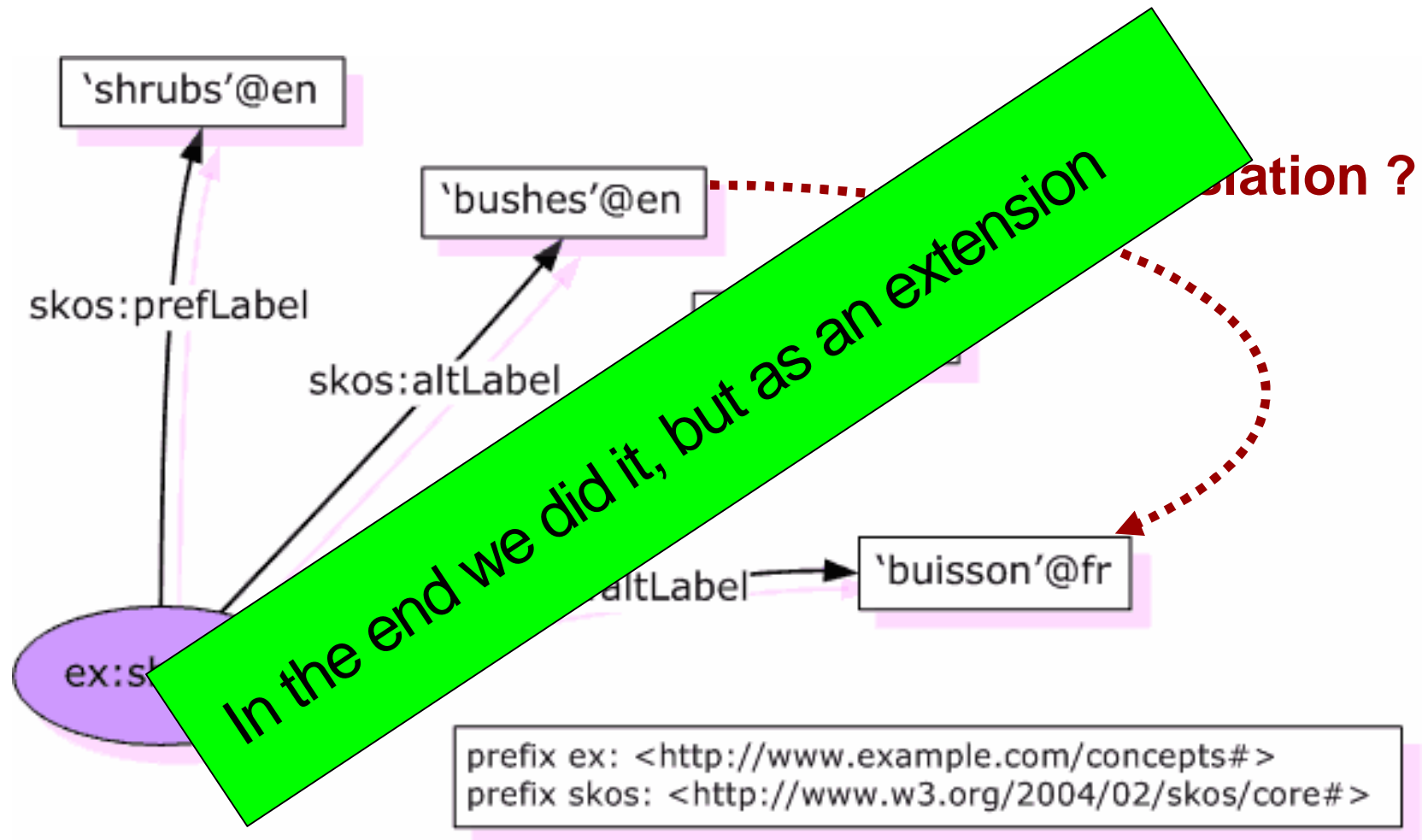
Example issue: relationships between lexical labels

“R-RelationshipsBetweenLabels

**Representation of links between labels associated
to concepts**

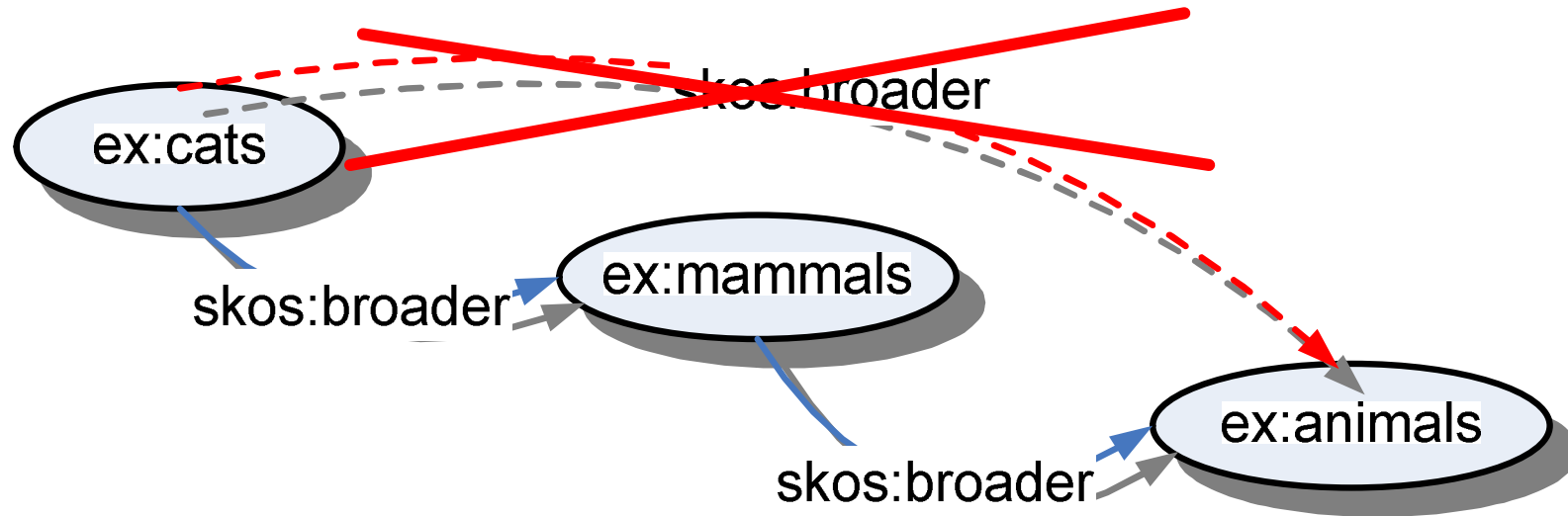
- In previous SKOS spec, labels are literals
- Literals cannot be subject of an RDF property

Example issue: relationships between lexical labels



Semantics of broader

- Is **skos:broader** "transitive"?



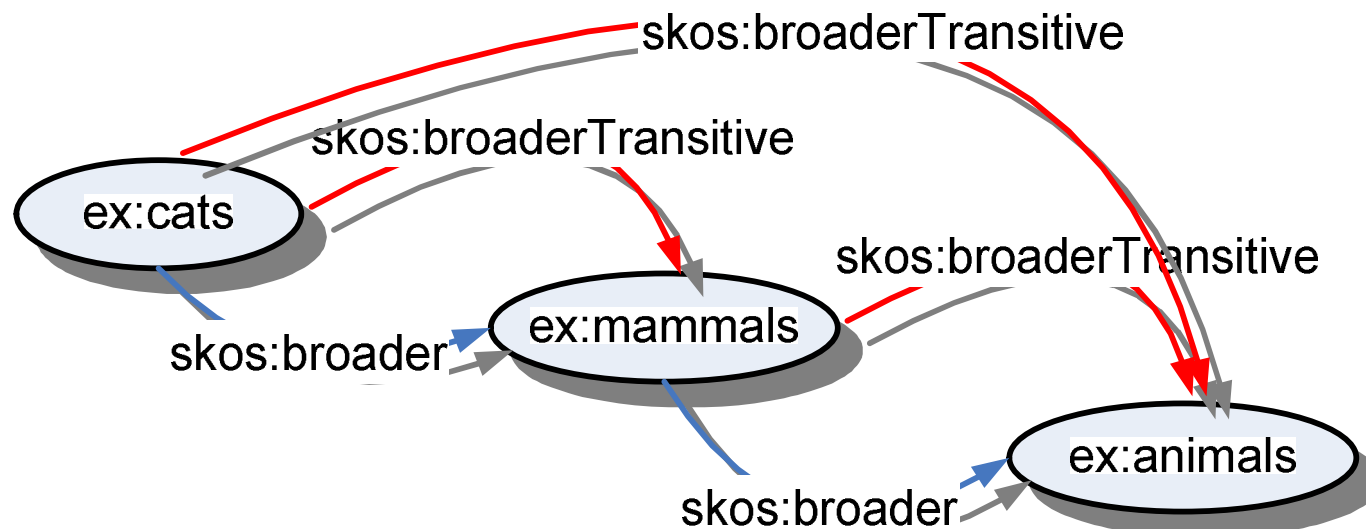
- It can be wrong, sometimes!
- **skos:broader** is not transitive in general

Semantics of broader

skos:broader has a super-property

skos:broaderTransitive with semantics of “has ancestors”

- 1: every **broader** implies a **broaderTransitive**
- 2: **broaderTransitive** is transitive!



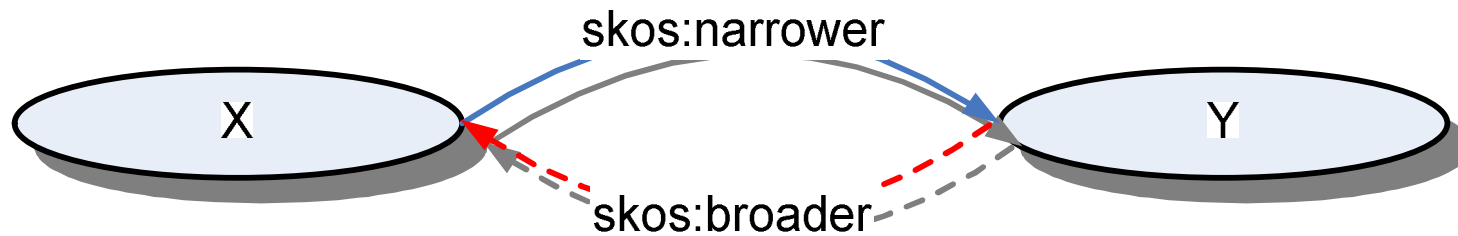
Semantics of broader

- skos:broader is not (a sub-property of) rdfs:subClassOf!
- Problem: people having KOSs with transitive hierarchies may just assert skos:broaderTransitive statements

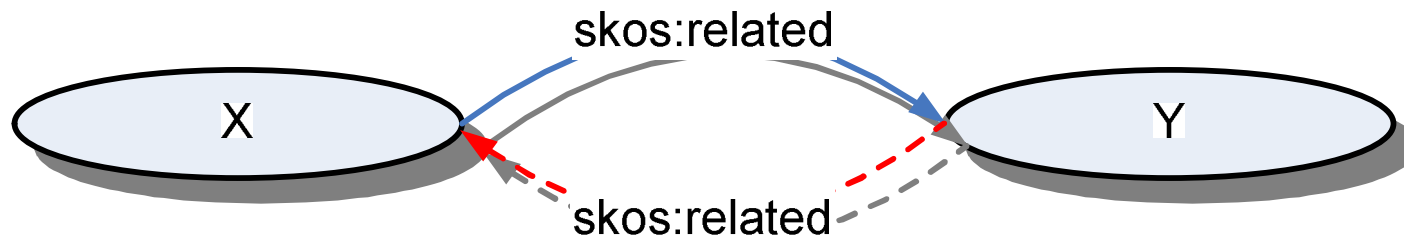
Interop . at risk!

Other semantics of relations

- **broader** and **narrower** are inverse of each other



- **related** is symmetric



Semantics of SKOS

- This tells what should be explicit or not in a SKOS conversion, and what can (shall) be inferred from it
 - Important for building applications using SKOS

Interop . enabler

- Beware: this sometimes requires reasoning!

Interop . at risk!

Is that damn thing useful?

- At least it's there!
 - A proposed standard to represent KOS on the SW
- It allows to publish KOSs
 - Simple, with minimal commitment
 - For most KOS features, conversion is smooth
- It allows to develop applications with re-usable & interoperable components
 - It can also be extended

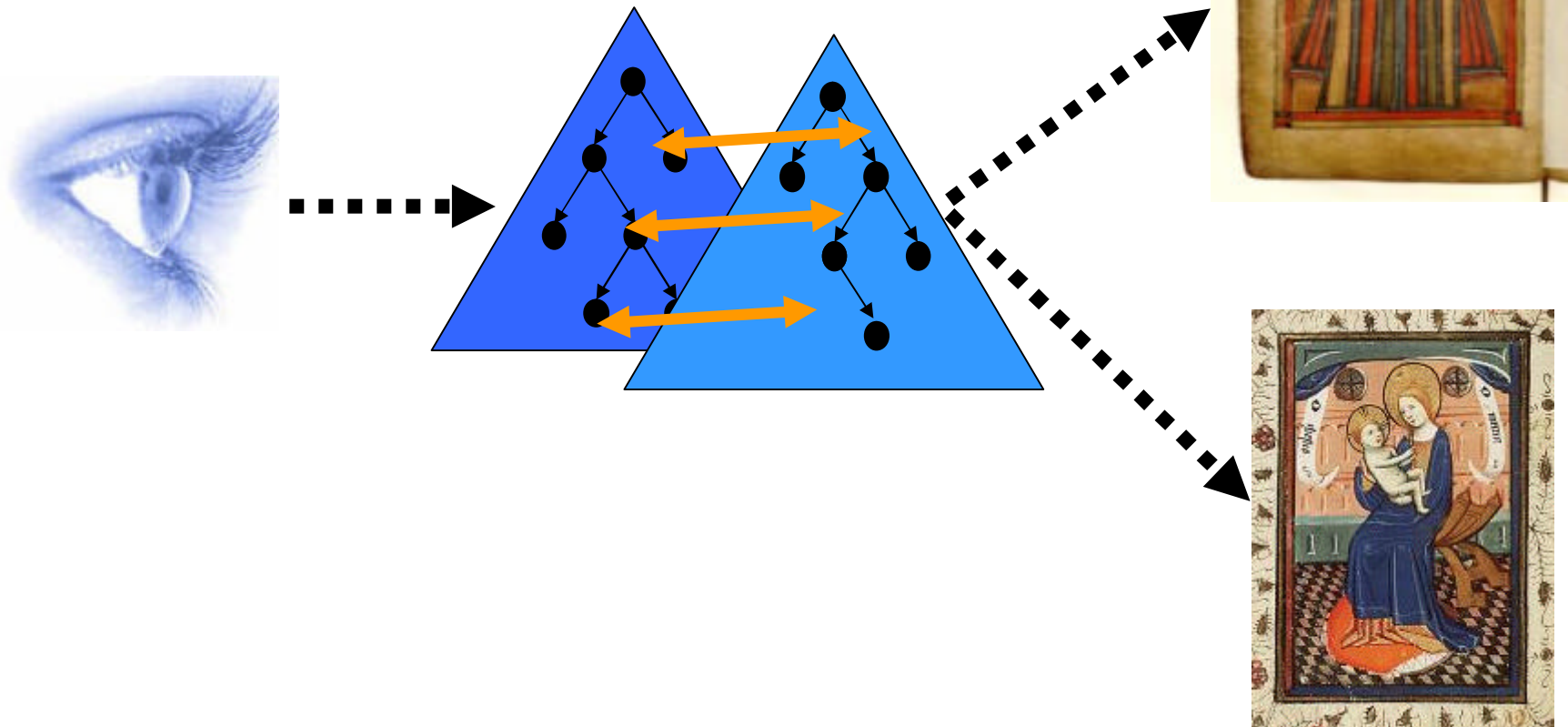
Conclusion

- ✍ Despite some issues, SKOS contributes to enhance interoperability of KOSs

Thanks!

- Reminder: comments highly welcome on SKOS
SKOS Reference: <http://www.w3.org/TR/skos-reference>
SKOS Primer: <http://www.w3.org/TR/skos-primer>
- Some pointers:
SKOS: www.w3.org/2004/02/skos/
Europeana: www.europeana.eu
Libris: libris.kb.se
eCulture (semantic search): eculture.cs.vu.nl
STITCH (vocabulary alignment and repository):
stitch.cs.vu.nl/demo.html

SKOS Demos: searching, browsing and repositories



- <http://eculture.cs.vu.nl:33333/MANDRA-SV-ICE-mandraNewNONE>
- <http://www.europeana.eu/portal/thought-lab.html>