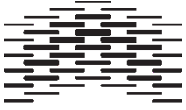




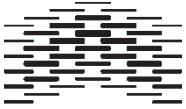
Linked data and Topic Maps

Lecture 3 at Masaryk University



Content

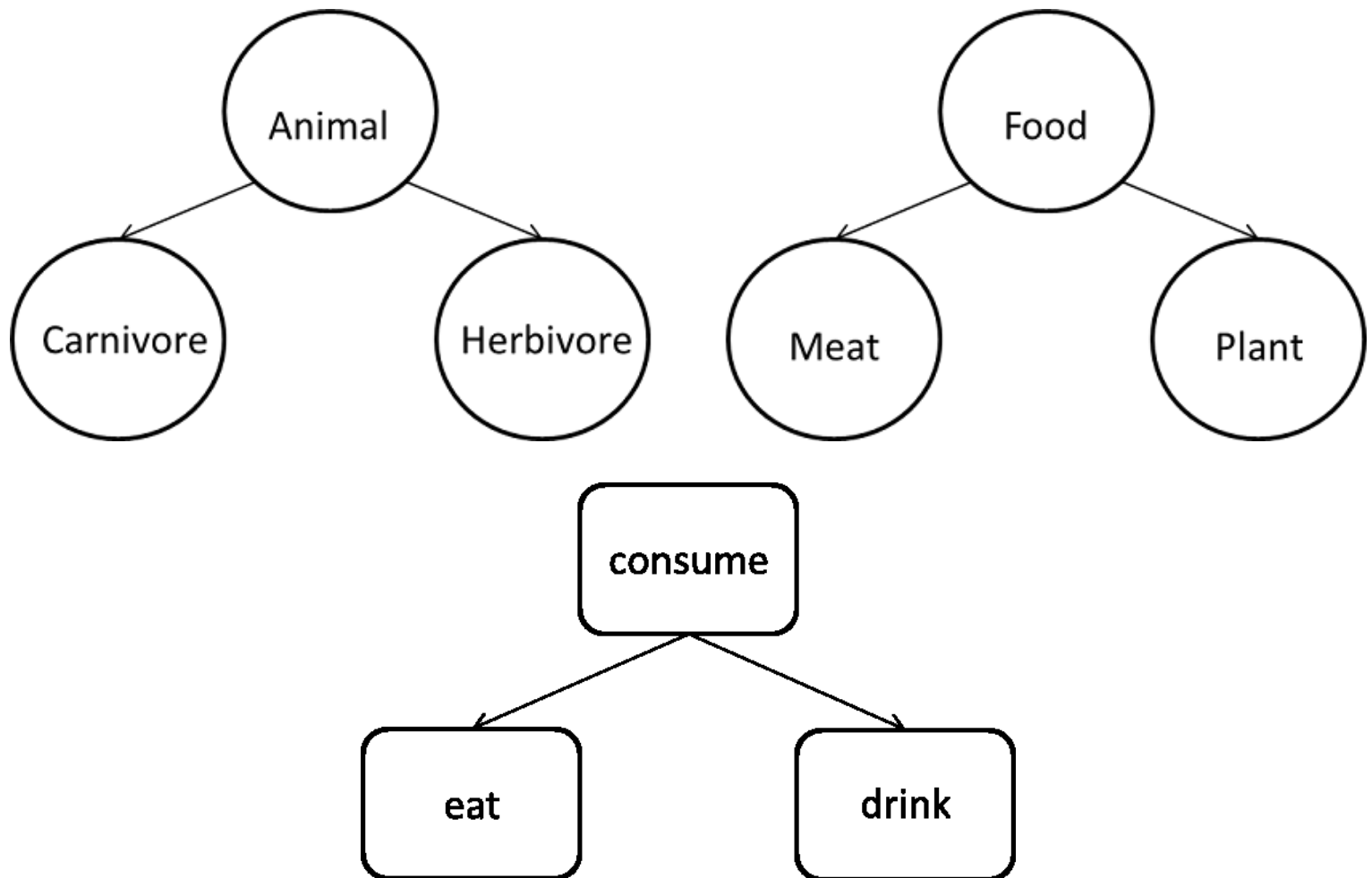
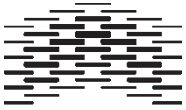
- ontology modelling revisitet
- interoperability recapitulated
- the data silo problem
- linked data
- Topic Maps

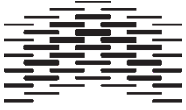


Ontology modelling in RDF, things to remember

- classes represent the collection of objects or individuals
- classes may contain other classes
- instances can belong to multiple classes
- sometimes things are better represented as properties...

Start with creating the class hierarchy!

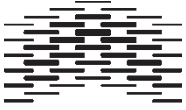




The Semantic Web

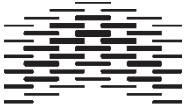
Facilitates

- data sharing
- data merging
- data reuse

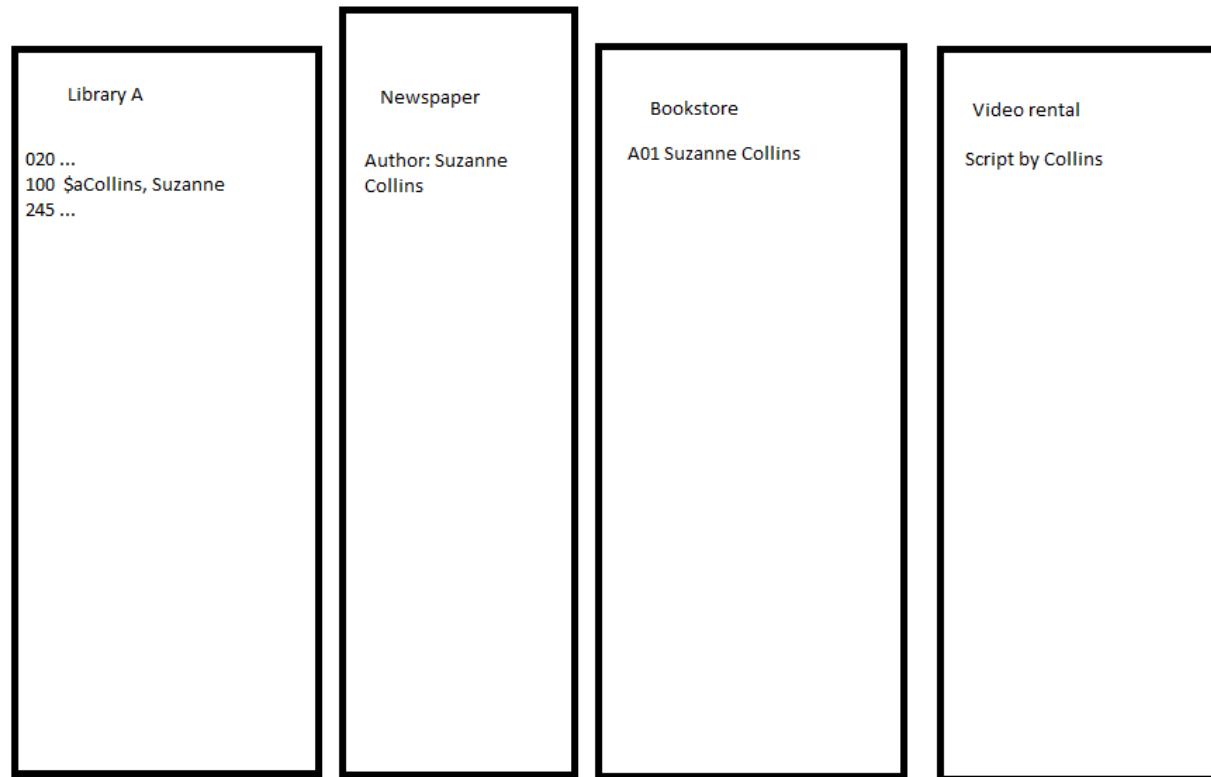


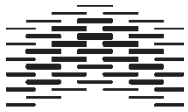
Challenges to the Web

- too much noise
- **internal systems with bad communication capabilities (data silos)**
- large costs of communication



Silo problem

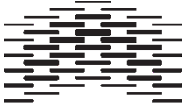




OSLO AND AKERSHUS
UNIVERSITY COLLEGE
OF APPLIED SCIENCES

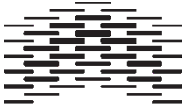
Tim Berners-Lee on the next Web





Linked data - bottom up approach to the semantic web

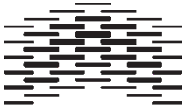
- Solve the silo problem by:
 - making your data available in RDF format (as SPARQL end points)
 - linking them to other existing data



The basics of linked data

1. Use [URIs](#) to identify things.
2. Use [HTTP URIs](#) so that people can look up those names.
3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL)
4. Include links to other URIs. so that they can discover more things

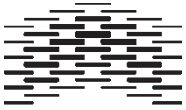
Tim Berners-Lee (2006). [Linked data - design issues](#)



1. Use URIs as names for things

Unique identifiers are essential for finding and referring to "things"

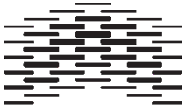
- URL - Uniform Resource Locator, known as a "web address", the same content may exist on several addresses
- URN - Uniform Resource Name, less well known. States the unique "name" of a resource. ISBN is a commonly used example



2. Use http URIs so that people can look up those names

Use the Web's standard protocol for hypertext transfer rather than other naming scheme.

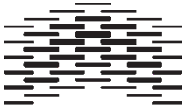
- facilitates access to resource describing the "thing"
- secures decentralized resource management
- enables "dereferencing"



Dereferencing URIs and content negotiation

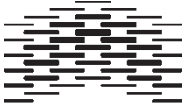
1. lookup URI
 - a. URI identifies the wanted information resource
 1. server sends an http response code 200 as well as the resource
 - b. URI identifies a resource as a non-information resource
 1. servers sends an http response code 303 along with a URI to a description of a resource representing the non-information resource
 2. client asks to get the representation resource in specified format, e.g. XML/RDF-format
 3. server sends client the RDF/XML-document

See more on [How to publish Linked Data on the Web](#)



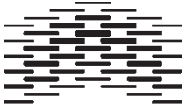
3. When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL)

- An Excel spreadsheet or a scanned PDF is better than nothing, but using standards is better



4. Include links to other URIs so that they can discover more things

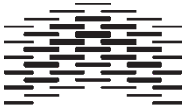
- If there are no links the data will remain in their silos...



Levels of linked open data (one to five stars)

1. Data openly available on the web
2. Data available in a machine-readable format
3. Data available in a non-proprietary format
4. Data available in W3C-format (RDF and SPARQL)
5. Data linked to other people's data





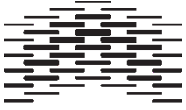
RDF revisited

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:dc="http://purl.org/dc/elements/1.1/">
<rdf:Description
rdf:about="http://www.jbi.hio.no/bibin/dig_korg/sem_web.htm">
<dc:Creator>Nils Pharo</dc:Creator>
<dc:Subject>Semantic Web</dc:Subject>
</rdf:Description> </rdf:RDF>
```

equals:

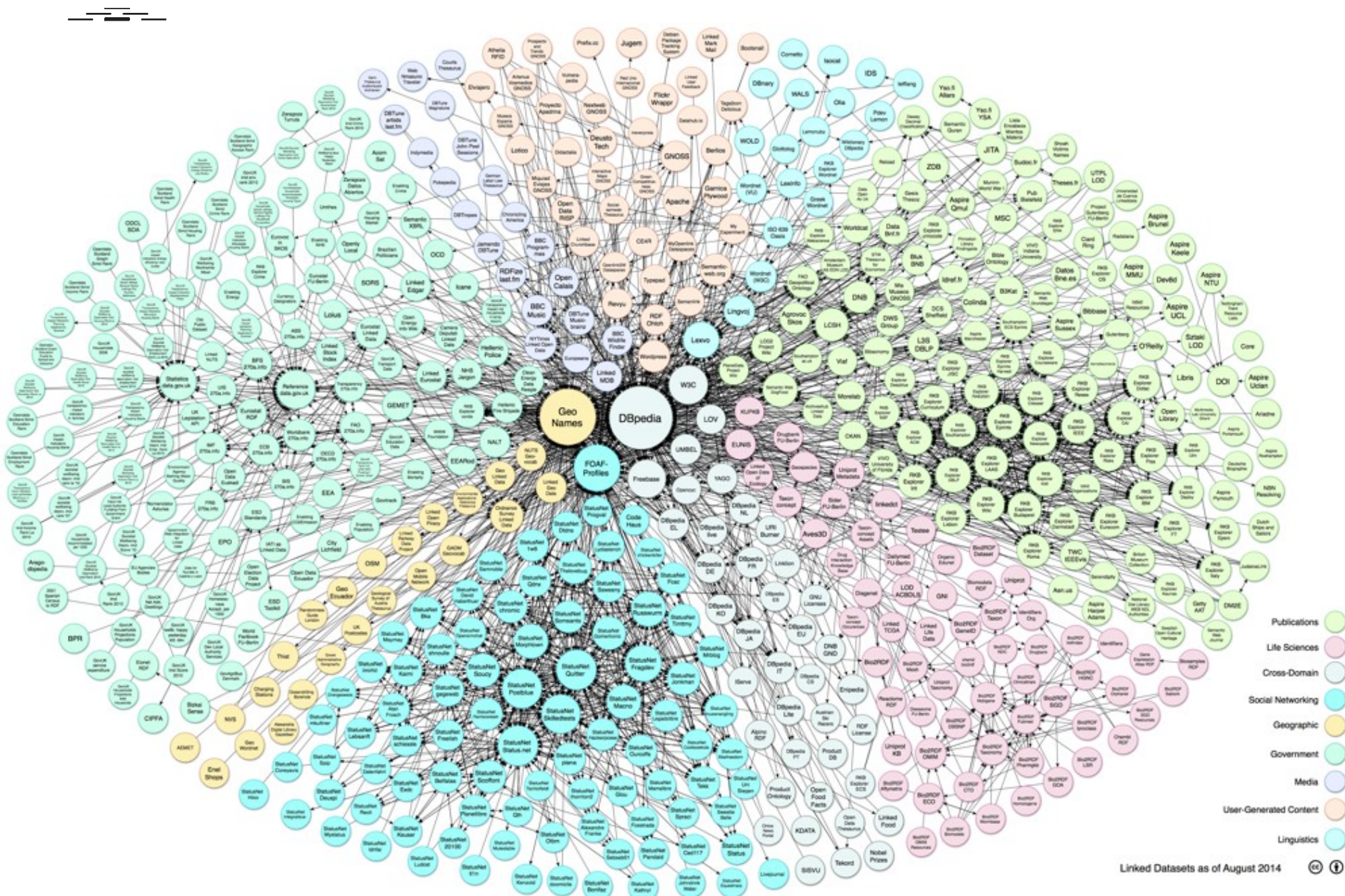
```
<@prefix dc: <http://purl.org/dc/elements/1.1/>.
<@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
<http://www.jbi.hio.no/bibin/dig_korg/sem_web.htm> dc:Creator "Nils
Pharo" ;
dc:Subject "Semantic Web" .
```

This is valid RDF, but it does not link with other data sets

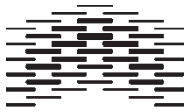


OSLO AND AKERSHUS
UNIVERSITY COLLEGE
OF APPLIED SCIENCES

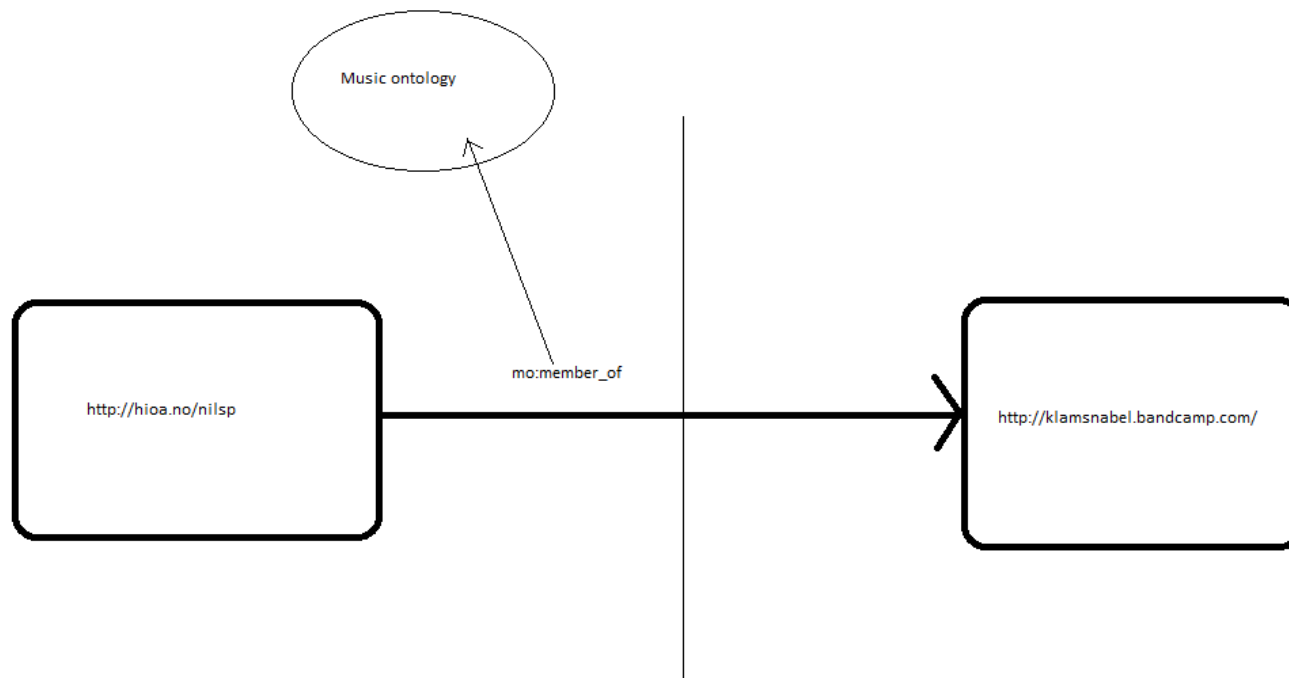
Linked data means connecting data sets!

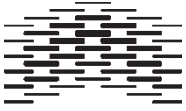


Linked Datasets as of August 2014



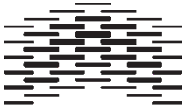
Example





Three types of RDF links

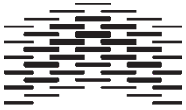
- relationship links
- identity links
- vocabulary links



Relationship links

purpose is to reuse other information sources to enrich your data

```
<http://hioa.no> <http://xmlns.com/foaf/spec/based_near>  
<http://dbpedia.org/page/Oslo>
```

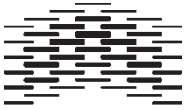


Identity links

Coupling resources that speak about the same things, establish URI aliases

```
<http://dbpedia.org/page/Oslo>  
<http://www.w3.org/2002/07/owl#sameAs>  
<http://www.geonames.org/3143244/>
```

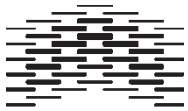
- explicates different opinions
- facilitates traceability
- secures robustness



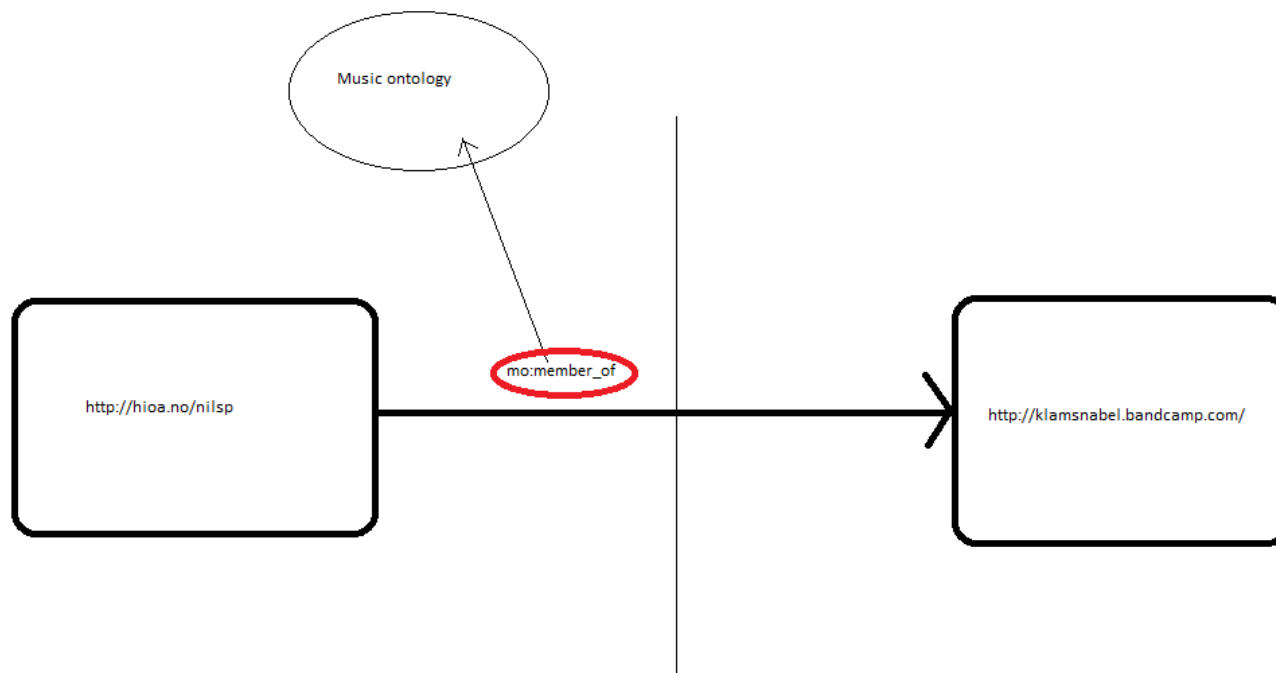
Vocabulary links

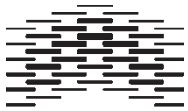
- Basic principle: reuse existing vocabularies instead of creating your own!
- Describe new vocabularies in RDFS/OWL
- create mappings between terms from different vocabularies

owl:equivalentClass can be used to map between classes



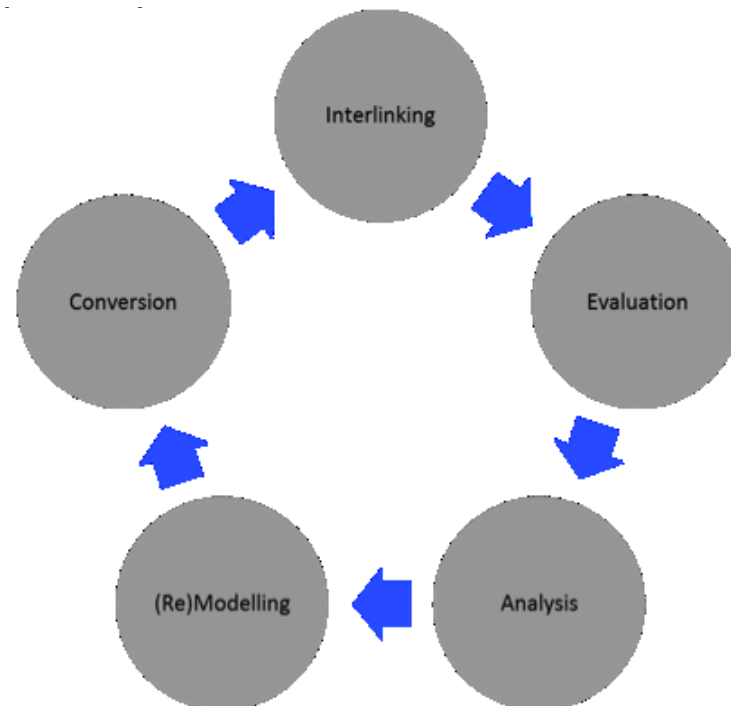
Vocabulary links example

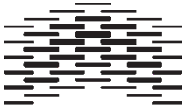




Linked data in practice, challenges

The linked data re





Conversion problems, example

Record A

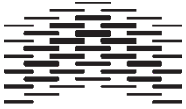
=110\$aMayhem
=24510\$aMediolanum capta est
=700 0\$aManiac
=700 0\$aBlasphemer
=700 0\$aHellhammer
=700 0\$aNecrobuttercher
=710 0\$aMayhem
=7400 \$aCarnage
=7400 \$aNecrolust
=7400 \$aDeathcrush
=7400 \$aAncient skin
=7400 \$aFreezing moon
=7400 \$aFall of seraphs
=7400 \$aSilvester Anfang
=7400 \$aChainsaw gutsfuck
=7400 \$aFrom the dark past
=7400 \$aI am thy labyrinth
=7400 \$aSymbols of bloodswords
=7400 \$aPure fucking Armageddon
=900 0\$aEriksen, Rune\$aBlasphemer
=900 0\$aStubberud, Jørn\$aNecrobuttercher
=900 0\$aKristiansen, Sven-Erik\$aManiac
=900 0\$aBlomberg, Jan Axel\$aHellhammer

Record B

=110\$aMayhem
=24510\$aGrand declaration of war
=700 1\$aManiac
=700 1\$aBlasphemer
=700 1\$aHellhammer
=700 1\$aNecrobuttercher
=700 1\$aFinstad, Børge
=700 0\$aManiac\$aTo Daimonion
=700 0\$aManiac\$aA time to die
=700 0\$aManiac\$aView from Nihil
=700 0\$aBlasphemer\$aTo Daimonion
=700 0\$aBlasphemer\$aA time to die
=700 0\$aBlasphemer\$aView from Nihil
=700 0\$aManiac\$aA grand declaration of war
=700 0\$aManiac\$aA bloodsword and a colder sun
=700 0\$aBlasphemer\$aA grand declaration of war
=700 0\$aManiac\$aIn the lies where upon you lay
=700 0\$aManiac\$aCompletion in science of agony
=700 0\$aBlasphemer\$aA bloodsword and a colder sun
=700 0\$aBlasphemer\$aIn the lies where upon you lay
=700 0\$aManiac\$aCrystalized pain in deconstruction
=700 0\$aBlasphemer\$aCompletion in science of agony
=700 0\$aBlasphemer\$aCrystalized pain indeconstruction
=710 0\$aMayhem
=900 1\$aNecro\$aNecrobuttercher

Record C

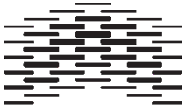
=110\$aMayhem
=24510\$aDeathcrush
=700 1\$aManiac
=700 1\$aMessiah
=700 1\$aAvnskog, Erik
=700 1\$aButcher, Necro
=700 1\$aAarseth, Øystein
=700 0\$aSchnitzler, Conrad\$aSilvester Anfang
=710 0\$aMayhem
=710 0\$aMayhem\$aNecrolust
=710 0\$aMayhem\$aDeathcrush
=710 0\$aVenom\$aWitching hour
=710 0\$aMayhem\$aWeird Manheim
=710 0\$aMayhem\$aChainsaw gutsfuck
=710 0\$aMayhem\$aPure fucking armageddon
=900 1\$aEuronymous\$aAarseth, Øystein



Philosophy of open data

certain data should be freely available to anyone

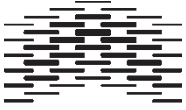
- non-textual data such as:
 - maps, formulaes, genomes
- textual data
 - government data, facts, public library records?



Open data arguments

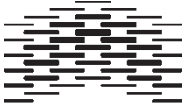
- data belong to the human race
- data was funded by public money
- data was created by government
- facts cannot legally be copyrighted
- openness accelerates progress

See also [Wikipedia](#)



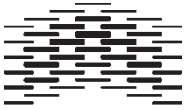
The Linking open data project

- goal: to convert data that are available under open licences to RDF
- [The state of the LoD cloud 2014](#)



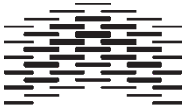
Exploring linked open data

- [Swoogle](#)
- [Semantic Web search engines](#)



Library data as linked open data

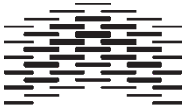
- descriptive bibliographic metadata (title, edition and document languages)
- authority control (author and title languages)
- content indexing (subject languages)



Using RDF to represent subject languages - SKOS

"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."

- Simple Knowledge Organisation System
- vocabulary for expressing controlled vocabularies in RDF
- can be used for modelling ontologies up to thesaurus "level"
- W3C Recommendation, 18. August 2009
- [SKOS home page](#)

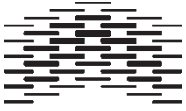


SKOS concept example

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:skos="http://www.w3.org/2004/02/skos/core#">

  <skos:Concept
    rdf:about="http://www.example.com/concepts#abattoirs">
    <skos:prefLabel>abattoirs</skos:prefLabel>
    <skos:altLabel>slaughterhouse</skos:altLabel>
    <skos:hiddenLabel>abatoirs</skos:hiddenLabel>
    <skos:hiddenLabel>abbatoirs</skos:hiddenLabel>
    <skos:hiddenLabel>abbattoirs</skos:hiddenLabel>
  </skos:Concept>

</rdf:RDF>
```



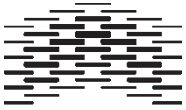
SKOS hierarchical relationship example

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:skos="http://www.w3.org/2004/02/skos/core#">

  <skos:Concept rdf:about="http://www.example.com/concepts#mammals">
    <skos:prefLabel>mammals</skos:prefLabel>
    <skos:broader rdf:resource="http://www.example.com/concepts#animals"/>
  </skos:Concept>

  <skos:Concept rdf:about="http://www.example.com/concepts#animals">
    <skos:prefLabel>animals</skos:prefLabel>
    <skos:narrower rdf:resource="http://www.example.com/concepts#mammals"/>
  </skos:Concept>

</rdf:RDF>
```



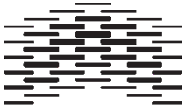
SKOS associative relationships example

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:skos="http://www.w3.org/2004/02/skos/core#">

  <skos:Concept rdf:about="http://www.example.com/concepts#birds">
    <skos:prefLabel>birds</skos:prefLabel>
    <skos:related rdf:resource="http://www.example.com/concepts#ornithology"/>
  </skos:Concept>

  <skos:Concept rdf:about="http://www.example.com/concepts#ornithology">
    <skos:prefLabel>ornithology</skos:prefLabel>
    <skos:related rdf:resource="http://www.example.com/concepts#birds"/>
  </skos:Concept>

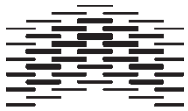
</rdf:RDF>
```



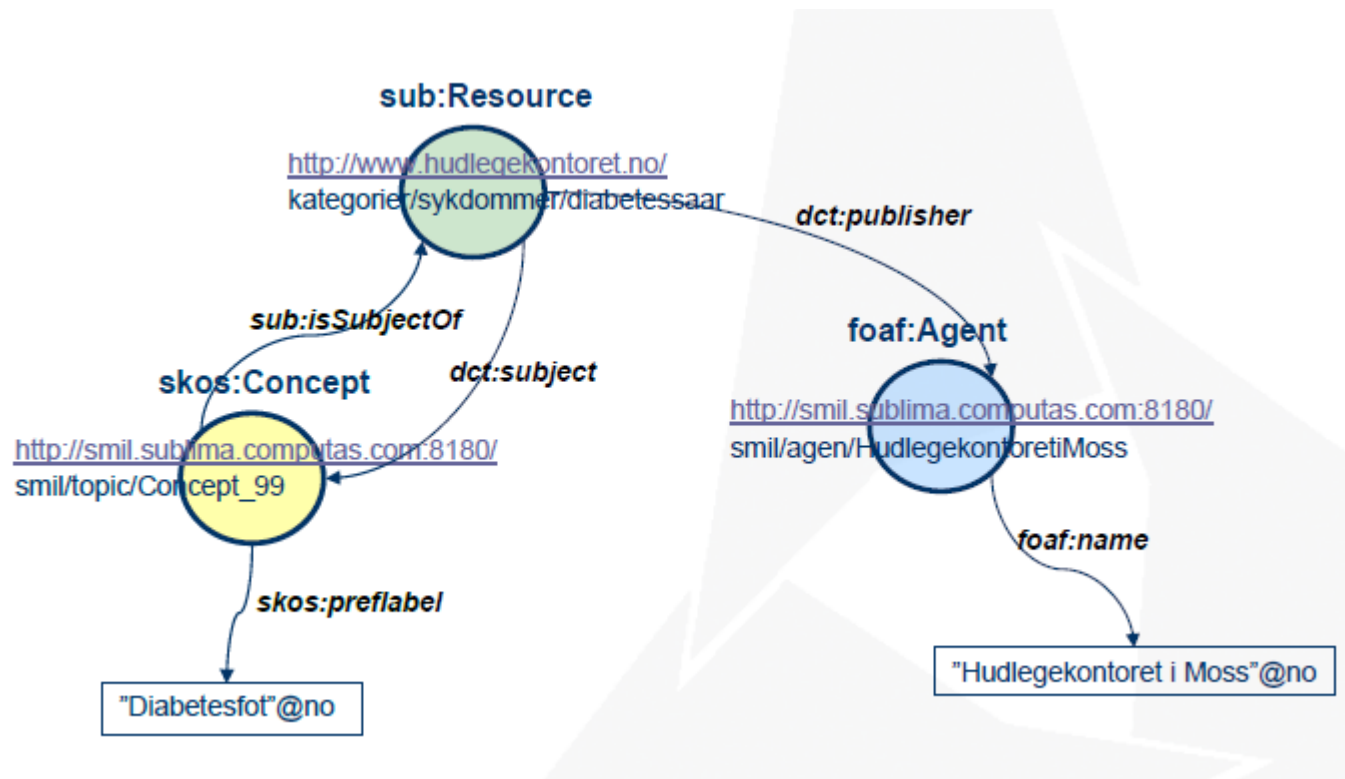
SKOS is concept-oriented

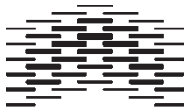
In contrast to termed-based thesauri, SKOS focus on concepts which may have several labels. In a standard thesaurus a term would relate to another term using relationships. See also the SKOS [FAQ](#)

- Examples
- [The Integrated Public Service Vocabulary](#)
- Library of Congress have made their [subject headings](#) available

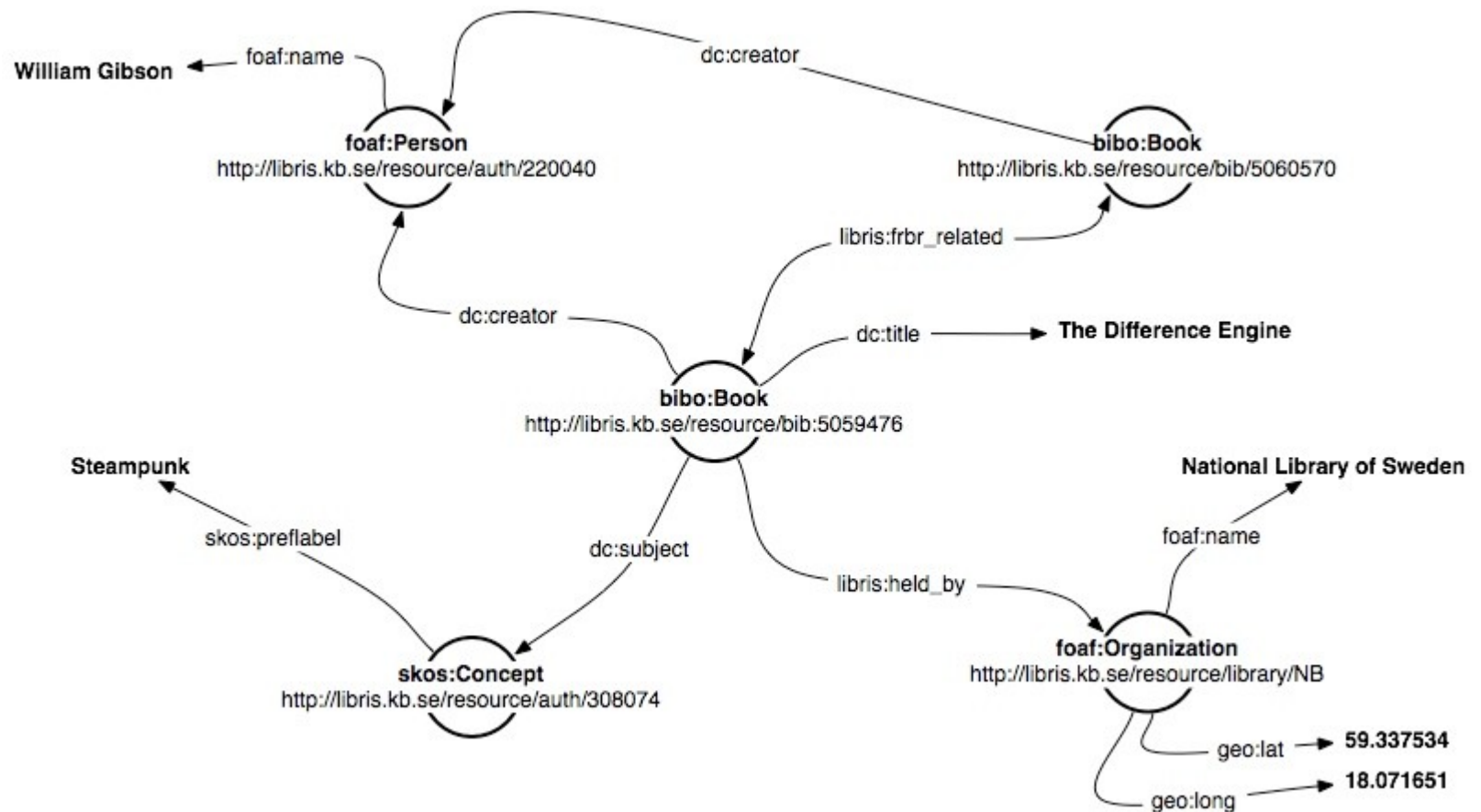


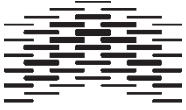
SKOS used in a subject portal





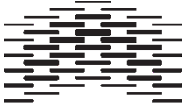
SKOS used in Libris





Characteristics of examples

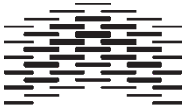
- a rich variety of name spaces
- [foaf](#), sub, dct, dc, libris, [bibo](#)



Using RDF to represent document and work languages

Many attempts have been made...

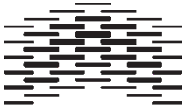
- the bibliographic ontology
- BIBFRAME
- Schema.org with extensions



Bibliographic description - from records to graphs?

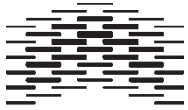
MARC has been used for describing bibliographic records for a long time

- international standard since 1973
- billions of MARC records exist
- developed in the pre-web world

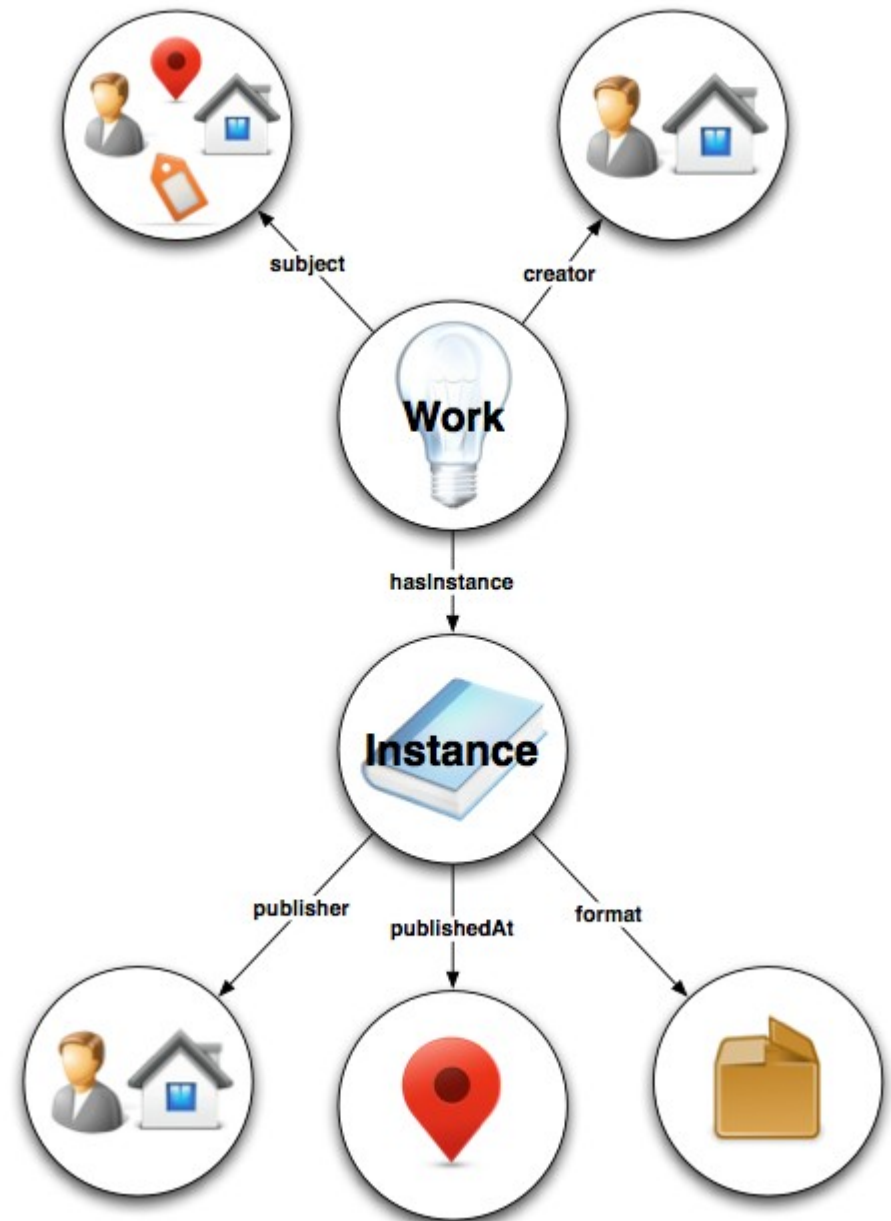


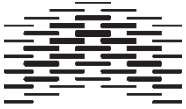
BIBO - the bibliographic ontology

- [BIBO specification](#)
- primarily developed for handling citations and bibliographic references
- can be used for simple bibliographic description



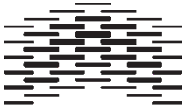
BIBFRAME Model





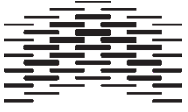
BIBFRAME

- bibframe.org
- [Library of Congress report](#)
- planned to be successor of MARC
- reflects the FRBR model
- develops a new namespace
- [Transformation tools](#)



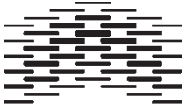
Schema.org's BIB extensions model

- [Schema.org](https://schema.org/) - initiative for creating structured data embedded in web pages
- supported by Google, Bing, Yahoo and Yandex
- light weight, cross-sectoral schema which is extensible for new domains, thus
- compatible with RDF
- [Report from OCLC](#)



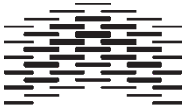
Discussion

What are the pros and cons of the BIBFRAME approach?



Topic maps

- a standard for organising digital content
- ISO certified in 2002 (ISO-standard 13250)
- Used for structuring web sites and a large variety of knowledge management purposes



Background

inspired by back-of-book indexes

X

XML. See Extensible

Markup Language (XML)

XML Topic Maps (XTM) 61,
72, 78

XML web services. See services

XPointer 99

XTM. See XML Topic Maps
(XTM)

XUL widgets

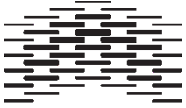
widgets 37

See also Extensible User

interface Language

(XUL)

From Passin (2004). The explorer's guide to the Semantic Web



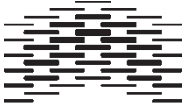
Two-layer model

(index)

Metadata layer

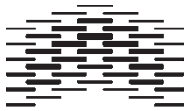
Information layer

(content)



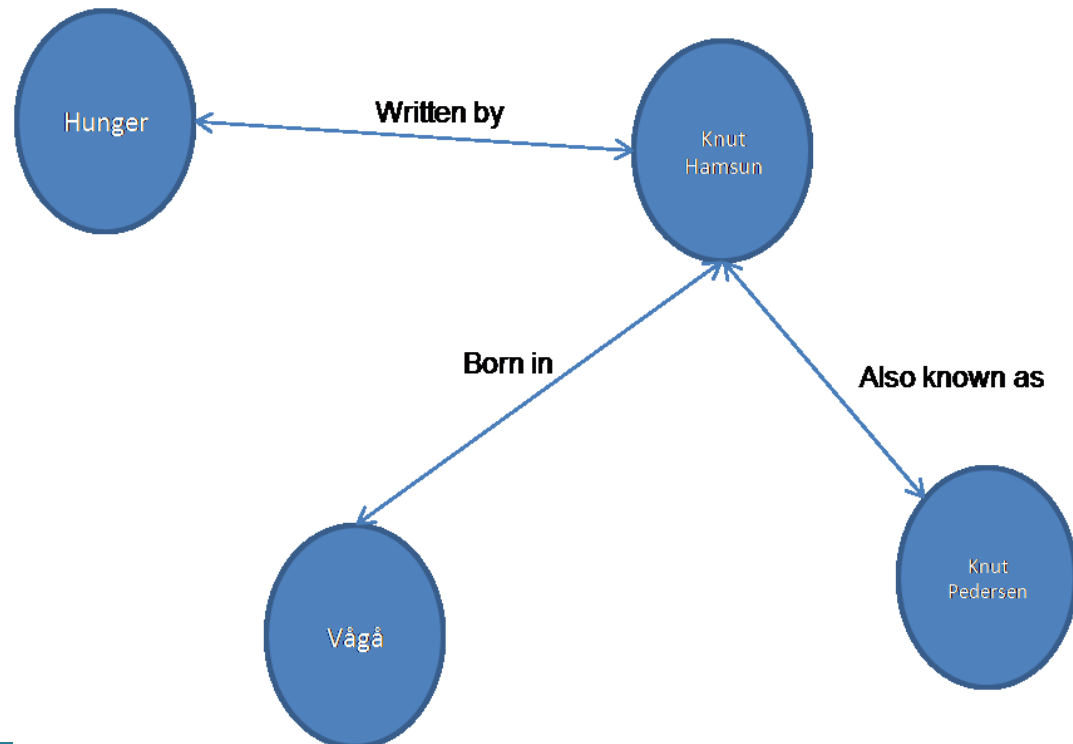
The information layer

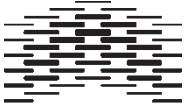
- Contains information (sic) - *occurences*
- independent of location, format or form
- not necessarily digital



The metadata layer

- contains *topics* and *associations* which ties topics together



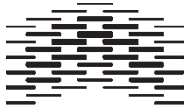


The basic elements of topic maps

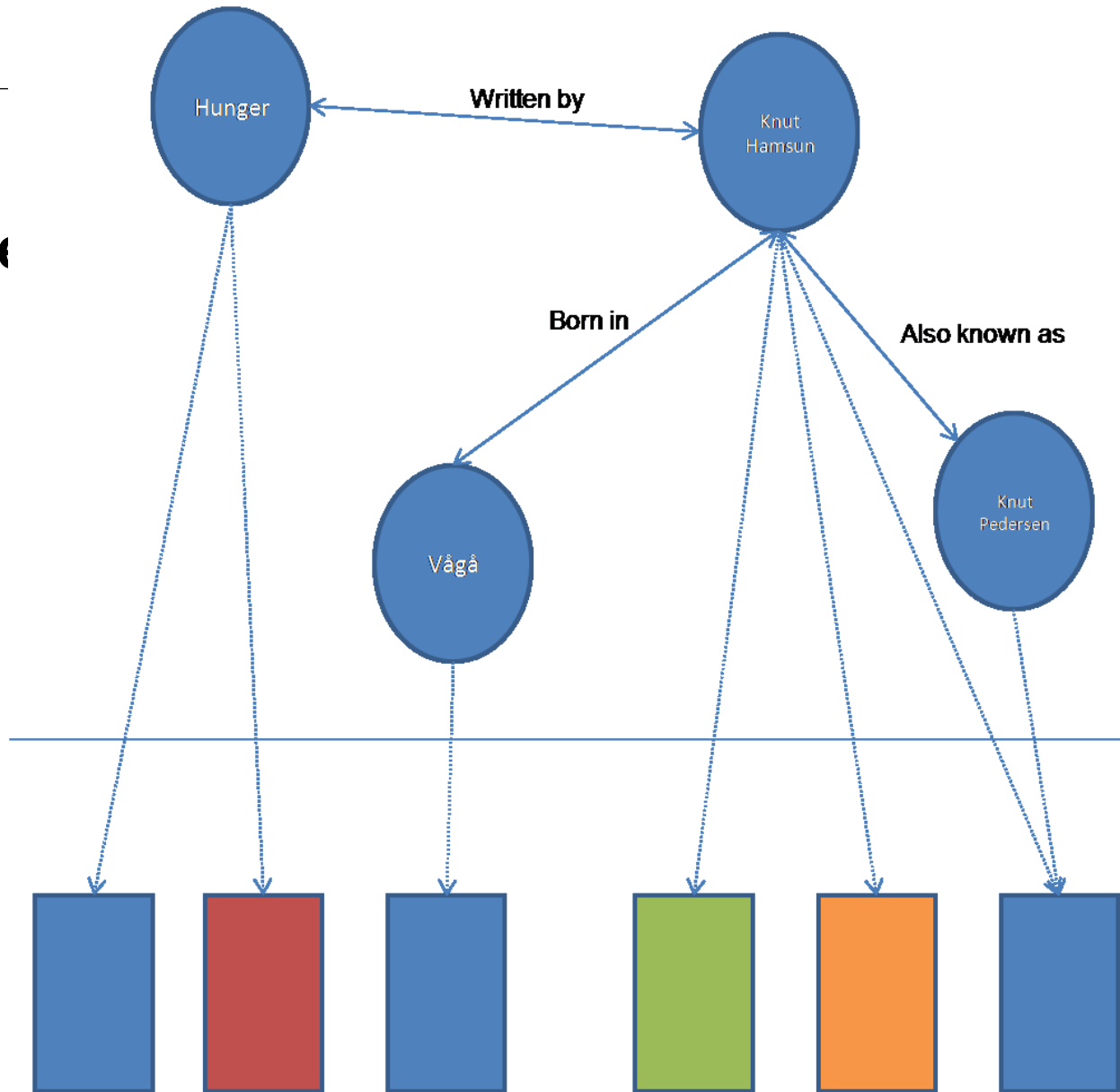
Topic maps consist of

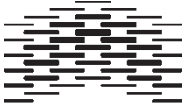
- Topics
- Associations
- Occurrences

i.e. [the TAO of topic maps](#) (Pepper, 2002)



Occurrence

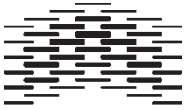




Topic maps and ontologies

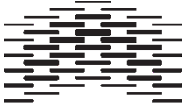
an ontology contains the topic map's input

- minimum requirement: some topics with associations



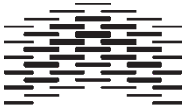
Topic types

- Topics can be typed; Hamsun is an *author*, Vågå is a *place*, Hunger is a *title*. This is comparable to classes in RDF
- Topic types are also topics



Association types

- Association types: Hamsun was *born in* Vågå, Hunger was *written by* Hamsun
- Cf properties in RDF
- Association types are also topics

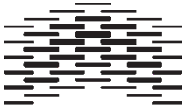


Occurrence types

– Occurrence types:

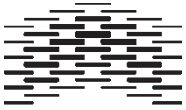
<http://nobelprize.org/nobel_prizes/literature/laureates/1920/hamsun-bio.html>
is a *biography* of Hamsun. <http://en.wikipedia.org/wiki/Knut_Hamsun> is an
encyclopedia article

– Occurrence types are also topics



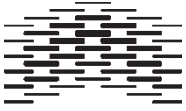
Topic map syntaxes

- HyTM - SGML based
- XTM - XML Topic Maps
- CTM - Compact syntax for Topic Maps



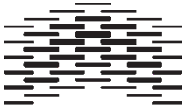
XTM example 1

```
<topic id="nils">
  <instanceOf>
    <topicRef xlink:href="#employee"/>
  </instanceOf>
  <instanceOf>
    <topicRef xlink:href="#teacher"/>
  </instanceOf>
  <baseName>
    <baseNameString>Nils Pharo</baseNameString>
  </baseName>
  <occurrence>
    <instanceOf>
      <topicRef xlink:href="#description"/>
    </instanceOf>
    <resourceData>Nils has worked at Oslo UC since
1997</resourceData>
  </occurrence>
</topic>
```



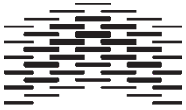
XTM example 2

```
<topic id="ouc">
  <instanceOf>
    <topicRef xlink:href="#institution"/>
  </instanceOf>
  <subjectIdentity>
    <subjectIndicatorRef
xlink:href="http://home.hio.no/~nilsp/psi/hio.psi"/>
  </subjectIdentity>
  <baseName>
    <baseNameString>Oslo University College</baseNameString>
  </baseName>
  <occurrence>
    <instanceOf>
      <topicRef xlink:href="#website"/>
    </instanceOf>
    <resourceRef xlink:href="http://www.hio.no"/>
  </occurrence>
</topic>
```



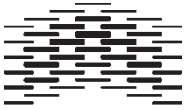
XTM example 3

```
<association id=" nils-ouc-association">
  <instanceOf>
    <topicRef xlink:href="#employment"/>
  </instanceOf>
  <member>
    <roleSpec><topicRef xlink:href="#employee"/></roleSpec>
    <topicRef xlink:href="#nils"/>
  </member>
  <member>
    <roleSpec><topicRef xlink:href="#employer"/></roleSpec>
    <topicRef xlink:href="#ouc"/>
  </member>
</association>
```



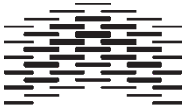
Characteristics of topic maps

- topics have names
- topics are knit together using associations
- a topic may be categorised by an unlimited number of topic *types*
- topic types are topics on a higher level of abstraction
- association types are association on a higher level of abstraction
- occurrences may be external or internal to the topic map
- topics can be disambiguated using subject indicators



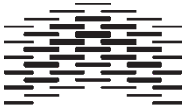
Topic and topic types

- a topic type defines a class/category of things
- topic types *need* instances!
- domain dependent
- Choose an appropriate level of generality
 - "Countries" is better than "Countries in South-East Asia"
 - The domain of the topic map tells you which countries it includes
- Don't make it too general!



Type hierarchies

- topic types can be arranged in hierarchies
- subtype/supertype hierarchies have a specific syntax
- If A is a superclass of B, then
 - Both A and B must be classes
 - If C is an instance of B, it must also be an instance of A
 - If C is a subclass of B, it must also be a subclass of A, (in which case an instance of C is also an instance of B and an instance of A)



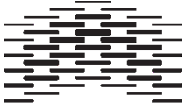
Topic maps and RDF

"Surface" level similarities

- both are standards for ontology modelling
- both use XML
- both use URIs for securing identity
- both have constraint and query languages

Important differences

- they are optimized for different purposes; TM for human reading, RDF for machine processing
- RDF is document centric whereas TM is subject centric



OSLO AND AKERSHUS
UNIVERSITY COLLEGE
OF APPLIED SCIENCES

- Thank you!
- Any questions?