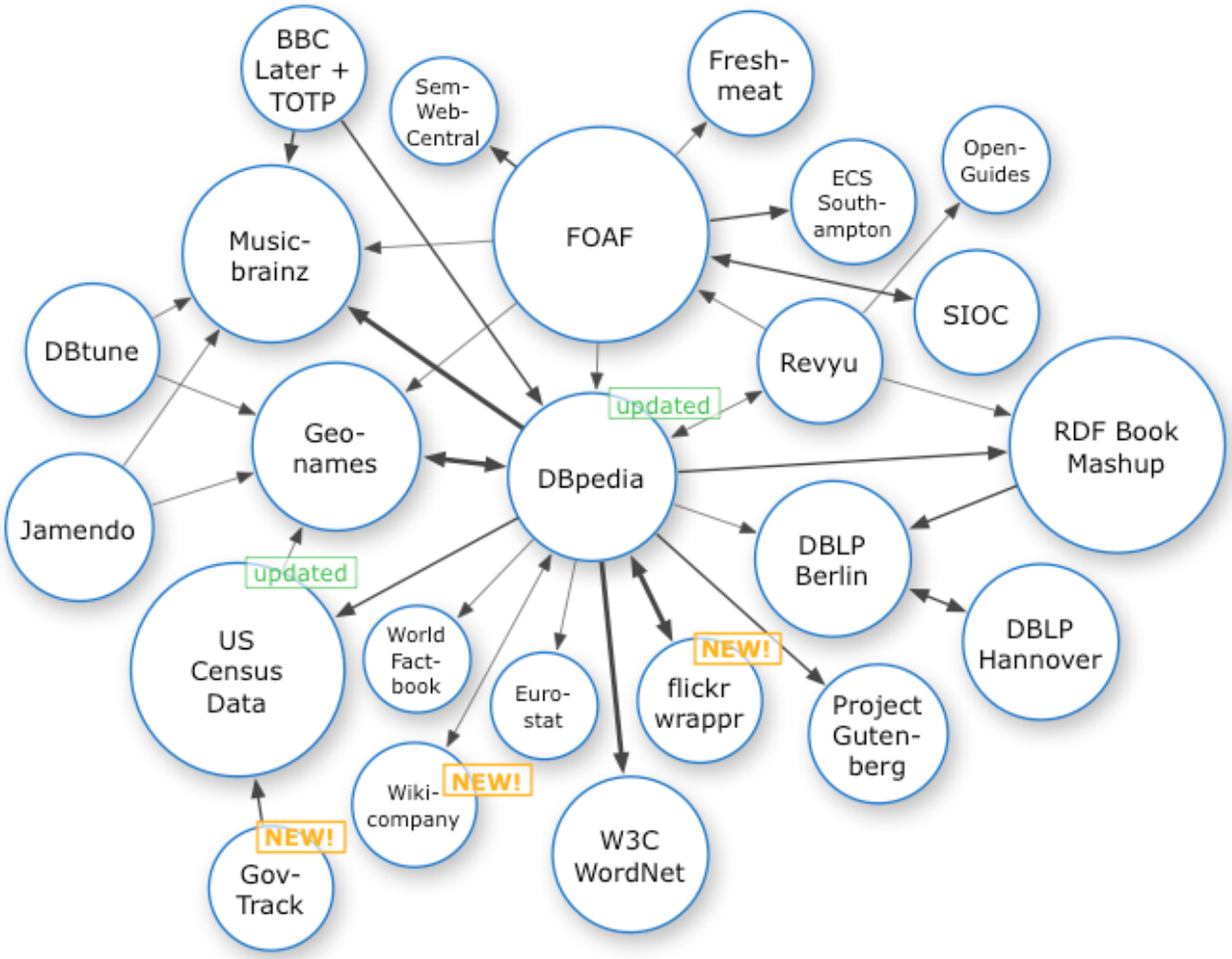
Recent Developments in Web-enabled Vocabularies

Tom Baker

MPG EScience Seminar on Metadata Infrastructures

Berlin, 15 October 2008

Linked Data

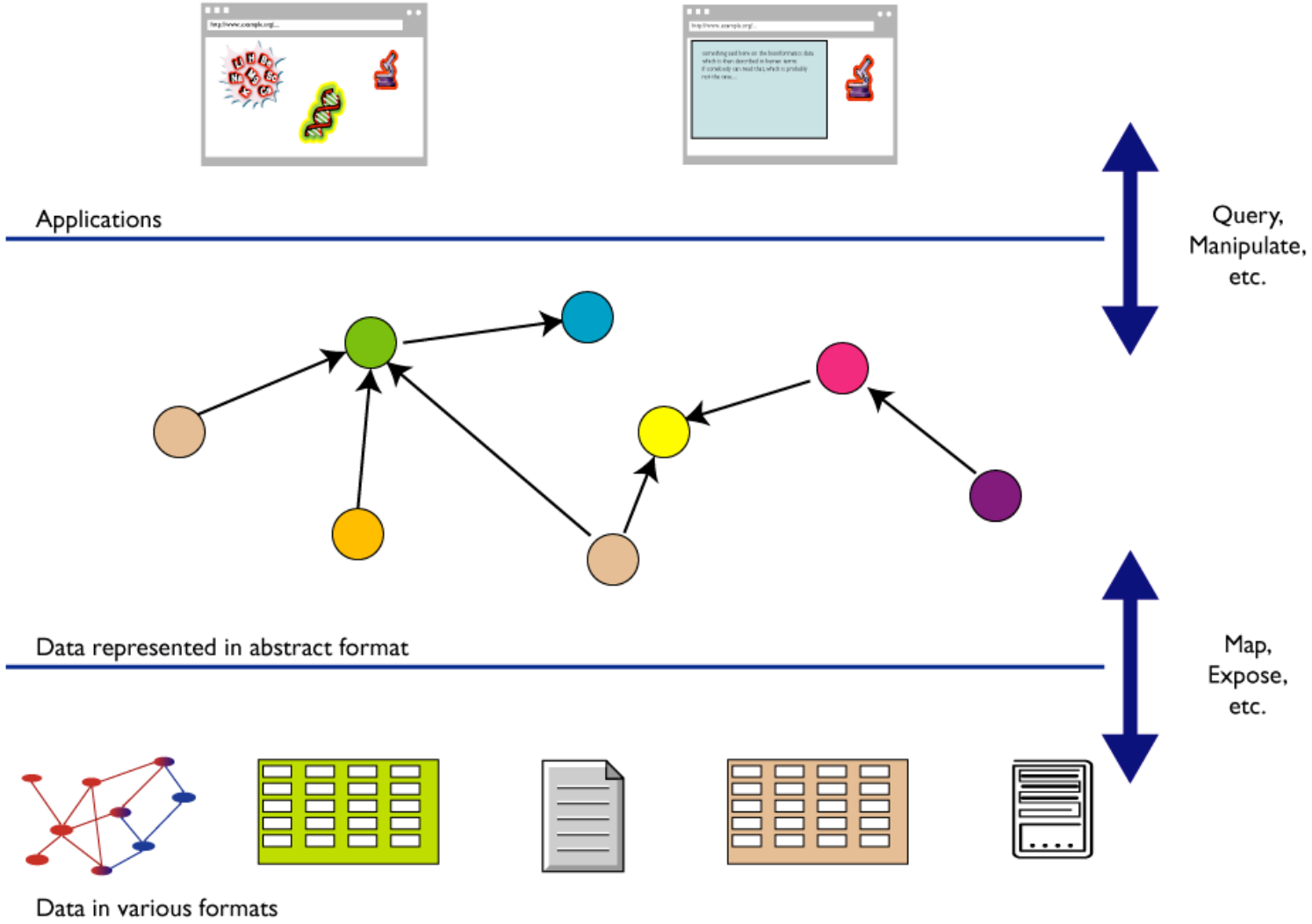


http://dbpedia.org

Open Linked Data

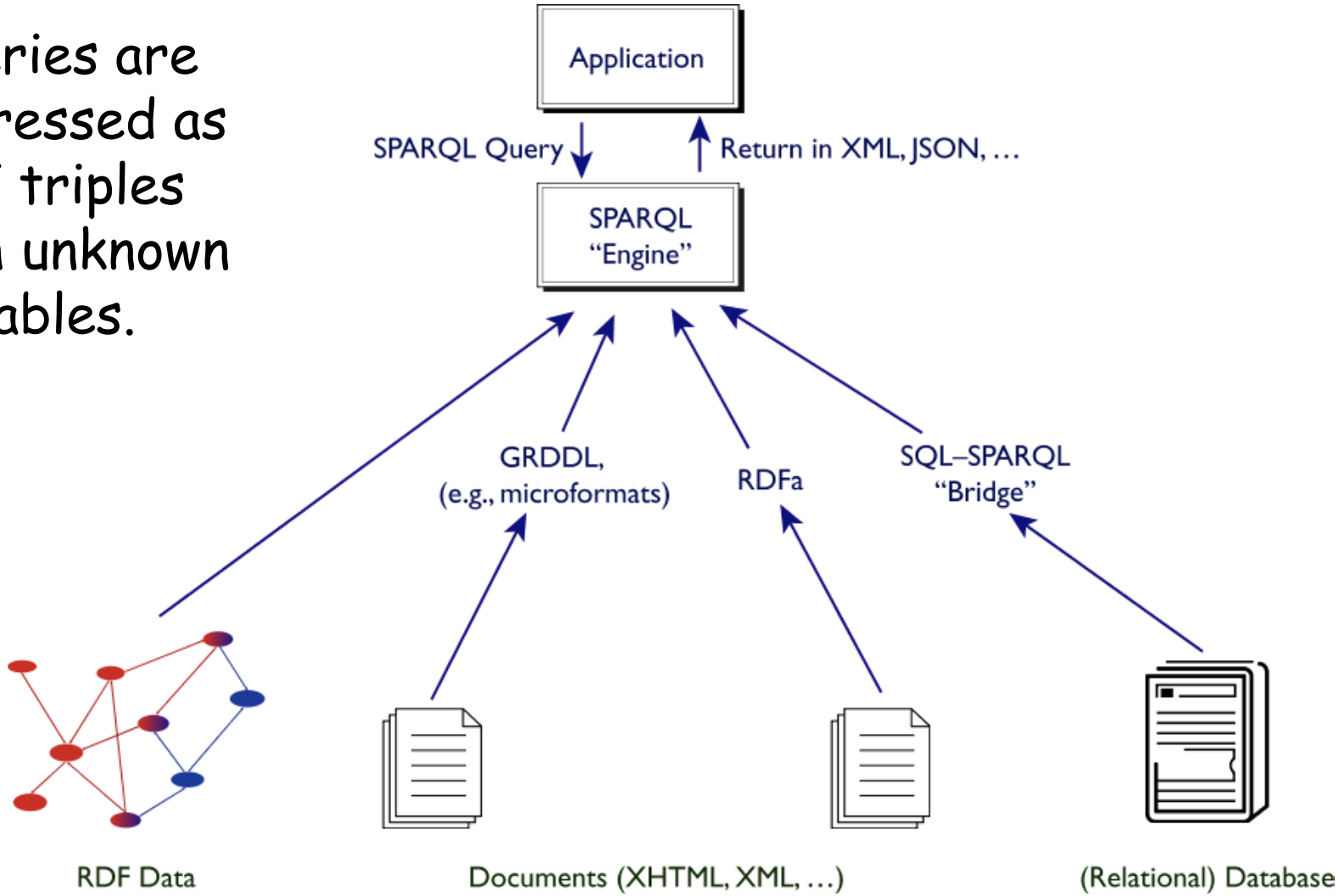
- Value of information is very much a function of what it links to (Tim Berners-Lee)
- Four rules (for maximizing unplanned re-use)
  - 1. Identify things with URIs.
  - 2. Use HTTP URIs.
  - 3. Serve information on the Web against the URIs.
  - 4. Link related material.
- http://www.w3.org/DesignIssues/LinkedData.html

RDF as a common format for merging data

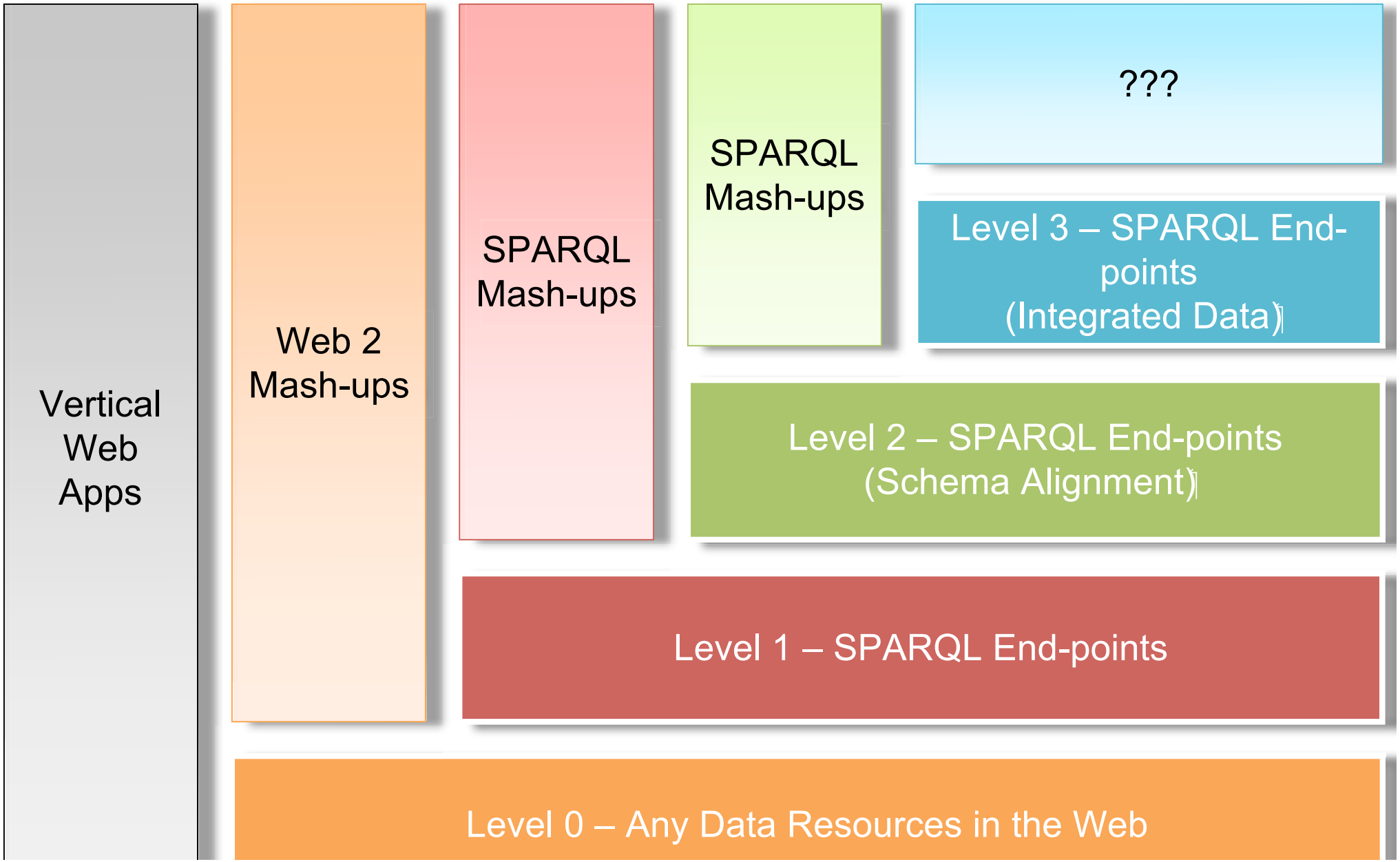


Merged data queried with SPARQL

Queries are expressed as RDF triples with unknown variables.



Data Web Layer Cake (Alistair Miles)



Interoperability Levels for Dublin Core metadata

- 4: Description Set Profile Interoperability
  - Shared formal vocabularies and constraints in records
- 3: Description Set syntactic interoperability
  - Shared formal vocabularies in exchangeable records
- 2: Semantic interoperability
  - Shared vocabularies based on formal semantics
- 1: Informal interoperability
  - Shared vocabularies defined in natural language

Getting data into triples

- **GRDDL** (Gleaning Resource Descriptions from Dialects of Languages)
  - Mechanism for extracting structured data from XML and XHTML documents and converting that data (via an XSLT script) into RDF
- **RDFa** ("RDF-in-attributes")
  - Embeds RDF-structured data into Web pages
  - Extends XHTML with attributes for carrying RDF data
  - Data can be cut-and-paste between RDF-aware applications
- **Microformats?**
  - Small XML formats to embed in Web pages
  - Work well for well-defined contact (hCard) or calendar (hCal) information, but fields not designed to be shared across formats
  - GRDDL can be used to extract RDF triples from Microformats

Tagging

- “Taggers” create “tags” about “resources”
  - Collectively, taggers create “folksonomies”
- Problem: to represent tagging data re-usably
  - Freely created tags are not associated with structured ontologies, hence not normalized
    - Countless variants, typos...
    - Conceptualizations may differ even when purposes same
    - Often no clear need to share idiosyncratic tags
  - Research topics: models for formalizing folksonomies, formal expression of tagging data

Extract data from wiki infoboxes...

WIKIPEDIA
The Free Encyclopedia

navigation
- Main page
- Contents
- Featured content
- Current events
- Random article

search
Go Search

interaction
- About Wikipedia
- Community portal
- Recent changes
- Contact Wikipedia
- Donate to Wikipedia
- Help

toolbox
- What links here
- Related changes
- Upload file
- Special pages
- Printable version
- Permanent link
- Cite this page

languages
- Afrikaans
- Alemannisch
- Anglo-Saxon
- العربية
- Aragonés

Charles Darwin

From Wikipedia, the free encyclopedia
(Redirected from Charles darwin)

For other people of the same surname, and places and things named after Darwin, see Darwin.

Charles Robert Darwin (February 12, 1809 – April 19, 1882) was an English naturalist,[1] who realised and demonstrated that all species of life have evolved over time from common ancestors through the process he called natural selection.[1] The fact that evolution occurs became accepted by the scientific community and the general public in his lifetime, while his theory of natural selection came to be widely seen as the primary explanation of the process of evolution in the 1930s,[1] and now forms the basis of modern evolutionary theory. In modified form, Darwin's scientific discovery remains the foundation of biology, as it provides a unifying logical explanation for the diversity of life.[2]

Darwin developed his interest in natural history while studying medicine at Edinburgh University, then theology at Cambridge.[3] His five-year voyage on the *Beagle* established him as an eminent geologist whose observations and theories supported Charles Lyell's uniformitarian ideas, and publication of his journal of the voyage made him famous as a popular author. Puzzled by the geographical distribution of wildlife and fossils he collected on the voyage, Darwin investigated the transmutation of species and conceived his theory of natural selection in 1838.[4] Although he discussed his ideas with several naturalists, he needed time for extensive research and his geological work had priority.[5] He was writing up his theory in 1858 when Alfred Russel Wallace sent him an essay which described the same idea, prompting immediate joint publication of both of their theories.[6]

His 1859 book *On the Origin of Species* established evolution by common descent as the dominant scientific explanation of diversification in nature. He examined human evolution and sexual selection in *The Descent of Man, and Selection in Relation to Sex*, followed by *The Expression of the Emotions in Man and Animals*. His research on plants was published in a series of books, and in his final book, he examined earthworms and their effect on soil.[7]

In recognition of Darwin's pre-eminence, he was one of only five 19th century UK non-royal personages to be honoured by a state funeral,[8] and was buried in Westminster Abbey, close to John Herschel and Isaac Newton.[9]

Contents [hide]

1 Biography
- 1.1 Early life
- 1.2 Journey of the Beagle
- 1.3 Inception of Darwin's evolutionary theory
- 1.4 Overwork, illness, and marriage
- 1.5 Preparing the theory of natural selection for publication
- 1.6 Publication of the theory of natural selection
- 1.7 Reaction to the publication
- 1.8 Descent of Man, sexual selection, and botany

Charles Darwin

Charles Robert Darwin (1809-1882). At the age o
Charles Darwin had just published
On the Origin of Species.

Born	12 February 1809 Mount House, Shrewsbury, Shropshire, England
Died	19 April 1882 (aged 73) Down House, Downe, Kent, Engl
Residence	England
Nationality	British
Fields	Naturalist
Institutions	Royal Geographical Society
Alma mater	University of Edinburgh University of Cambridge
Academic advisors	Adam Sedgwick John Stevens Henslow

Query on merged data in DBPedia

UNIVERSITÄT LEIPZIG pediaX

Query Wikipedia

This semantic database contains over 10 million statements extracted from the English Wikipedia.

search for queries | Most popular | Upcoming

- Tennis players from Moscow
- Sitcoms set in NYC
- People influenced by Friedrich Nietzsche
- Soccer player with tricot number 11 from club with stadium with >40000 seats born in a country with more than 10M inhabitants
- Film music composer born 1965
- Films longer than 5 hours
- Space Missions
- People being 1.80m tall
- List of Web browser software
- Battles in Saxony
- Mayors of US cities higher than 1000m
- Hip hop CDs from Texas Artists
- Pictures of American guitarists
- Scientists and their doctoral advisors
- Planes and their designers built in the 1st decate of the 20th century

<< 1 >>

More Information: at dbpedia.org and in the paper What have Innsbruck and Leipzig in common? Extracting Semantic from Wiki Content.

Contact: AKSW Workgroup @ BIS / Universität Leipzig

Scientists and their doctoral advisors

Modify this query or create your own!

Click on a column header to sort results on this page.

69 results found in 0.212s

Nr.	?advisee	?image	?advisor	?image_advisor
1	Charles Darwin		Adam Sedgwick	
2	William Hopkins		Adam Sedgwick	

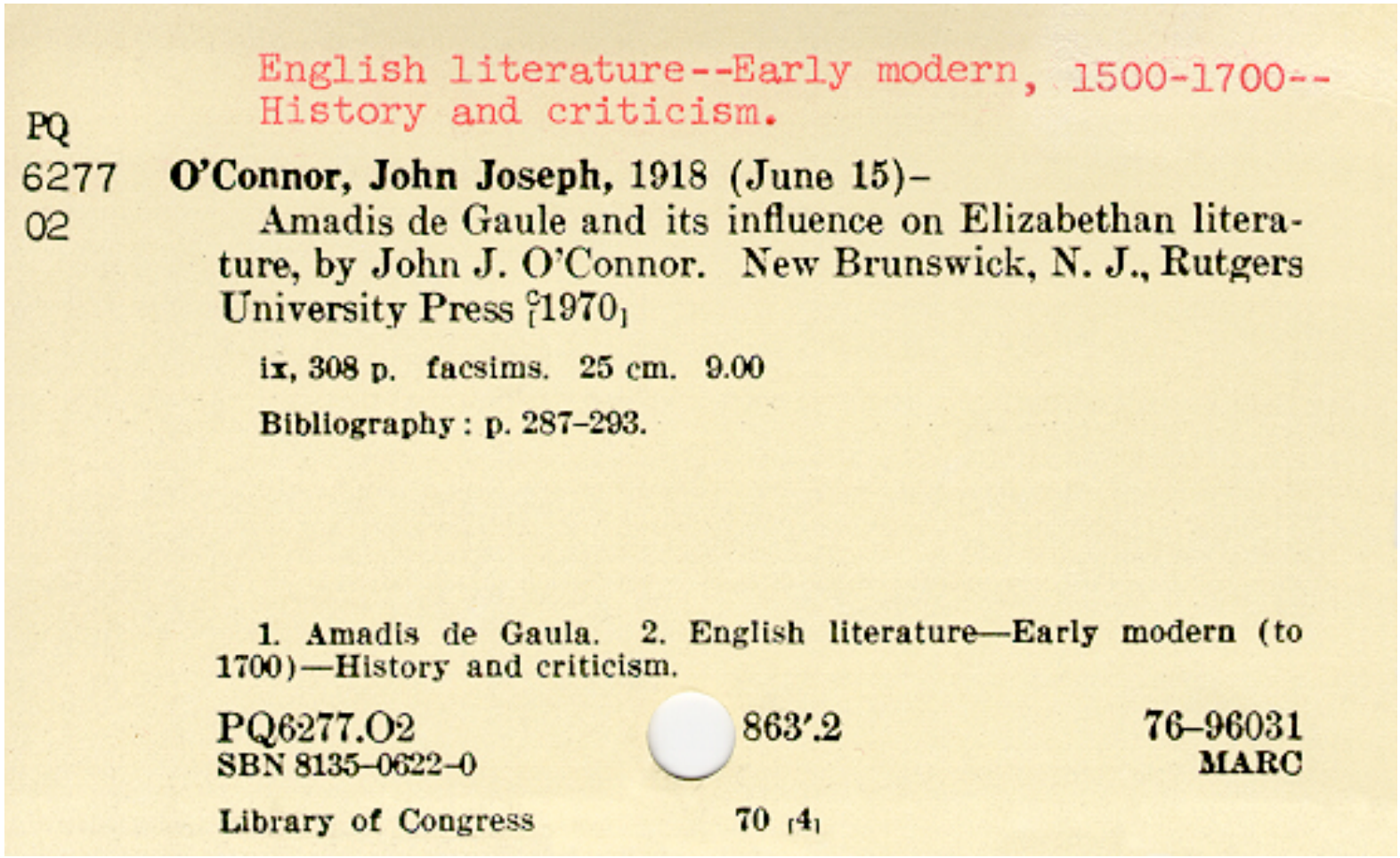
Language of data, written with URIs

- URIs are the words in a new language of data
  - “classes”/“properties” like “nouns”/“verbs”
    - `http://purl.org/dc/terms/title`
  - When grammatically used, URIs are the basis of meaningful interoperability
- Established Domain Name System (DNS)
  - Global delegation of maintenance authority
  - Web infrastructure provides “dictionary” of context-independent terms

Creating shared RDF vocabularies

- Vocabularies "born RDF"
  - Simple Knowledge Organization System (SKOS)
  - Web Ontology Language (OWL)
  - DCMI Metadata Terms
- Translation of pre-Web vocabularies into RDF
  - MARC Relators (roles of agents with respect to resources) - Done!
  - RDA (successor to Anglo-American Cataloging Rules)
  - Library of Congress value vocabularies in SKOS
  - FRBR (Func. Requirements for Bibliographic Records)
  - CIDOC CRM

Library catalog card



English literature--Early modern, 1500-1700--
History and criticism.

PQ
6277
02

O'Connor, John Joseph, 1918 (June 15)–

Amadis de Gaule and its influence on Elizabethan literature, by John J. O'Connor. New Brunswick, N. J., Rutgers University Press [©1970]

ix, 308 p. facsims. 25 cm. 9.00

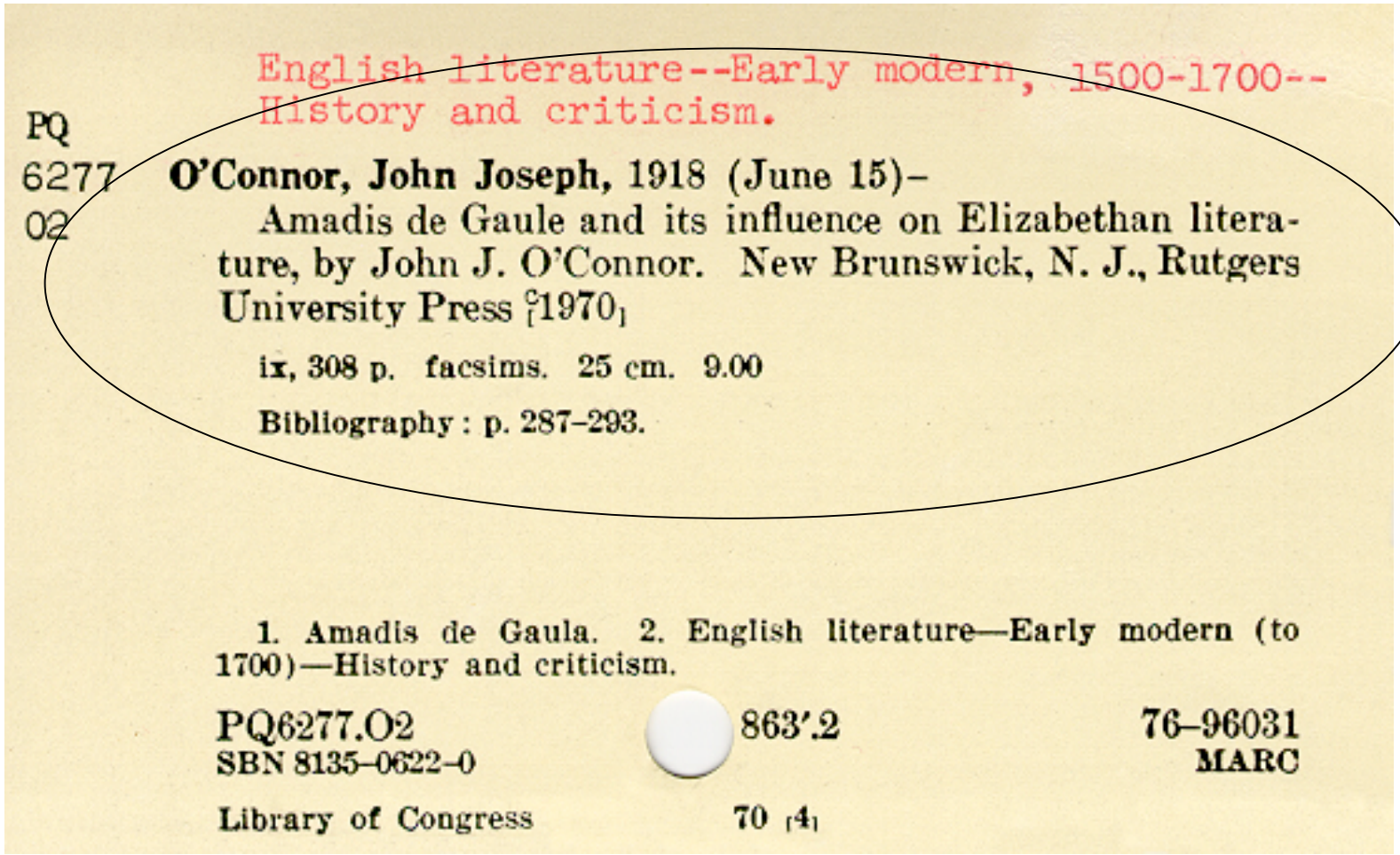
Bibliography: p. 287–293.

1. Amadis de Gaula. 2. English literature—Early modern (to 1700)—History and criticism.

PQ6277.O2 863'.2 76–96031
SBN 8135–0622–0 MARC

Library of Congress 70 [4]

Library catalog card – bibliographic description



English literature--Early modern, 1500-1700--
History and criticism.

PQ
6277
02

O'Connor, John Joseph, 1918 (June 15)–

Amadis de Gaule and its influence on Elizabethan literature, by John J. O'Connor. New Brunswick, N. J., Rutgers University Press [c1970]

ix, 308 p. facsims. 25 cm. 9.00

Bibliography: p. 287–293.

1. Amadis de Gaula. 2. English literature—Early modern (to 1700)—History and criticism.

PQ6277.O2 863'.2 76–96031
SBN 8135–0622–0 MARC

Library of Congress 70 [4]

Library catalog entities in RDF

- Resource Description and Access (RDA)
- Successor standard to Anglo-American Cataloging Rules (AACR)
- Based on FRBR
- Joint DCMI/RDA task force
- Prototype RDF vocabularies for bibliographic information
- Early stages

Making implicit relations explicit

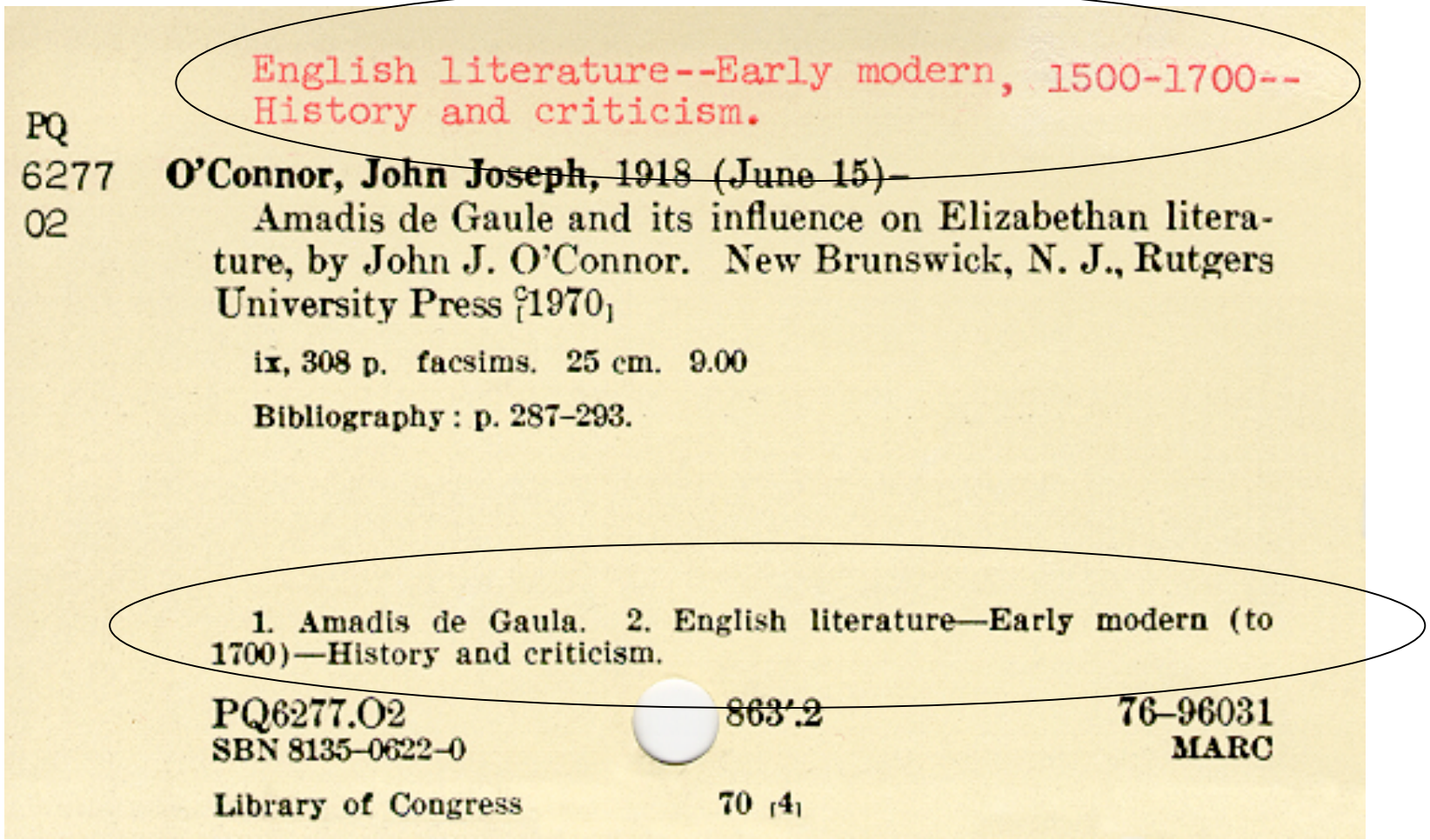
000 01683cam a22003253a 450
001 5235121
005 20041110071040.0
006 m | d |
007 cr||n|||||||n
008 860306s1743 enk|||| s 00| ||eng d
035 __ |**a** ESTCT161273
040 __ |**a** Uk-ES |**c** Uk-ES |**d** CStRLIN |**d** Thomson Gale |**d** NIC
130 0_ |**a** Bible. |**p** N.T. |**p** 1 Corinthians. |**l** English. |**p** Authorised. |**k** Selections.
245 14 |**a** The Christian's epinikion, or song of triumph |**h** [e ource]. |**b** A paraphrase on ch. XV. of St. Pa
260 __ |**a** London : |**b** printed for M. Cooper, |**c** 1743.
500 __ |**a** The dedication is signed: T. Morell.
500 __ |**a** Includes the text in the Authorised version.
500 __ |**a** Reproduction of original from the British Library.
510 4_ |**a** English Short Title Catalog, |**c** ESTCT161273.
533 __ |**a** Electronic reproduction. |**b** Farmington Hills, Mich. : |**c** Thomson Gale, |**d** 2003. |**n** Available via the World
700 1_ |**a** Morell, Thomas, |**d** 1703-1784.
730 0_ |**a** Christian's epinikion, or song of triumph.
856 40 |**u**
http://encompass.library.cornell.edu/cgi-bin/checkIP.cgi?access=gateway_standard%26url=http://galenet.galeg
|**z** Full text online

Record describes two versions

Original Version

Digital Version

Library catalog card – subject headings



English literature--Early modern, 1500-1700--
History and criticism.

PQ
6277
02

O'Connor, John Joseph, 1918 (June 15)–

Amadis de Gaule and its influence on Elizabethan literature, by John J. O'Connor. New Brunswick, N. J., Rutgers University Press [c1970]

ix, 308 p. facsims. 25 cm. 9.00

Bibliography: p. 287–293.

1. Amadis de Gaula. 2. English literature—Early modern (to 1700)—History and criticism.

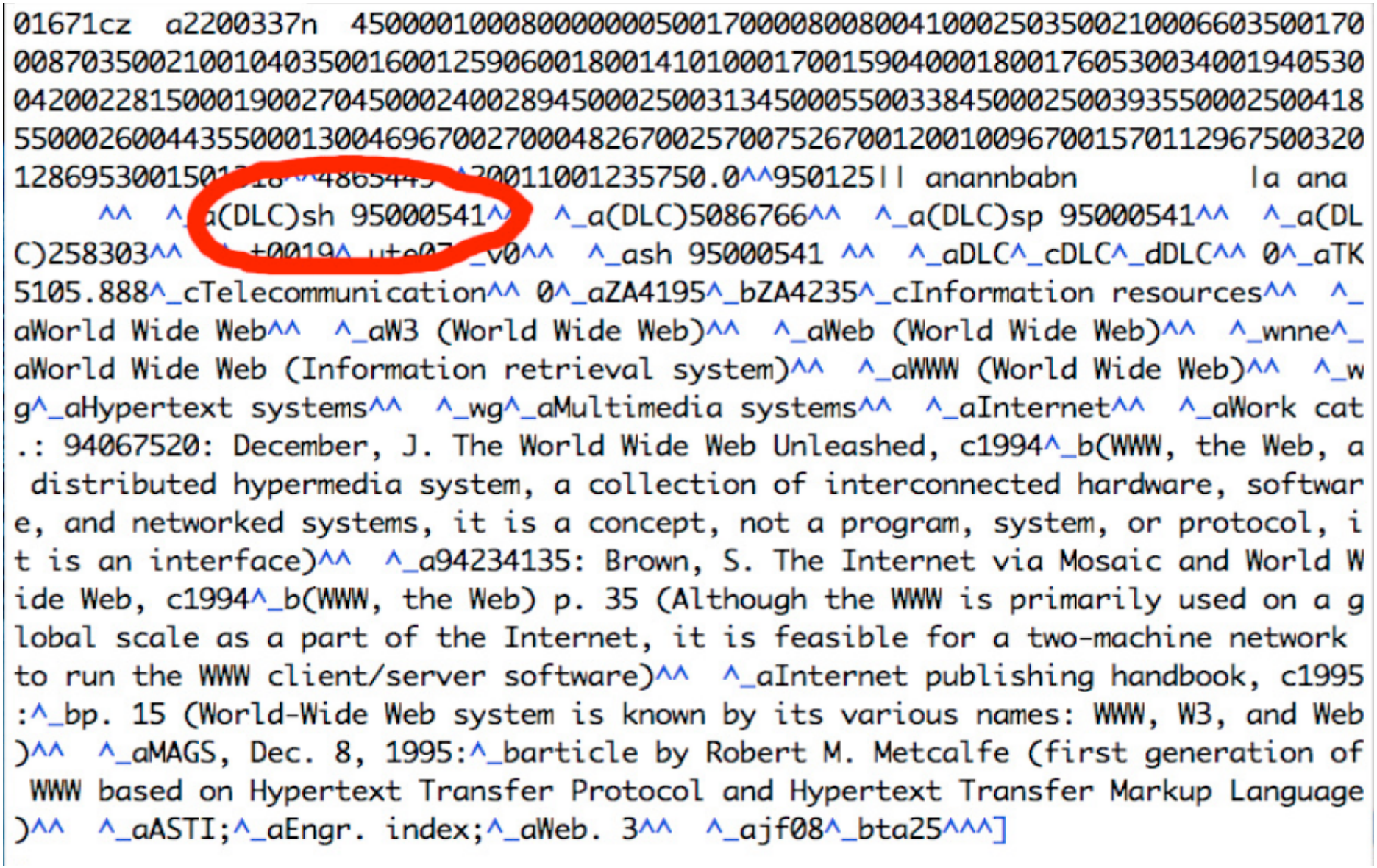
PQ6277.O2 863'.2 76–96031
SBN 8135–0622–0 MARC

Library of Congress 70 [4]

Converting Library of Congress Subject Headings...

```
01671cz  a2200337n  450000100080000000500170000800800410002503500210006603500170
008703500210010403500160012590600180014101000170015904000180017605300340019405300
042002281500019002704500024002894500025003134500055003384500025003935500025004185
550002600443550001300469670027000482670025700752670012001009670015701129675003200
1286953001501318^^4865449^^20011001235750.0^^950125|| anannbabn          |a ana
     ^^  ^_a(DLC)sh 95000541^^  ^_a(DLC)5086766^^  ^_a(DLC)sp 95000541^^  ^_a(DL
C)258303^^  ^_t0019^_ute07^_v0^^  ^_ash 95000541 ^^  ^_aDLC^_cDLC^_dDLC^^ 0^_aTK
5105.888^_cTelecommunication^^ 0^_aZA4195^_bZA4235^_cInformation resources^^  ^_
aWorld Wide Web^^  ^_aW3 (World Wide Web)^^  ^_aWeb (World Wide Web)^^  ^_wnne^_
aWorld Wide Web (Information retrieval system)^^  ^_aWWW (World Wide Web)^^  ^_w
g^_aHypertext systems^^  ^_wg^_aMultimedia systems^^  ^_aInternet^^  ^_aWork cat
.: 94067520: December, J. The World Wide Web Unleashed, c1994^_b(WWW, the Web, a
 distributed hypermedia system, a collection of interconnected hardware, softwar
e, and networked systems, it is a concept, not a program, system, or protocol, i
t is an interface)^^  ^_a94234135: Brown, S. The Internet via Mosaic and World W
ide Web, c1994^_b(WWW, the Web) p. 35 (Although the WWW is primarily used on a g
lobal scale as a part of the Internet, it is feasible for a two-machine network
to run the WWW client/server software)^^  ^_aInternet publishing handbook, c1995
:^_bp. 15 (World-Wide Web system is known by its various names: WWW, W3, and Web
)^^  ^_aMAGS, Dec. 8, 1995:^_barticle by Robert M. Metcalfe (first generation of
 WWW based on Hypertext Transfer Protocol and Hypertext Transfer Markup Language
)^^  ^_aASTI;^_aEngr. index;^_aWeb. 3^^  ^_ajf08^_bta25^^^]
```

Using an existing code as basis for URIs...



...and representing it in RDF

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
   xmlns:dcterms="http://purl.org/dc/terms/"
   xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
   xmlns:skos="http://www.w3.org/2004/02/skos/core#"
>
  <rdf:Description rdf:about="http://lcsh.info/sh95000541#concept">
    <skos:prefLabel xml:lang="en">World Wide Web</skos:prefLabel>
    <skos:narrower rdf:resource="http://lcsh.info/sh2003001415#concept"/>
    <skos:narrower rdf:resource="http://lcsh.info/sh97003254#concept"/>
    <skos:narrower rdf:resource="http://lcsh.info/sh2002000569#concept"/>
    <skos:inScheme rdf:resource="http://lcsh.info/"/>
    <dcterms:created rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2000-04-28</dcterms:created>
    <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
    <skos:broader rdf:resource="http://lcsh.info/sh88002671#concept"/>
    <skos:broader rdf:resource="http://lcsh.info/sh92002381#concept"/>
    <skos:related rdf:resource="http://lcsh.info/sh92002816#concept"/>
    <skos:editorialNote xml:lang="en">94234135: Brown, S. The Internet via Mosaic and World Wide Web, c1994 (WWW, the
    <skos:editorialNote xml:lang="en">ASTI; Engr. index; Web. 3</skos:editorialNote>
    <skos:editorialNote xml:lang="en">Work cat.: 94067520: December, J. The World Wide Web Unleashed, c1994 (WWW, the
    <skos:editorialNote xml:lang="en">MAGS, Dec. 8, 1995: article by Robert M. Metcalfe (first generation of WWW base
    <skos:editorialNote xml:lang="en">Internet publishing handbook, c1995: p. 15 (World-Wide Web system is known by :
    <dcterms:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2001-10-01T09:56:06</dcterms:modified>
    <skos:altLabel xml:lang="en">WWW (World Wide Web)</skos:altLabel>
    <skos:altLabel xml:lang="en">W3 (World Wide Web)</skos:altLabel>
    <skos:altLabel xml:lang="en">World Wide Web (Information retrieval system)</skos:altLabel>
    <skos:altLabel xml:lang="en">Web (World Wide Web)</skos:altLabel>
    <dcterms:LCC>TK5105.888</dcterms:LCC>
  </rdf:Description>
</rdf:RDF>
```

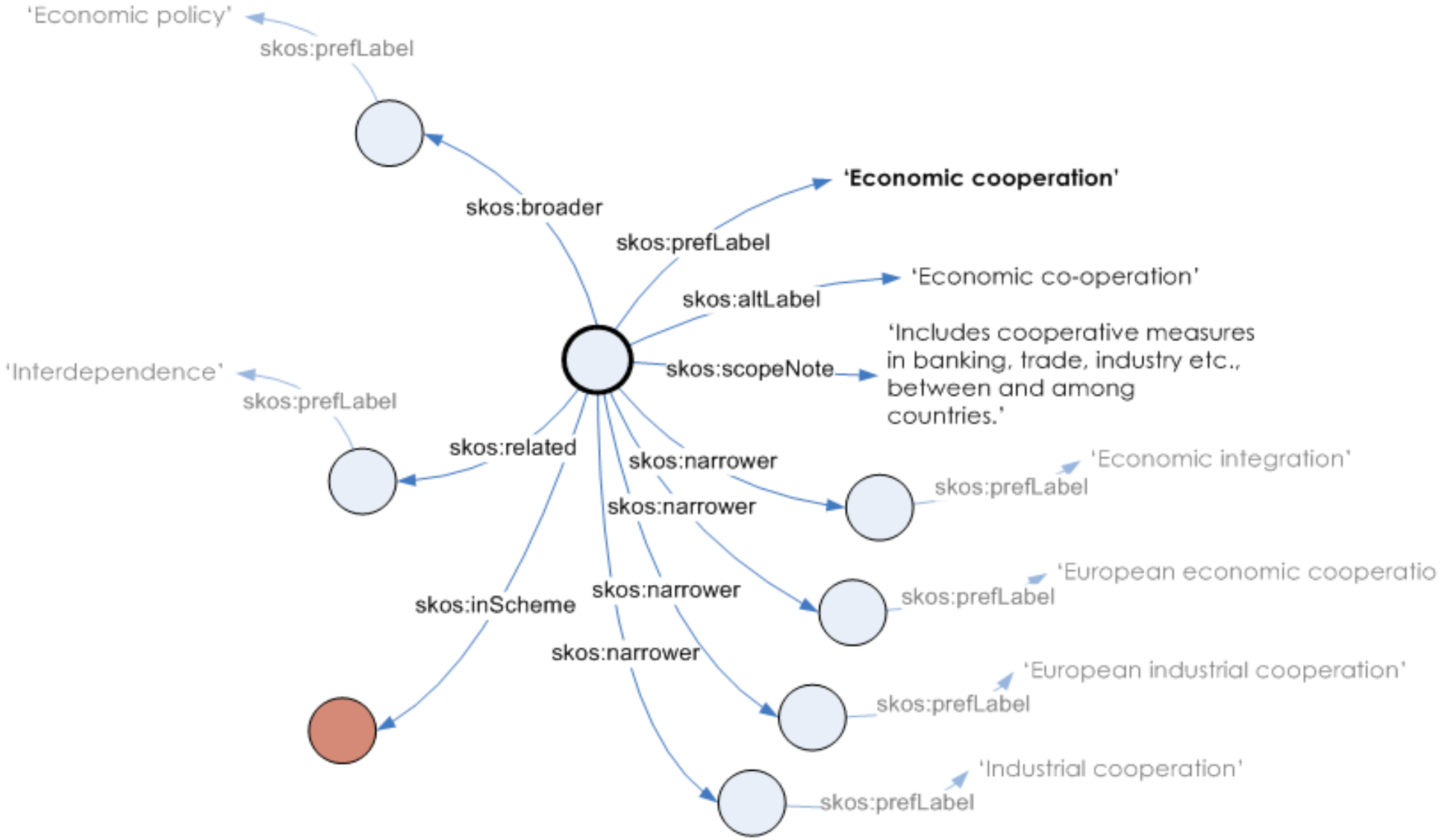
SKOS Example

UK Archival Thesaurus

Term:	**Economic cooperation**
Used For:	Economic co-operation
Broader terms:	Economic policy
Narrower terms:	Economic integration, European economic cooperation, European industrial cooperation, Industrial cooperation
Related terms:	Interdependence
Scope Note:	Includes cooperative measures in banking, trade, industry etc., between and among countries.

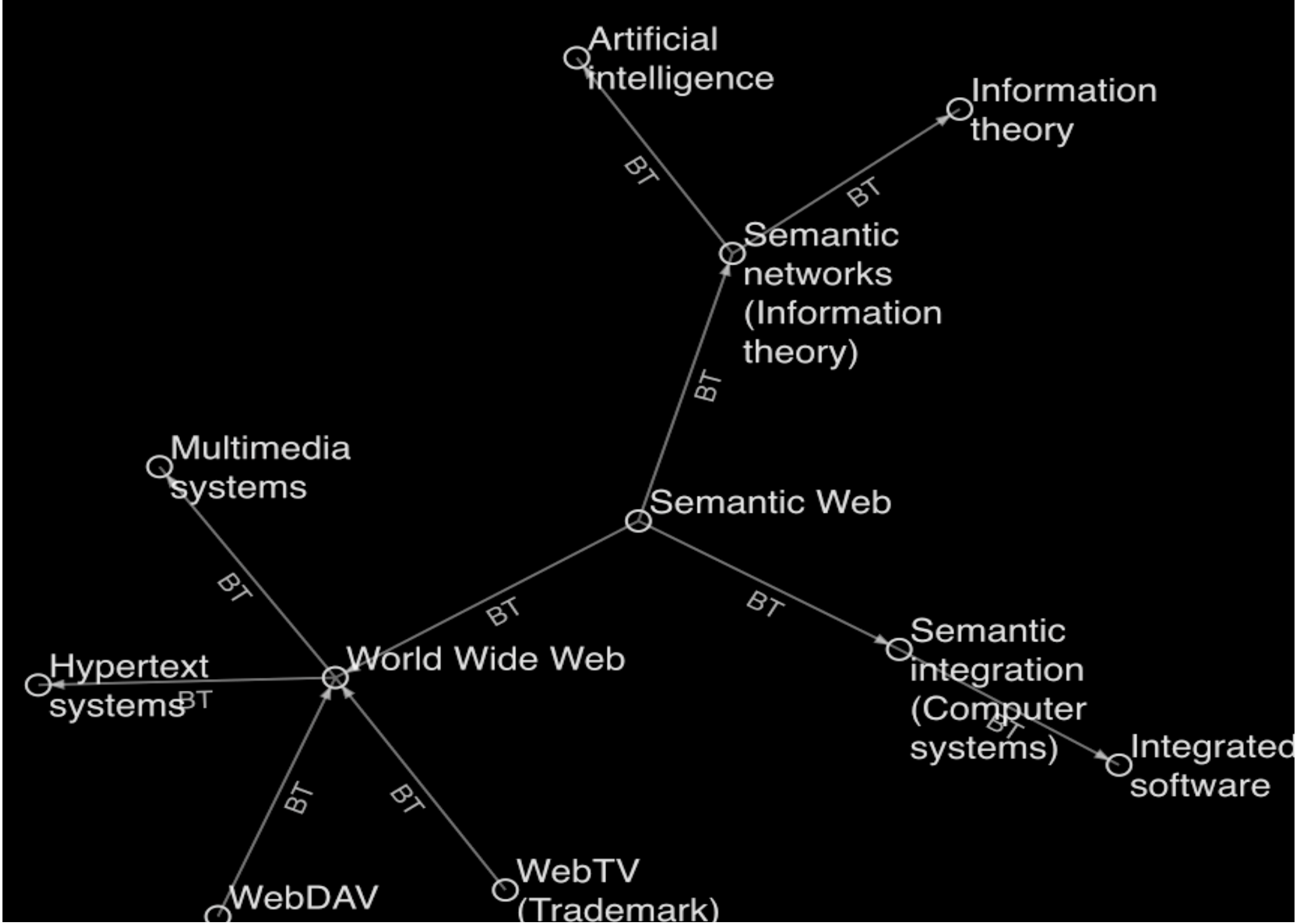
SKOS Example

UK Archival Thesaurus in SKOS



prefix skos: <http://www.w3.org/2004/02/skos/core#>

http://inkdroid.org/bzr/lcsh/docs/slides/


Artificial intelligence
Information theory
BT
BT
Semantic networks (Information theory)
BT
Multimedia systems
Semantic Web
BT
BT
BT
Semantic integration (Computer systems)
Hypertext systems
BT
World Wide Web
BT
Integrated software
BT
BT
WebDAV
WebTV (Trademark)

```
@prefix dc: <http://purl.org/dc/terms/> .

<http://lccn.loc.gov/99027665>
  dc:title "Weaving the Web : the original design
    and ultimate destiny of the World Wide Web by
    its inventor /" ;
  dc:creator "Berners-Lee, Tim." ;
  dc:creator "Fischetti, Mark." ;
  dc:type "text" ;
  dc:publisher "San Francisco : HarperSanFrancisco" ;
  dc:date "1999" ;
  dc:language "eng" ;
  dc:description "Includes index." ;
  dc:subject "Berners-Lee, Tim." ;
  dc:subject "World Wide Web" ;
  dc:identifier "URN:ISBN:0062515861 (cloth)" ;
  dc:identifier "URN:ISBN:006251587X (paper)" .
```

```
@prefix dc: <http://purl.org/dc/terms/> .

<http://lccn.loc.gov/99027665>
  dc:title "Weaving the Web : the original design
    and ultimate destiny of the World Wide Web by
    its inventor /" ;
  dc:creator "Berners-Lee, Tim." ;
  dc:creator "Fischetti, Mark." ;
  dc:type "text" ;
  dc:publisher "San Francisco : HarperSanFrancisco" ;
  dc:date "1999" ;
  dc:language "eng" ;
  dc:description "Includes index." ;
  dc:subject "Berners-Lee, Tim." ;
  dc:subject <http://loc.gov/lcsh/sh95000542> ;
  dc:identifier <urn:isbn:0062515861> ;
  dc:identifier <urn:isbn:006251587X> .
```

Interoperability Levels for Dublin Core metadata

- 4: Description Set Profile Interoperability
  - Shared formal vocabularies and constraints in records
- 3: Description Set syntactic interoperability
  - Shared formal vocabularies in exchangeable records
- 2: Semantic interoperability
  - Shared vocabularies based on formal semantics
- 1: Informal interoperability
  - Shared vocabularies defined in natural language

Interoperability Levels for Dublin Core metadata

- Level 1: Informal interoperability
  - Shared concepts with natural-language definitions
    - DCMI terms have carefully crafted, well-known definitions
  - No use of formal models or term URIs
  - Test: Is there a mapping to shared elements?
  - Example: IEEE LOM reuses some definitions and maps to 15-element “Dublin Core” (ISO 15836)

Interoperability Levels for Dublin Core metadata

- Level 2: Semantic Interoperability
  - Correct use of formal RDF graph model with conformant vocabularies (eg DCMI Metadata terms)
  - Use of URIs and formal semantic relationships between terms (eg subproperties)
  - Test: Is there a mapping to RDF triples?
  - Examples:
    - All RDF data (by definition)
    - All RDF data extracted from non-RDF formats (eg via GRDDL transforms)
    - All XHTML or HTML data using RDFa or DC-HTML/2008.

Interoperability Levels for Dublin Core metadata

- Level 3: Description set syntactic interoperability
  - Level-2-compatible data packaged in bounded description sets (records) as per DCMI Abstract Model (DC-AM)
  - Conventions for citing vocabulary encoding schemes (controlled vocabularies)
  - Test: Is there a mapping to “Expressing Dublin Core metadata using the DC-Text format”?
  - Examples: All data using DC-AM-compatible specifications, such as DC-DS-XML.

Interoperability Levels for Dublin Core metadata

- Level 4: Description Set Profile Interoperability
  - Level-3-compatible data that follows the specification “Description Set Profiles: A constraint language for Dublin Core Application Profiles”
  - Additional interoperability via shared Functional Requirements and Domain Model (“Singapore Framework for Dublin Core Application Profiles”)
  - Test: Is there a mapping to DSP constraints?
  - Examples:
    - Scholarly Works Application Profile