語意網與鏈結資料

柯皓仁 教授
國立臺灣師範大學圖書資訊學研究所

Outline

- Semantic Web Overview
- Linked Data in Brief
- Web of Data – Results of Linked Data
- Resource Description Framework (RDF)
- Principles of Linked Data
- Four Rules, Five Stars, and A Plan
- Implementation Framework
- Conclusion

SEMANTIC WEB OVERVIEW

鏈結(開放)資料

- 鏈結(開放)資料 = Linked (Open) Data, L(O)D
- Linked Data乃是一組最佳實務，用以將結構化資料發布在Web上，並加以串聯
- Linked Data 是實現語意網(Semantic Web)的具體方案
- 美國國會圖書館的一段Linked Open Data短片…呃，希望上完課後你全懂了
- 從語意網談起…

Weaving the Web – Vision of Tim Berners-Lee

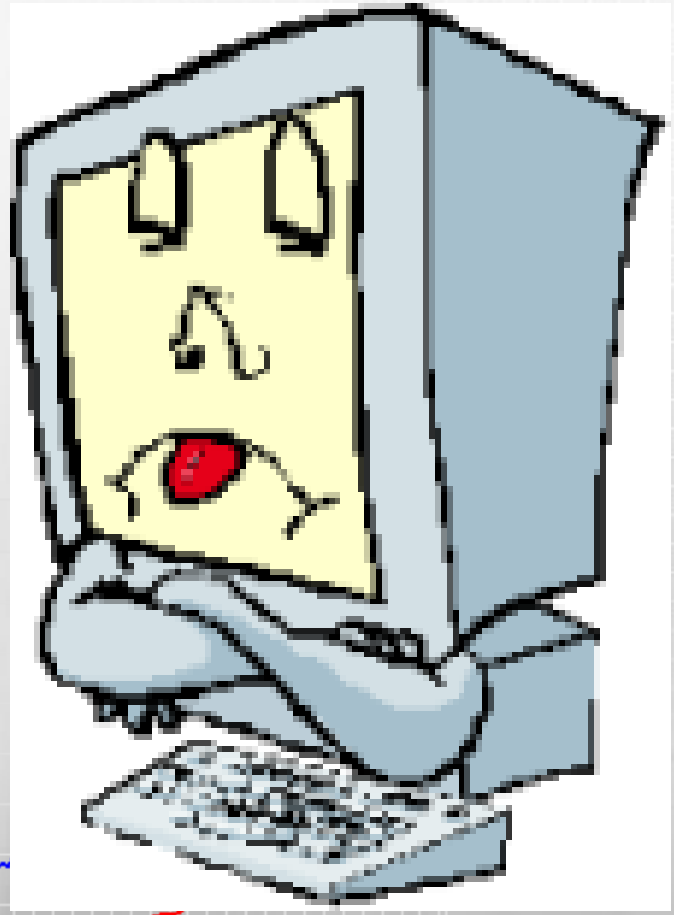
- The first step is putting data on the Web in a form that machines can naturally understand, or converting it to that form. This creates what I call a Semantic Web (SW) – a web of data that can be processed directly or indirectly by machines

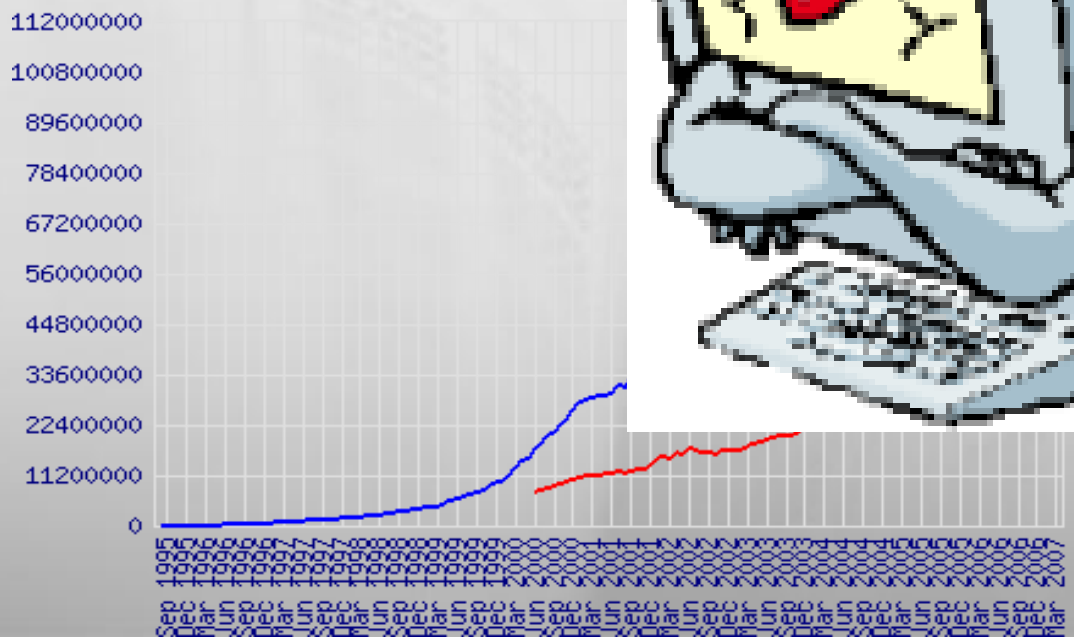
  -- Tim Berners-Lee *Weaving the Web*, Harper San Francisco, 1999

- Require that there be a machine-understandable semantics for some or all of the information presented in the WWW

Web Today

- Information repository.
- Simplicity – HTML
- Primarily for human interpretation and use.





Semantic Web Applications

- Software agents can then be built that can understand the product information the Web sites provide
- Meta-online stores can then be constructed with little effort, and this technique will also enable complete market transparency in various dimensions of diverse product properties

Semantic Web Applications (Cont.)

- The low-level programming of wrappers based on text extraction and format heuristics will be replaced by semantic mappings that translate different formats used to represent products and can be used to navigate and search automatically for the required information

Meta-online Store – Findbook.tw

御手洗潔的問候

作　　者：島田莊司
編／譯者：郭清華
出版社：皇冠文化
ISBN：9573322897
出版日：2006.12.18
分　　級：普級
語　　言：中文
規　　格：中文平裝

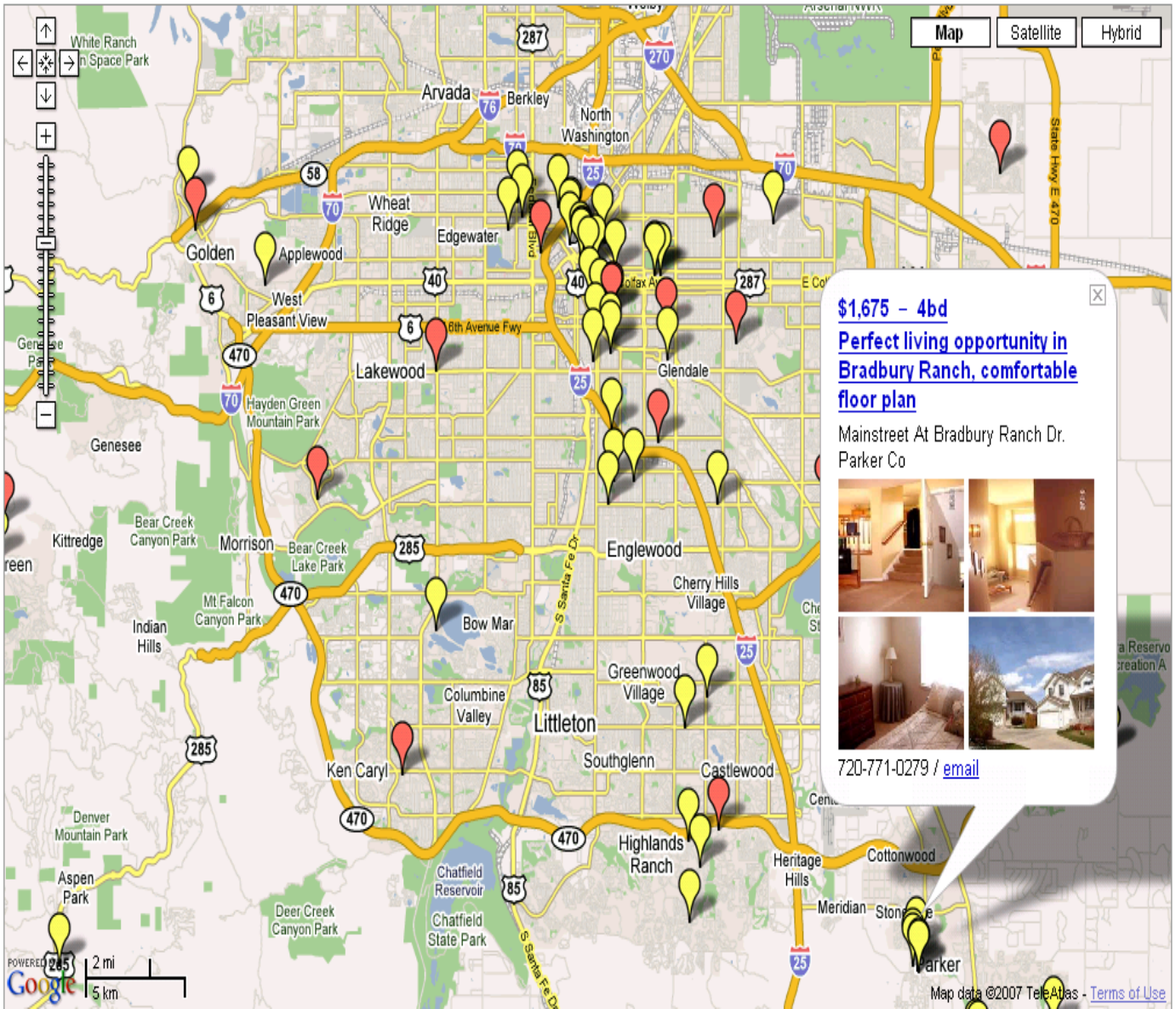
定　　價：280元
特　　價：82折!230元
用紅利價：79折!220元 (回饋4點)
運　　費：0元(限到店取貨)

參考庫存量:9本

加入購物清單
加入我的書店
推薦給好友


臺灣大學整合查詢系統 - Microsoft Internet Explorer
檔案(F) 編輯(E) 檢視(V) 我的最愛(A) 工具(T) 說明(H)
http://metacat.ntu.edu.tw/metacat/app
上一頁 搜尋 我的最愛
網址(D) http://metacat.ntu.edu.tw/metacat/app
移至 連結
搜尋 登入
網頁翻譯 信箱 知識+ 拍賣 交友 購物 股市 新聞 家族 服務
Met@Cat
國立台灣大學圖書館整合查詢館藏服務
書名 / 期刊名 作者 標題 ISBN ISSN 關鍵字
天龍八部
查詢
目前查詢範圍：主要圖書館(三十萬冊以上)
檢查圖書館列表 更改設定

Housingmaps.com -- Mashup



pics	price	bd	description	city	date
	$1595	3bd	W/Loft New Appliancecs 1700+Sq Ft Remodeled Mtn Views	Highlands R	5/27
	$1600	3bd	Near Dtc, Great Home/Neighborhood, 1900 Sft of Living Space	Aurora	5/27
	$1675	4bd	Perfect living opportunity in Bradbury Ranch, comfortable floor plan	Parker	5/27
	$1500	3bd	3BR/ 2BA House Avail. 6/16 Must See!	Boulder	5/26
	$1800	2bd	Denver Dry Lofts - Live in the heart of Downtown!	Denver	5/26
	$1850	4bd	House for Rent; Great condition, great area; July 1	Denver	5/26
	$1500	2bd	Total Remodel in Wash Park, 2BR, 2BA, Gorgeous!	Denver	5/26
	$1750	3bd	North Evergreen Townhome/2 Car attached garage	Evergreen	5/26
	$1650	3bd	Comfortable Ranch House, Martin Acres, Air Conditioned	Boulder	5/26
	$1695	4bd	Rock Creek home on Open Space	Superior	5/26
	$1875		Heart of Cherry Creek	Denver	5/26
	$1899	5bd	5 BR/ 3 BA mountain home	Evergreen	5/26
	$1675	3bd	Prestigious private residence in very conv. loc. Bad credit Ok.	Parker	5/26
	$1600	5bd	5 Bdrm, 4 Bath, 3 Car Garage, House For Rent	Denver	5/26
	$2000	3bd	Historic home	Denver	5/26
	$1500	5bd	Rent To Own Large Arvada Home - 2400 sq ft.	Arvada	5/26
	$1800	4bd	Du / Harvard Park Live On Park Perfect Location	Denver	5/26
	$1675	1bd	New Listing >Downtown Condo< Brand New w/Fitness Center, Doorman.	Denver	5/26

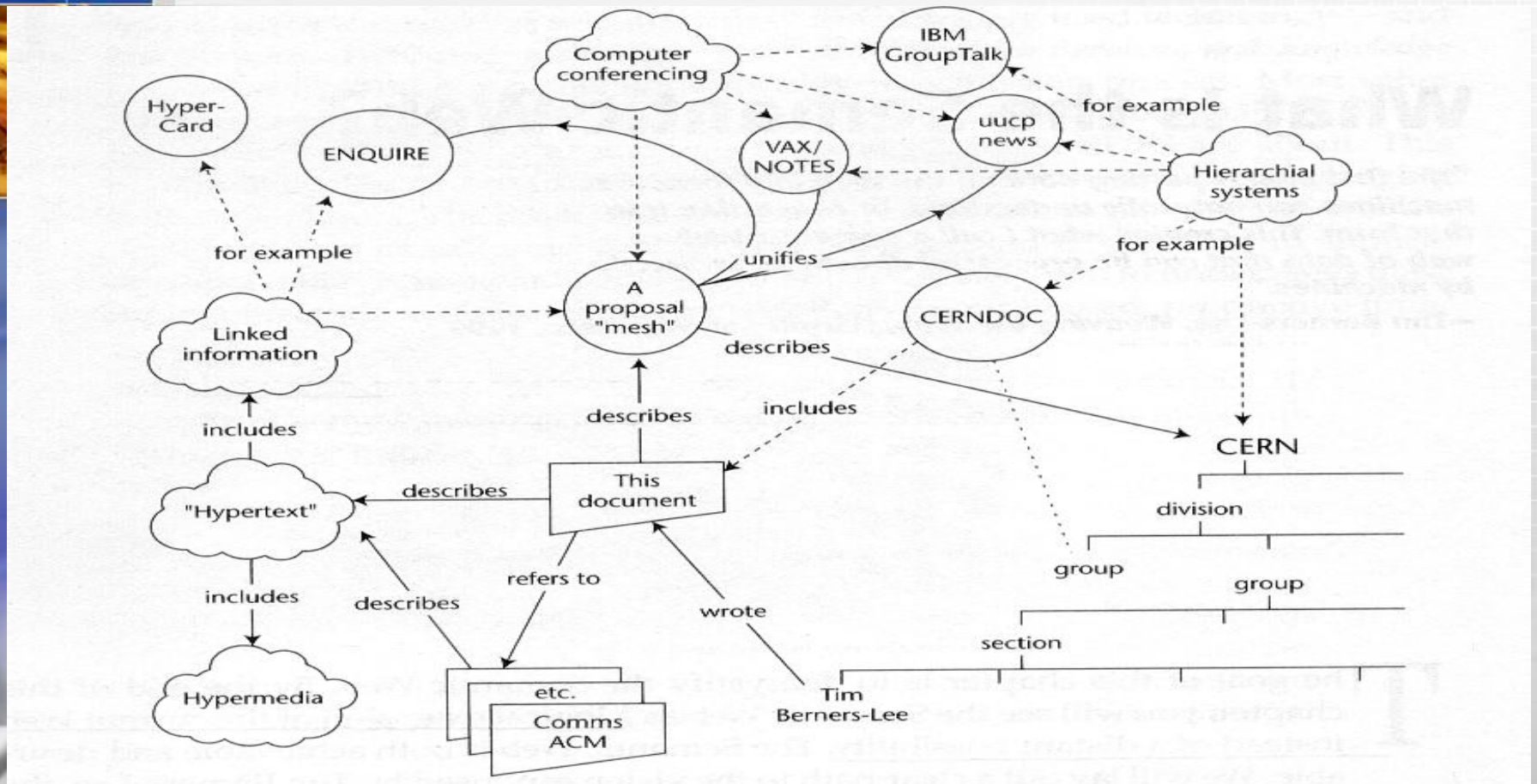
Toward Semantic Web

Information repository → Information-service provider

Vision of Tim Berners-Lee for the Future of the Web

- Make the Web understandable, and thus processable, by machines
- Original Web proposal to CERN
  - Relations between information items like "includes," "describes,", and "wrote" ➔ not currently captured on the Web
    - ✓ Use RDF to capture such relationship

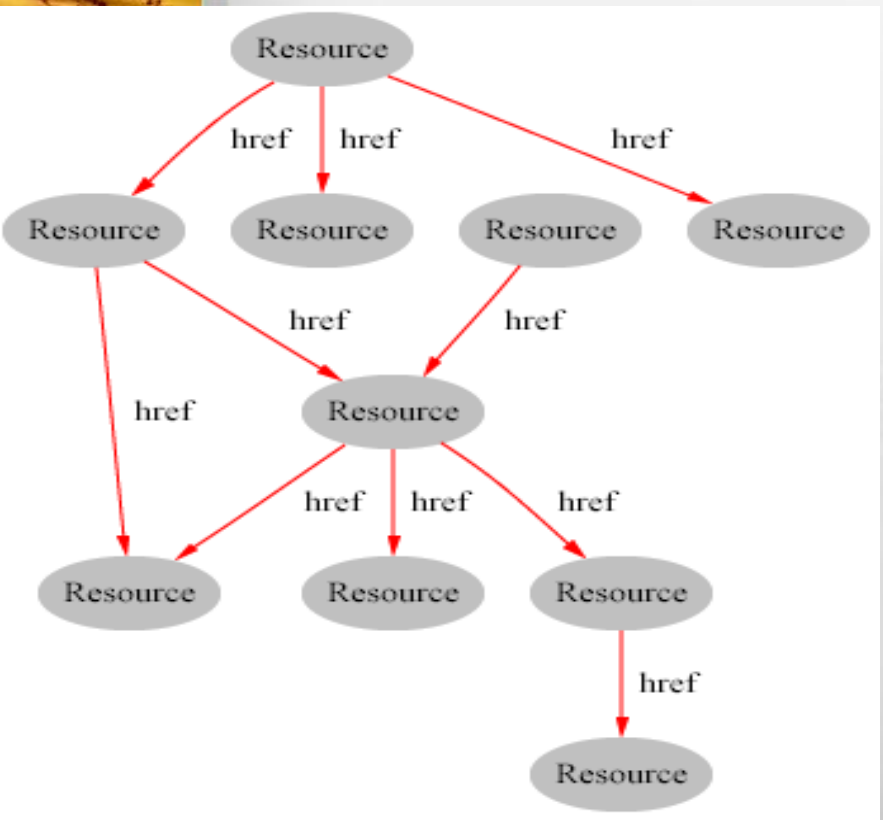
Original Web Proposal to CERN



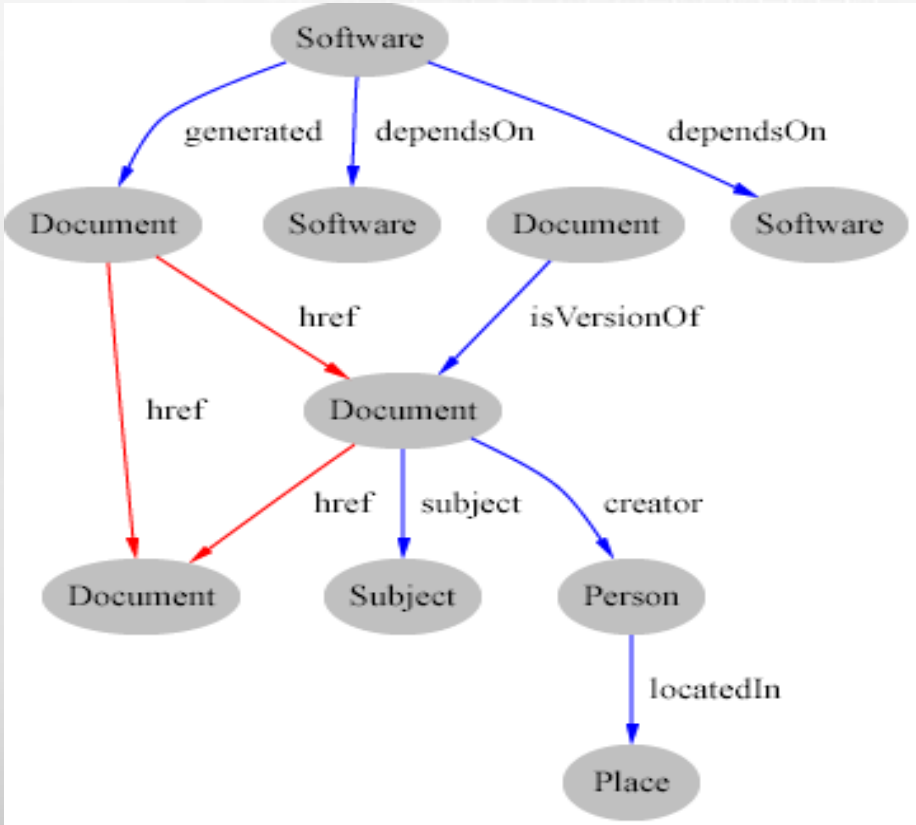
Encompass additional metadata above and beyond what is currently in the Web. This additional metadata is needed for machines to be able to process information on the Web

Web vs. Semantic Web

Web



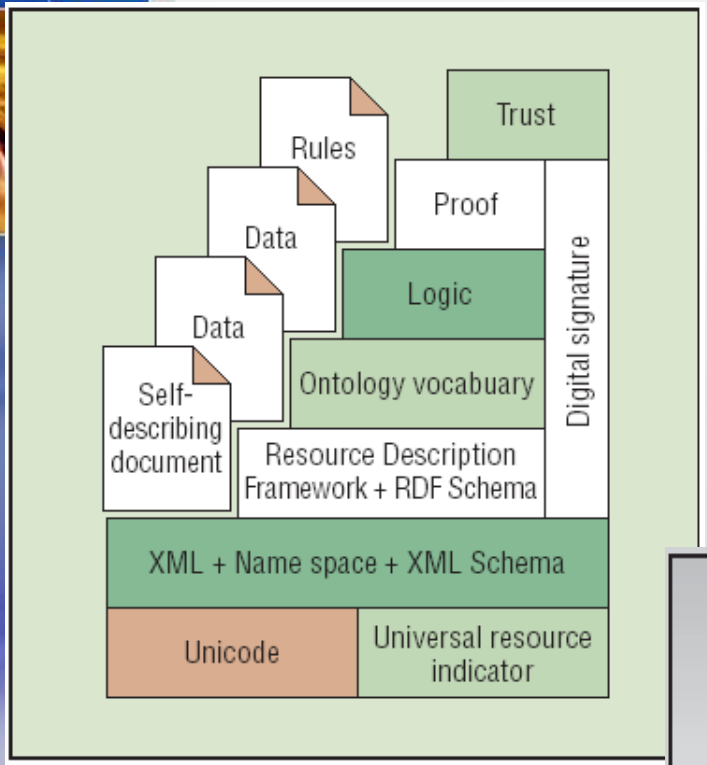
Semantic Web



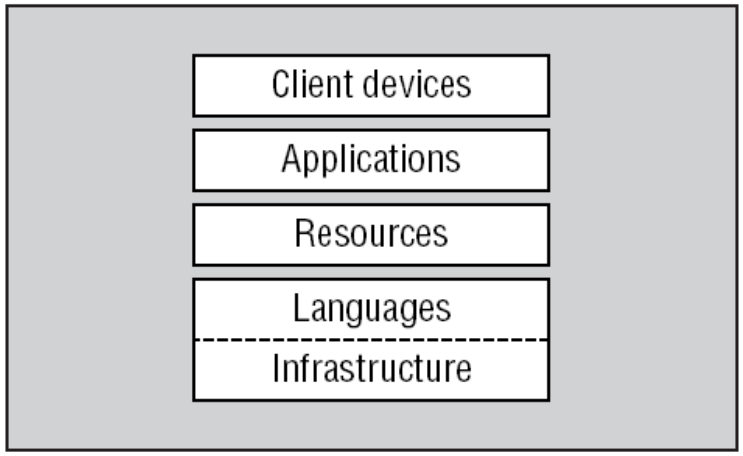
Definition of Semantic Web

- The Semantic Web is a Web that is able to describe things in a way that computers can understand
  - The Beatles was a popular band from Liverpool.
  - John Lennon was a member of the Beatles.
- The Semantic Web is not about links between web pages.
- The Semantic Web describes the **relationships between things** and the **properties of things**
- An infrastructure enables machines to COMPREHEND semantic documents and data

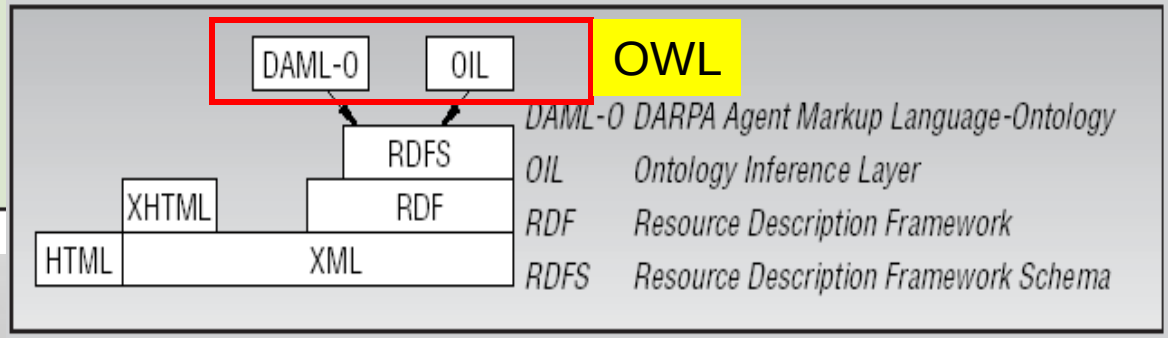
What achieves the Semantic Web?



The Semantic Web "layer cake"



A high-level layered view of the Semantic Web.



The layer language model for the Web.

LINKED DATA IN BRIEF

The Data Deluge

- Data production, consuming, and sharing
  - Amazon, Google, Genome, TheyWorkForYou
- Key questions
  - How best to provide access to data so it can be most easily reused?
  - How to enable the discovery of relevant data within the multitude of available data sets?
  - How to enable applications to integrate data from large numbers of formerly unknown data sources?

From Data Islands to a Global Data Space

- Linking data distributed across the Web requires a standard mechanism for specifying the existence and meaning of connections between items described in this data
  - RDF: a flexible way to describe things in the world – such as people, locations, or abstract concepts – and how they relate to other things
    - ✓ RDF links things, not just documents
    - ✓ RDF links are typed

Linked Data

- Just as hyperlinks in the classic Web connect documents into a single global information space, Linked Data enables links to be set between items in different data sources and therefore connect these sources into a single global data space.
  - Web of Data, Semantic Web
- The use of Web standards and a common data model make it possible to implement generic applications that operate over the complete data space. This is the essence of *Linked Data*.

Linked Data Principles

- A set of best practices for publishing and interlinking structured data on the Web
  - Sharing structured data on global scale
- Linked Data principles
  - Use URIs as names for things.
  - Use HTTP URIs, so that people can look up those names.
  - When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL).
  - Include links to other URIs, so that they can discover more things.

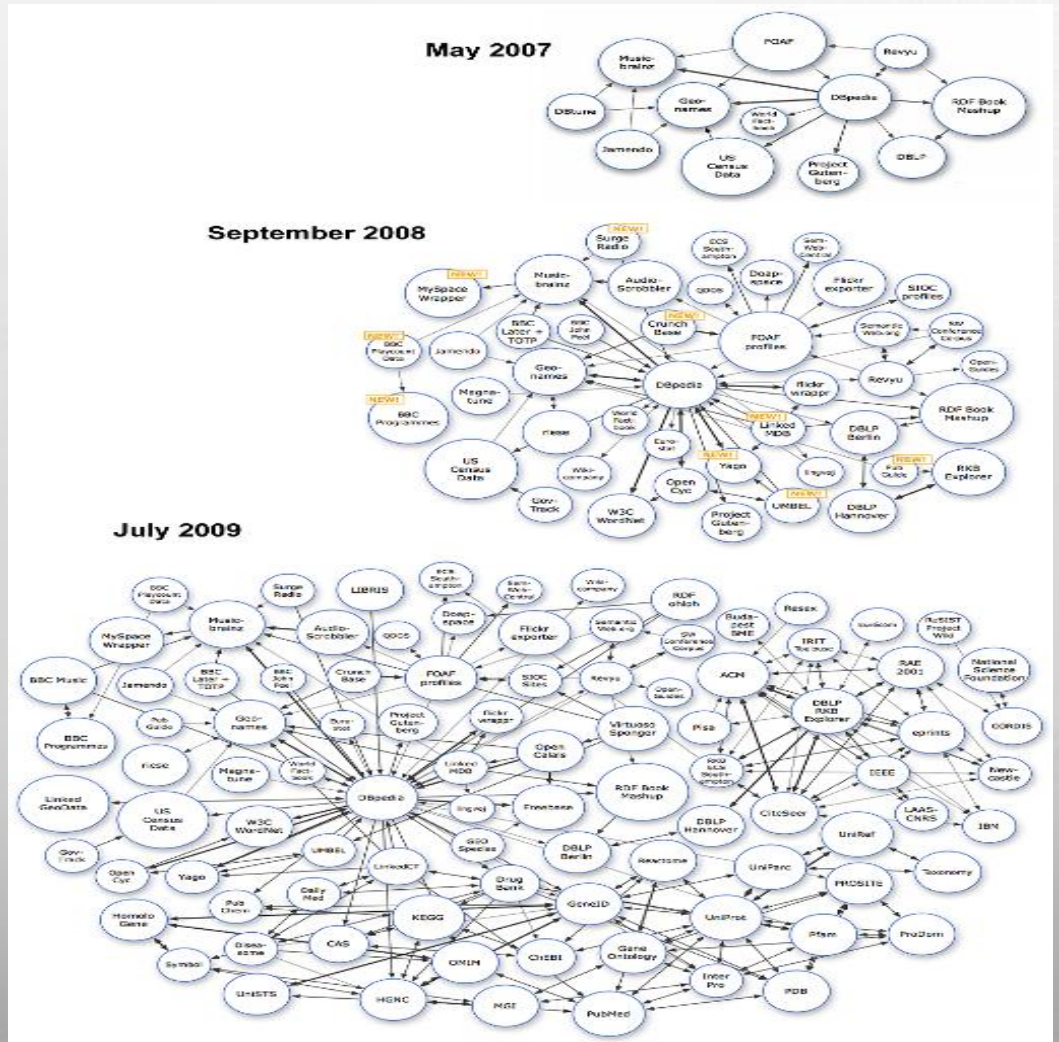
WEB OF DATA – RESULTS OF LINKED DATA

Web of Data Properties

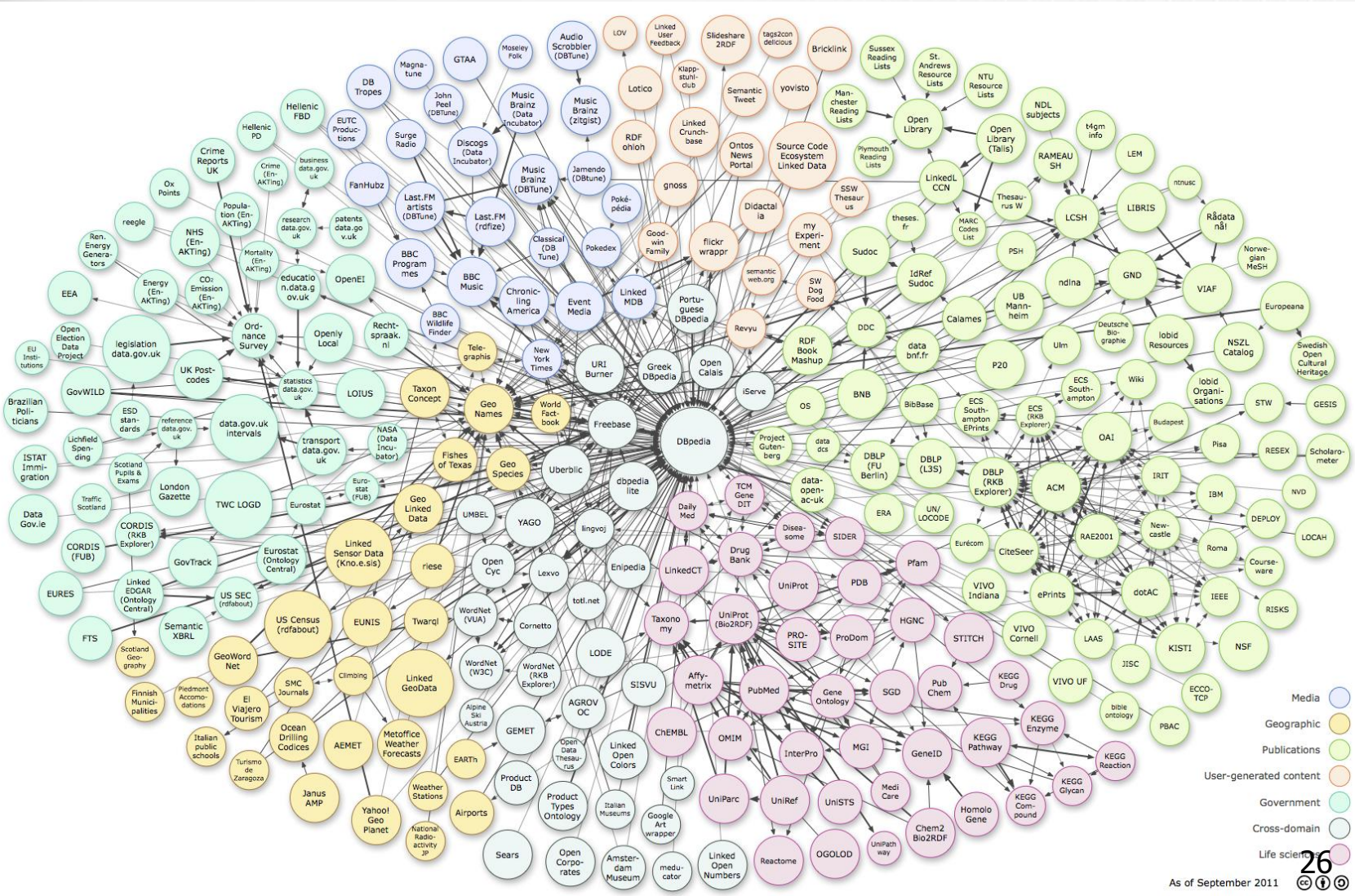
- The Web of Data is generic and can contain any type of data.
- Anyone can publish data to the Web of Data.
- The Web of Data is able to represent disagreement and contradictory information about an entity.

Bootstrapping the Web of Data

- W3C *Linking Open Data (LOD) project*



LOD Data Cloud as of September 2011



Topology of the Web of Data

Domain	Data Sets	Triples	Percent	RDF Links	Percent
Cross-domain	20	1,999,085,950	7.42	29,105,638	7.36
Geographic	16	5,904,980,833	21.93	16,589,086	4.19
Government	25	11,613,525,437	43.12	17,658,869	4.46
Media	26	2,453,898,811	9.11	50,374,304	12.74
Libraries	67	2,237,435,732	8.31	77,951,898	19.71
Life sciences	42	2,664,119,184	9.89	200,417,873	50.67
User Content	7	57,463,756	0.21	3,402,228	0.86
	203	26,930,509,703		395,499,896	

Cross-Domain Data

- *DBpedia*
  - A data set automatically extracted from publicly available Wikipedia dumps.
  - http://en.wikipedia.org/wiki/Birmingham ➔ http://dbpedia.org/resource/Birmingham
- *Freebase, UMBEL, YAGO, and OpenCyc*

從dbPedia開始

DBpedia Blog | Get Involved | Get Help

- About / News
- Applications
- Use Cases
- Datasets
- Online Access
- DBpedia Live
- Downloads
- Interlinking
- Development
- Support
- Publications
- Credits
- Contact / Imprint

DBpedia is a crowd-sourced community effort to extract structured information from Wikipedia and to make this information available on the Web. DBpedia allows you to make sophisticated queries against Wikipedia, and to link other data sets on the Web to Wikipedia data. We hope this will make it easier for the amazing amount of information in Wikipedia to be used in new and interesting ways, and that it might inspire new mechanisms for navigating, linking, and improving the encyclopedia itself.

News

DBpedia 3.8 released, including enlarged Ontology and additional localized Versions

Hi all,we are happy to announce the release of DBpedia 3.8. The most important improvements of the new release compared to DBpedia 3.7 are: 1. the new release is based on updated Wikipedia dumps dating from late May / early June 2012. 2. the DBpedia ontology is enlarged and the number of infobox to ontology mappings has [...]

DBpedia Spotlight has been selected for Google Summer of Code. Please apply now!

The Google Summer of Code (GSoC) is a global program that offers student developers (BSc,MSc,PhD) stipends to write code for open source software projects. It has had thousands of participants since the first edition in 2005, connecting prospective students with mentors from open source communities such as Debian, KDE, Gnome, Apache Software Foundation, Mozilla, etc. For [...]

DBpedia 3.7 released, including 15 localized Editions

Hi all, we are happy to announce the release of DBpedia 3.7. The new release is based on Wikipedia dumps dating from late July 2011. The new DBpedia data set describes more than 3.64 million things, of which 1.83 million are classified in a consistent ontology, including 416,000 persons, 526,000 places, 106,000 music albums, 60,000 films, 17,500 [...]

dbPedia More…

1.3. Demo Query Script for Text Search on Virtuoso

We published a simple script which was developed as a software study before the development of the Relfinder started.
We think that it will help you get familiar with SPARQL + String search on a Virtuoso server which hosts DBpedia.
The demo is deployed here and you can find the source code here.

For more interest, another demo is deployed here for getting familiar with Relfinder and how are its queries. The code for such demo exists here.

1.4. Example queries displayed with the Berlin SNORQL query explorer

- (*Complex query, might need 60sec*) All soccer players, who played as goalkeeper for a club that has a stadium with more than 40.000 seats and who are born in a country with more than 10 million inhabitants
- People who were born in Berlin before 1900
- German musicians with German and English descriptions
- German musicians who were born in Berlin
- French films
- First-person shooter computer games
- Luxury cars

OpenLibrary

Log in / Sign Up

SUBJECTS

AUTHORS ADD A BOOK

LISTS RECENTLY HELP

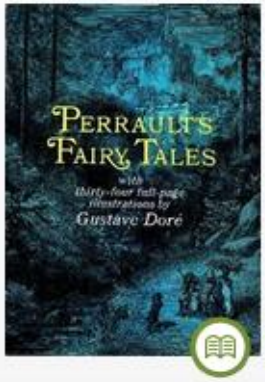
OPEN LIBRARY

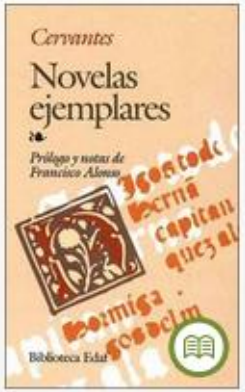
One web page for every book.

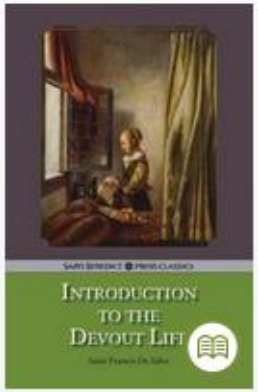
Search

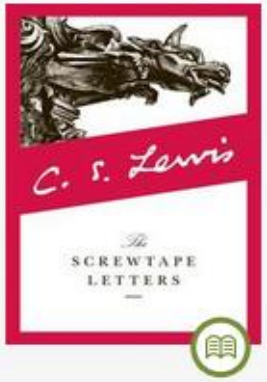
Show only eBooks More search options

Books to Read The World's classic literature at your fingertips. Over 1,000,000 free ebook titles available.

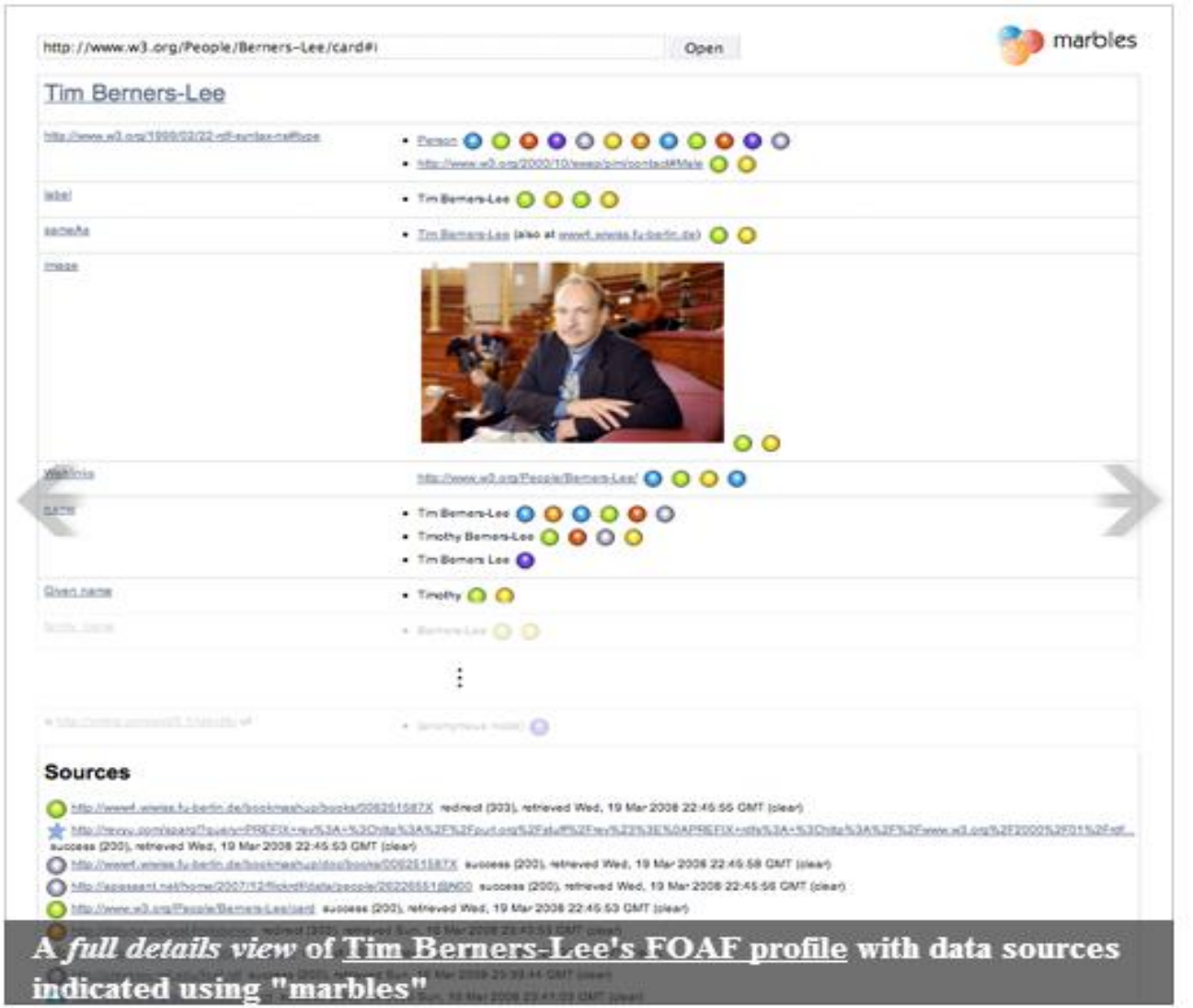








Linked Data Browser -- Marbles



A *full details view* of Tim Berners-Lee's FOAF profile with data sources indicated using "marbles"

Sig.Ma

SIG.MA SEMANTIC INFORMATION MASHUP

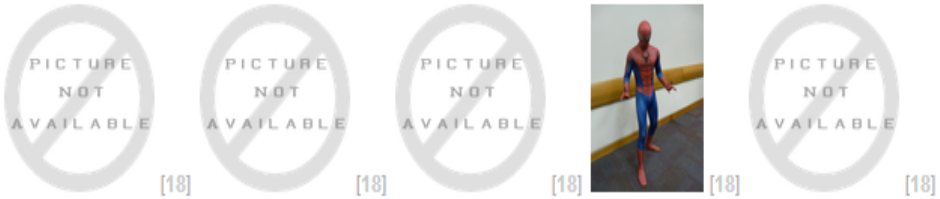
Help | About | Support

Version: 2.0.8

Spiderman | Add More Info | Start New | Order | Options | Use it

spiderman

picture:



given name: spiderman [1]

family name: spiderman [1]

comment: Spiderman 3 [2,4]

Spiderman [3,5,9]

Sosyal Medya ve Kişisel Pazarlama Üzerine [13]

Spider-Man is a fictional Marvel Comics superhero. The character was created by writer-editor Stan Lee and writer-artist Steve Ditko. He first appeared in Amazing Fantasy #15 (Aug. 1962). Lee and Ditko conceived of the character as an orphan being raised by his Aunt May and Uncle Ben, and as a teenager, having to deal with the normal struggles of adolescence in addition to those of a costumed crime fighter. Spider-Man's creators gave him super strength and agility, the ability to cling to most surfaces, shoot spider-webs using devices of his own invention which he called web-shooters, and react to danger quickly with his spider-sense, enabling him to combat his foes. When Spider-Man first appeared in the early 1960s, teenagers in superhero comic books were usually relegated to the role of sidekick to the protagonist. The Spider-Man series broke ground by featuring Peter Parker, a teenage high school student to whose self-obsessions with rejection, inadequacy, and loneliness young readers could relate. Unlike previous teen heroes such as Bucky and Robin, Spider-Man did not benefit from being the prot00E9g00E9 of any adult mentors like Captain America and Batman, and thus had to learn for himself that with great power there must also come great responsibility2014a line included in a text box in the final panel of the first Spider-Man story, but later retroactively attributed to his guardian, the late Uncle Ben. Marvel has featured Spider-Man in several comic book series, the first and longest-lasting of which is titled The Amazing Spider-Man. Over the years, the Peter Parker character has developed from shy, high school student to troubled but outgoing college student, to married high school teacher to, in

Sources (20) | Approved (0) | Rejected (0)

1 SPIDERMAN 6 facts | 2010-12-03
indice http://spidermanquat.blogspot.com/2008/12...

2 Spiderman 3 11 facts | 2010-07-15
indice http://it.bestshopping.com/prezzi/Spiderm...

3 Spiderman 11 facts | 2010-07-15
indice http://it.bestshopping.com/prezzi/Spiderm...

4 Spiderman 3 11 facts | 2010-07-15
indice http://it.bestshopping.com/prezzi/Spiderm...

5 Spiderman 11 facts | 2010-07-14
indice http://it.bestshopping.com/prezzi/Spiderm...

6 Untitled document 2 facts | 2013-03-20
OPENLINK http://sw.opencyc.org/2008/06/10/concept/...

7 Spiderman 5 facts | 2013-03-20
OPENLINK http://rdf.basekb.com/ns/m.0ykbxj

8 Spiderman 6 facts | 2013-03-20
OPENLINK http://rdf.basekb.com/ns/m.0zv9f5

9 Spiderman 11 facts | 2010-07-15
indice http://it.bestshopping.com/prezzi/Spiderm...

10 Spiderman 32 facts | 2009-09-13
indice http://www.mloknitting.com/?tag=spiderman

11 spiderman 7 facts | 2012-10-10
indice http://www.hisboyscanswim.com/tag/spiderm...

<- 1 2 -> | reject all | approve all

Geographic Data

- *Geonames*
  - An open-license geographical database that publishes Linked Data about 8 million locations
- *LinkedGeoData*
  - A Linked Data conversion of data from the *OpenStreetMap* project
  - Interlinking between Geonames, LinkedGeoData, and DBpedia
- Linked data versions of EuroStat, World Factbook, US Census data sets

LinkedGeoData Browser

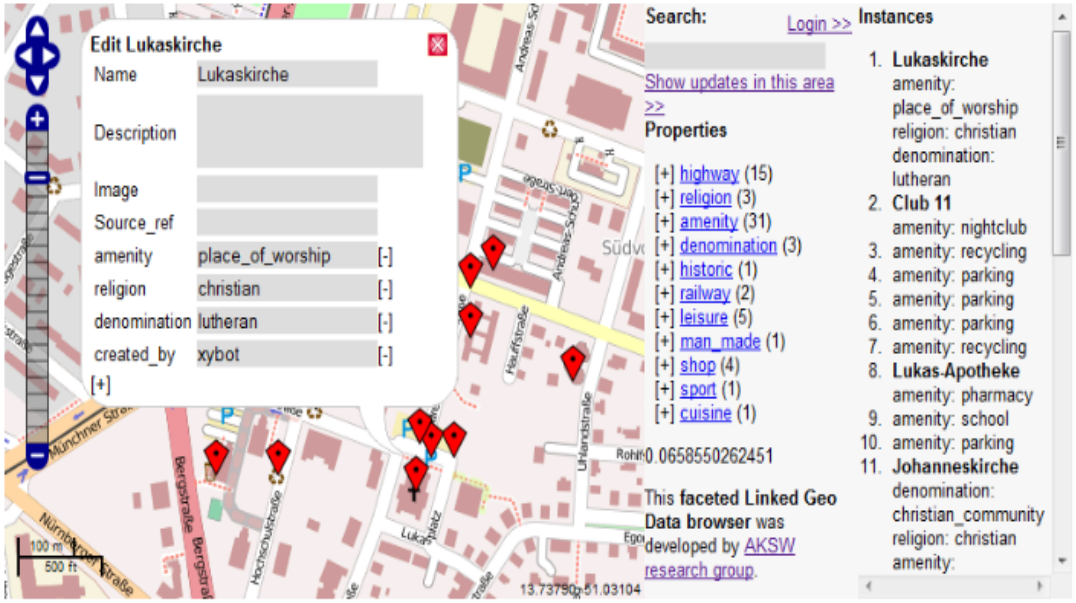
LinkedGeoData.org

Adding a spatial dimension to the Web of Data.

Project created by: AKSW

Linked Geo Data Browser and Editor

The LGD Browser and Editor (available at http://browser.linkedgeodata.org) allows to browse the world by using a slippy map. Once a region is selected, the browser analyzes the descriptions of nodes and ways in that region and generates facets for filtering. Once a facet or a specific facet value has been selected, matching elements are displayed as markers on the map and in a list. If the selected region is changed, these are updated accordingly.



If a user logs into the application by using her OSM credentials, the displayed elements can directly be edited in the map view. For this, the browser generates a dynamic form based on existing properties. The form also allows to add arbitrary additional properties. In oder

Media Data

- British Broadcasting Corporation (BBC)
  - http://www.bbc.co.uk/programmes
  - http://www.bbc.co.uk/music
- New York Times
  - http://data.nytimes.com/
- Thomas Reuters: *Calais*
  - http://www.opencalais.com/

Libraries and Education

- American LC (Subject Heading)
  - http://id.loc.gov/authorities/subjects/sh85042531
- OpenLibrary
  - http://openlibrary.org/
- DBLP
  - http://www.informatik.uni-trier.de/~ley/db/
- Talis Aspire
  - http://www.talisaspire.com/

RESOURCE DESCRIPTION FRAMEWORK (RDF)

What is RDF?

- Resource Description Framework
  - RDF is a language for representing information about **Web** resources (Ah…metadata)
  - RDF can be used to represent information about things that can be identified on the Web, even when they cannot be directly retrieved on the Web
  - Intended for situations in which information needs to be processed by applications, rather than being only displayed to people. (Ah… Semantic Web…)

RDF Data Model

- A model for describing
  - resources
  - named properties
  - property values
- (resource, property, value)

Resource

- Any object that is uniquely identifiable by an Uniform Resource Identifier (URI) plus optional anchor ids (URIrefs)
  - An entire Web page
  - Part of a Web page
  - **An object not directly accessible via the Web (e.g. a printed book, person, car…)**

Property and Values

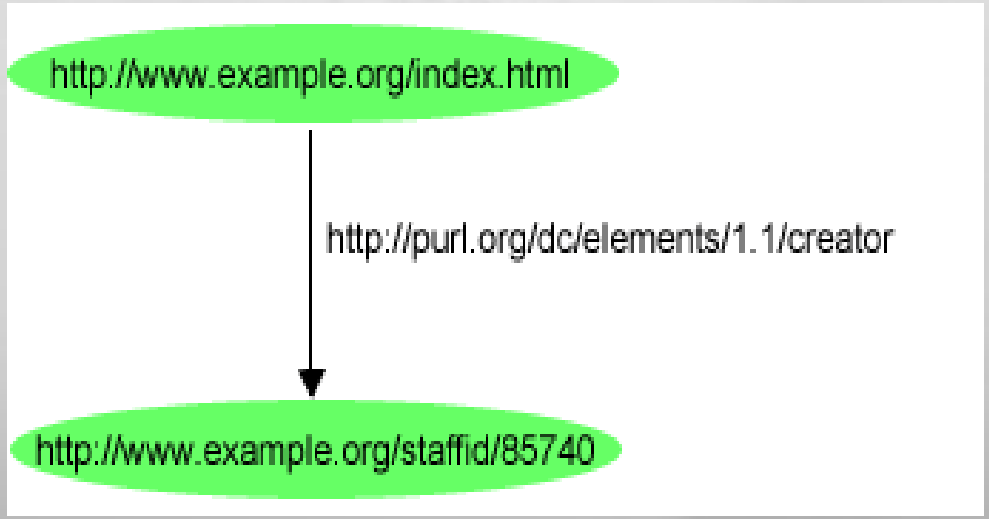
- A specific aspect, characteristics, attribute, or relation used to describe a resource
  - Specific meaning
  - Permitted values
  - Types of resources it can describe
  - Relationship with other properties
- Properties and values are identified by URIrefs
  - ***WHY???***

Statement

- A specific resource together with a named property plus the value of that property for that resource
  - Subject – Resource
  - Predicate – Property
  - Object – Property value
    - ✓ Literal or another resource
- Description: A collection of the properties that refers to the same resource

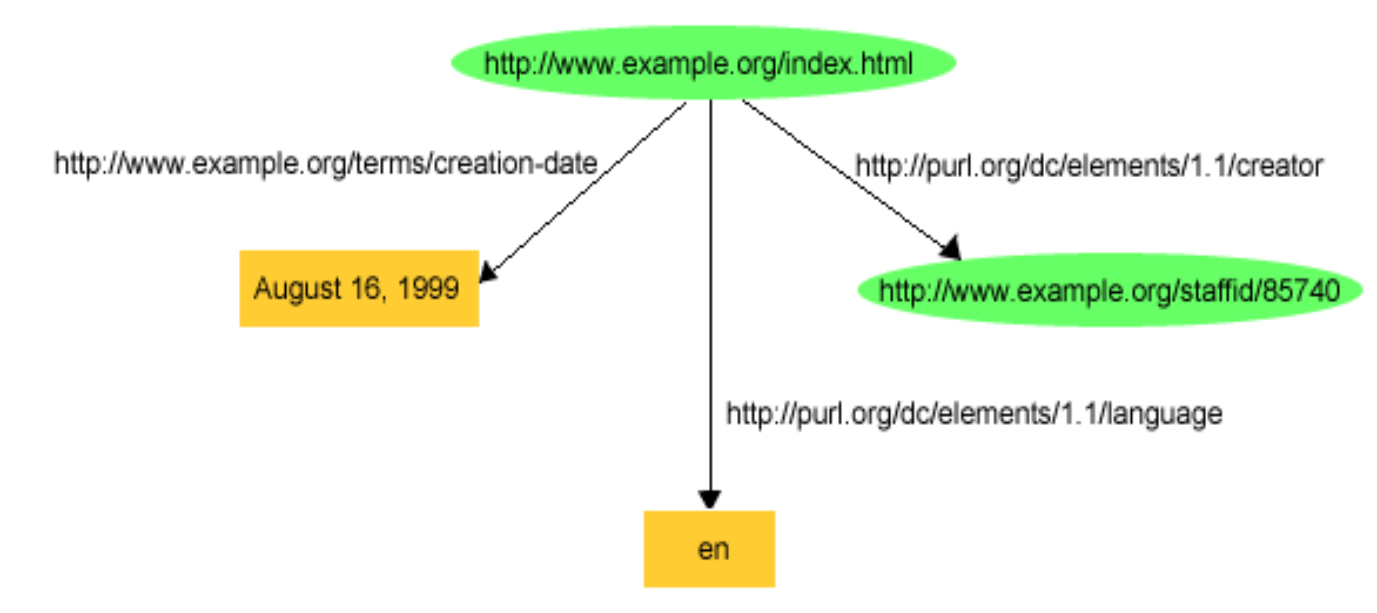
Example 1

- http://www.example.org/index.html has a creator whose value is John Smith
  - A subject http://www.example.org/index.html
  - A predicate http://purl.org/dc/elements/1.1/creator
  - An object http://www.example.org/staffid/85740



Example 2

- http://www.example.org/index.html has a creation-date whose value is August 16, 1999
- http://www.example.org/index.html has a language whose value is English



Objects in RDF statements may be either URIrefs, or constant values (called literals)

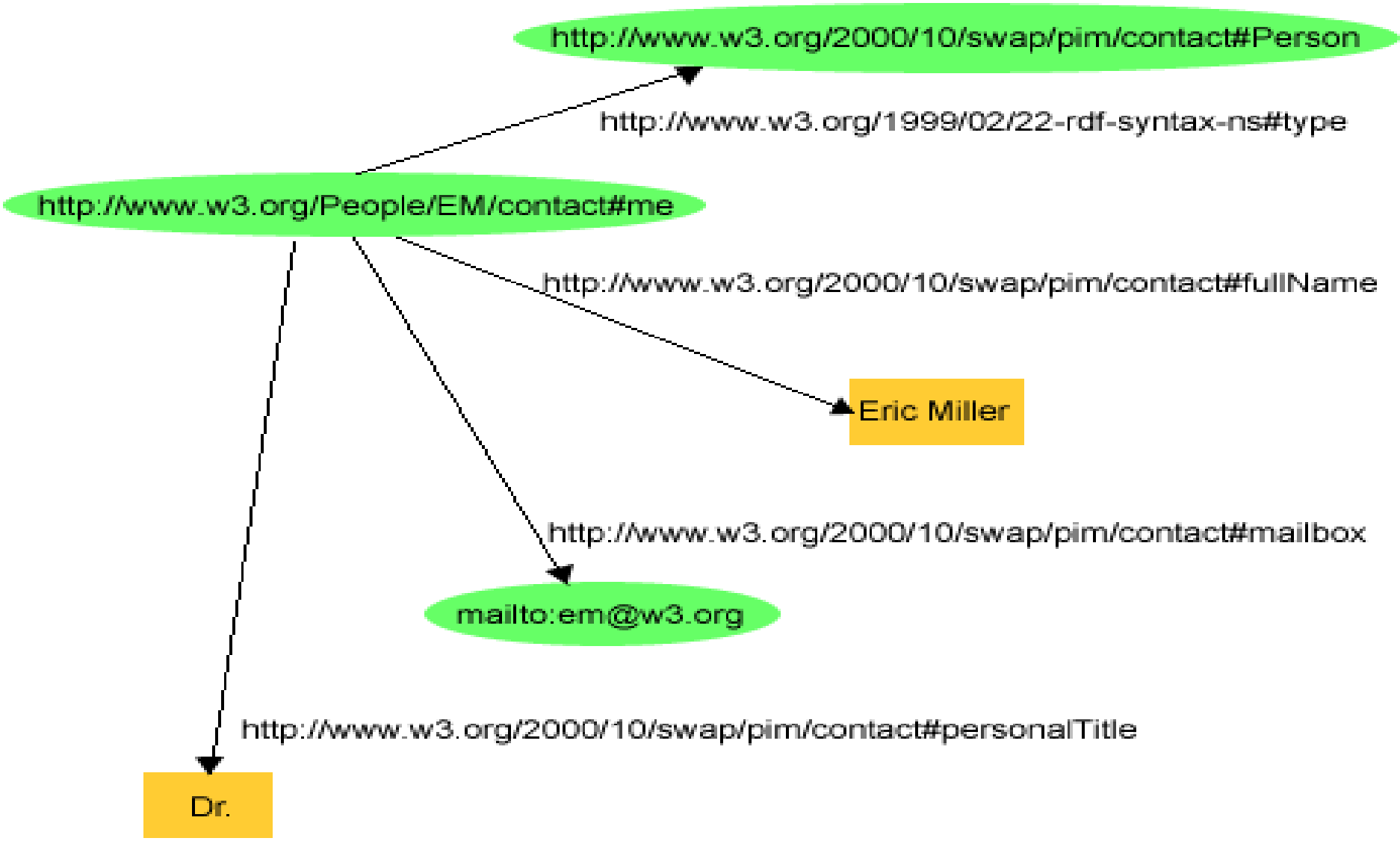
Triple Notation

- Triple Notation of Example 2
  - <http://www.example.org/index.html> <http://purl.org/dc/elements/1.1/creator> <http://www.example.org/staffid/85740> .
  - <http://www.example.org/index.html> <http://www.example.org/terms/creation-date> "August 16, 1999" .
  - <http://www.example.org/index.html> <http://purl.org/dc/elements/1.1/language> "en" .
- Shorthand Triple Notation – Giving prefix (Namespace)
  - ex:index.html dc:creator exstaff:85740 .
  - ex:index.html exterms:creation-date "August 16, 1999" .
  - ex:index.html dc:language "en" .

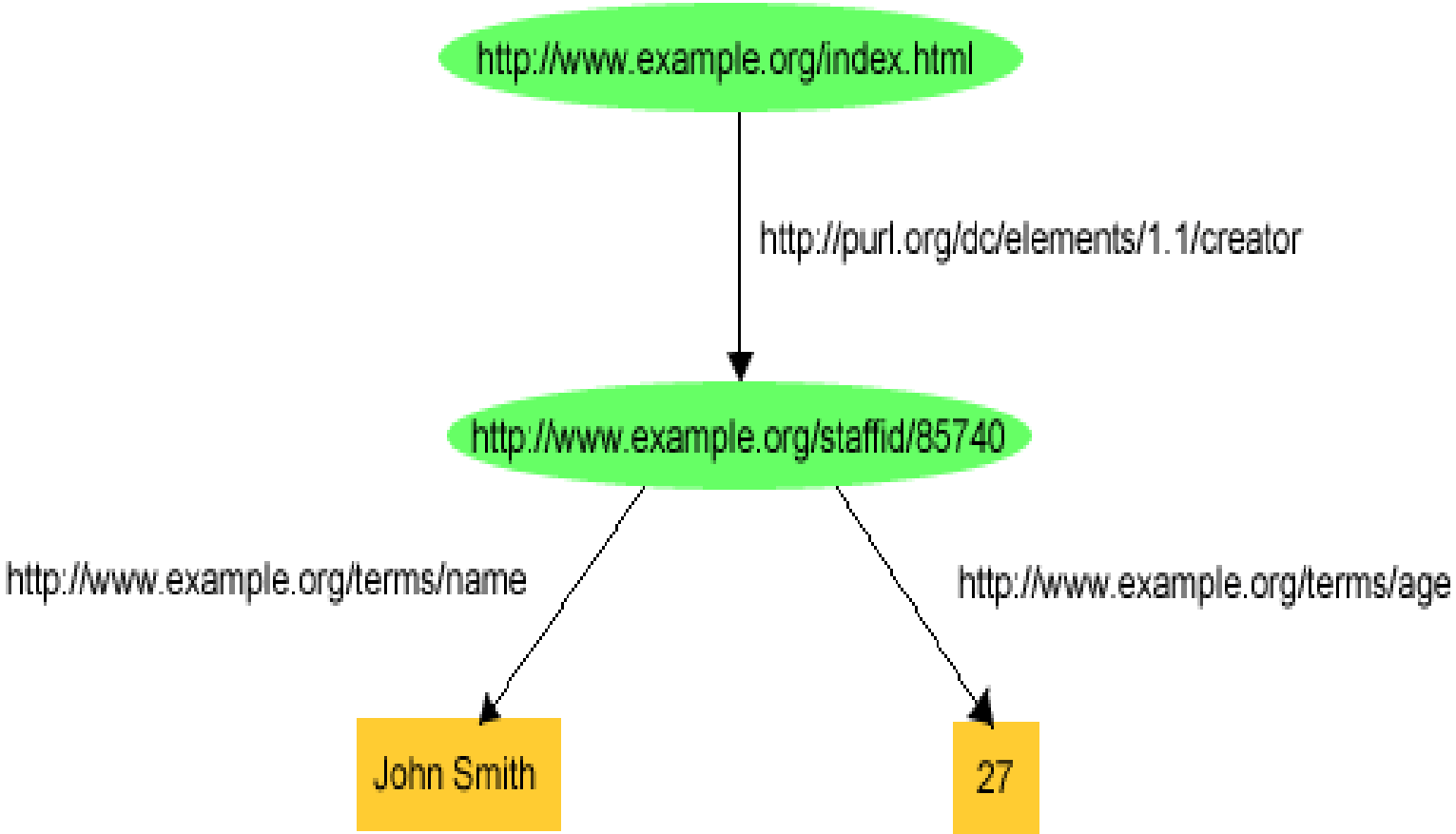
Example 3

- There is a Person identified by http://www.w3.org/People/EM/contact#me, whose name is Eric Miller, whose email address is em@w3.org, and whose title is Dr.

Example 3 (Cont.)



Example 4


http://www.example.org/index.html
http://purl.org/dc/elements/1.1/creator
http://www.example.org/staffid/85740
http://www.example.org/terms/name
http://www.example.org/terms/age
John Smith
27

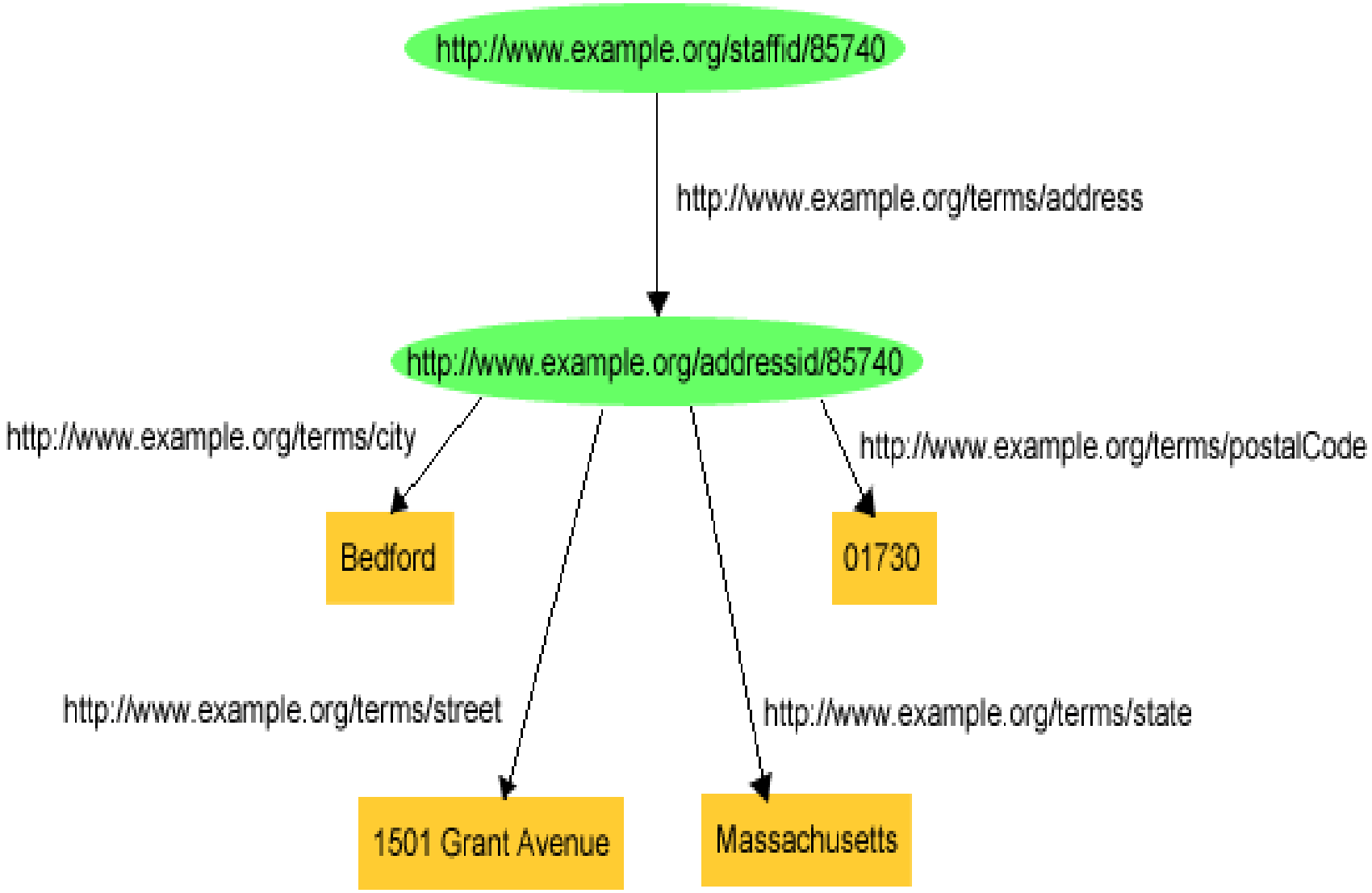
Why Use URIrefs?

- Unambiguous
- Using URIrefs as subjects, predicates, and objects in RDF statements supports the development and use of shared vocabularies on the Web, since people can discover and begin using vocabularies already used by others to describe things, reflecting a shared understanding of those concepts

Structured Property Value

- Structure property
  - Name: first name, middle name, last name
  - Address: street, city, state, postal code
- Structured information is represented in RDF by considering the aggregate thing to be described as a resource, and then making statements about that new resource
- This way of representing structured information in RDF can involve generating numerous "intermediate" URIrefs to represent aggregate concepts such as John's address.
  - Such concepts may never need to be referred to directly from outside a particular graph, and hence may not require "universal" identifiers

Breaking Up John's Address



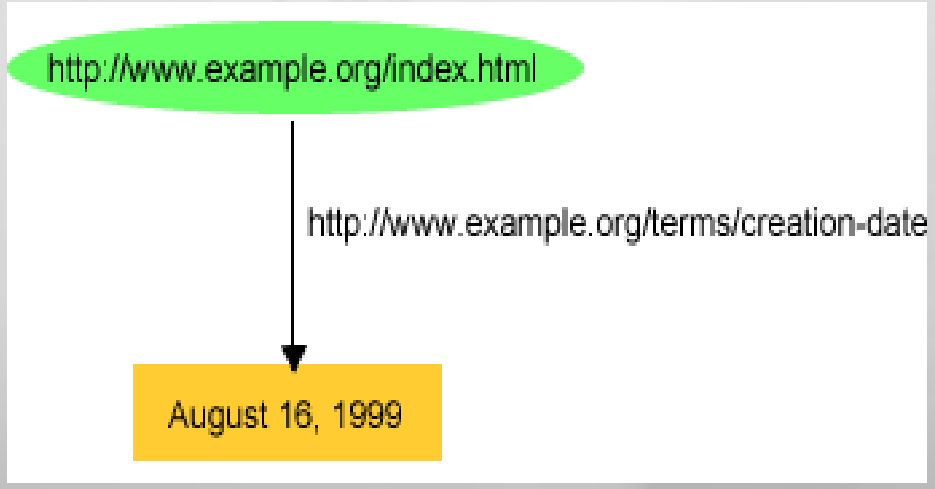
An XML Syntax for RDF: RDF/XML

- The RDF data model provides an abstract, conceptual framework for defining and using resource.
- A concrete syntax is also needed for the purposes of creating and exchanging this data model.
  - XML
  - Alternate syntaxes, like S-expression, are also possible

Describing a Web Page

http://www.example.org/index.html has a **creation-date** whose value is **August 16, 1999**

ex:index.html exterms:creation-date "August 16, 1999" .



ex1.xml

Schemas and Namespaces

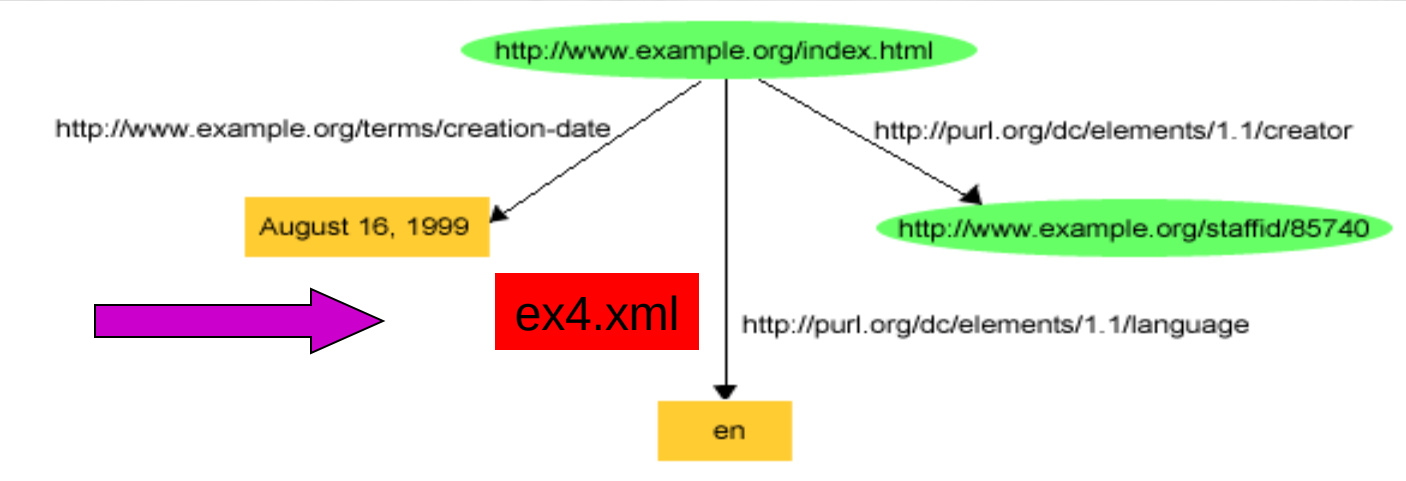
- Meaning in RDF is expressed through reference to a schema
  - Defines the terms used in RDF statements
    - ✓ Metadata element (like Dublin Core…)
- RDF uses the XML namespace facility for encoding schemas

Describing a Web Page (Cont.)

ex:index.html exterms:creation-date "August 16, 1999" .
ex:index.html dc:language "en" .

ex2.xml

ex:index.html dc:creator exstaff:85740 .
ex:index.html exterms:creation-date "August 16, 1999" .
ex:index.html dc:language "en" .



ex3.xml

ex4.xml

Basic Syntax

- Group multiple statements for the same resource into a Description element
  - about: described resource
    - ✓ about = URIref
- propertyElt: Property of a resource
  - Property names must be associated with a schema
    - ✓ By qualifying the element names with namespace prefix
- Value
  - Literal in value
  - resource in attribute
    - ✓ resource=id

Container

- Frequently, it is necessary to refer to a collection of resources
  - A work was created by more than one person
  - List the students in a course
  - List the software modules in a package
- RDF containers are used to hold such lists of resources or literals

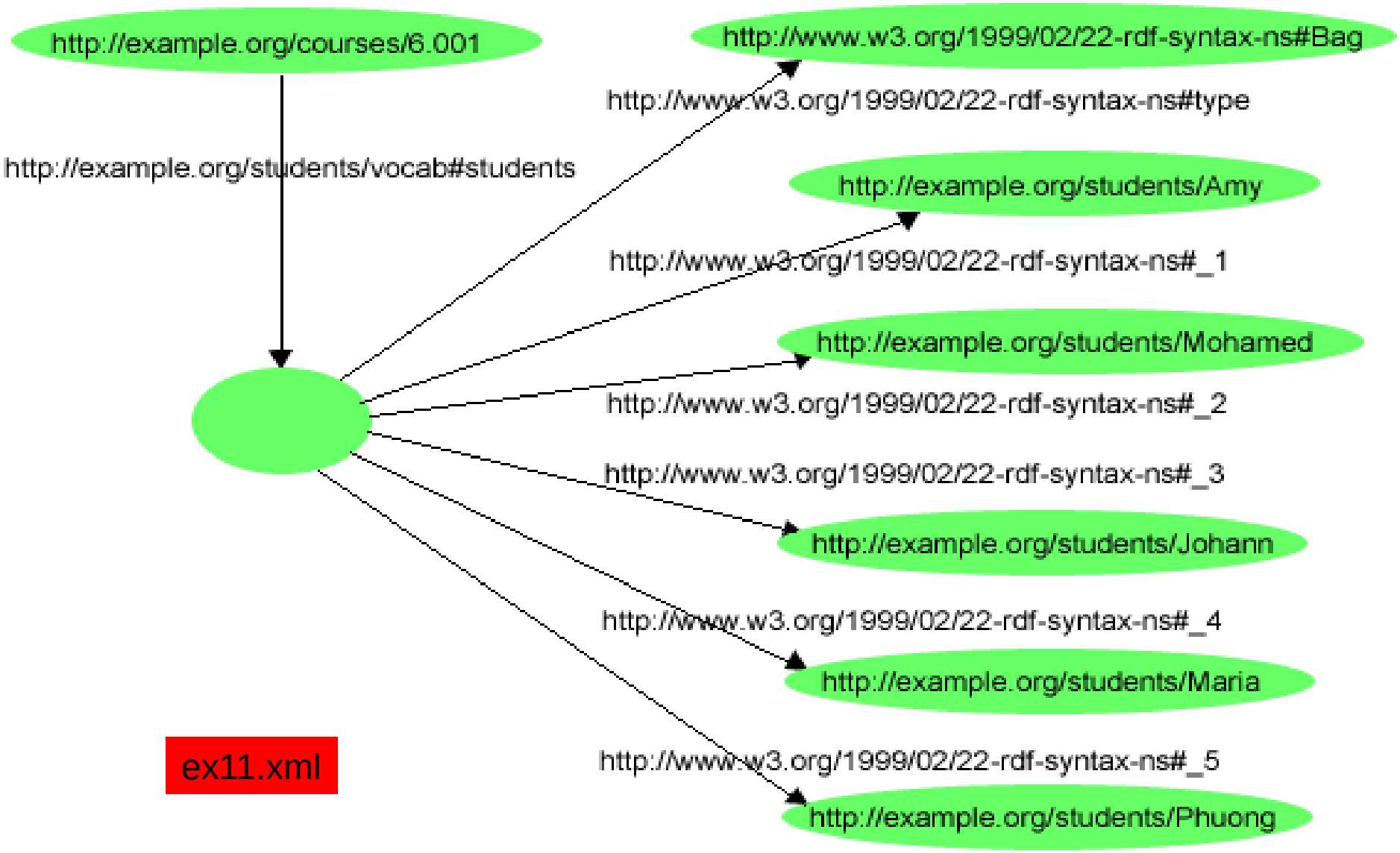
RDF Container Model (I)

- Bag
  - Unordered list of resources or literals
  - Multiple values and no significant order
- Sequence
  - Ordered list of resources or literals
  - Multiple values and with significant order
- Alternative (At least one is needed)
  - Alternatives for the (single) value of a property
  - choose any one of the items in the list as appropriate

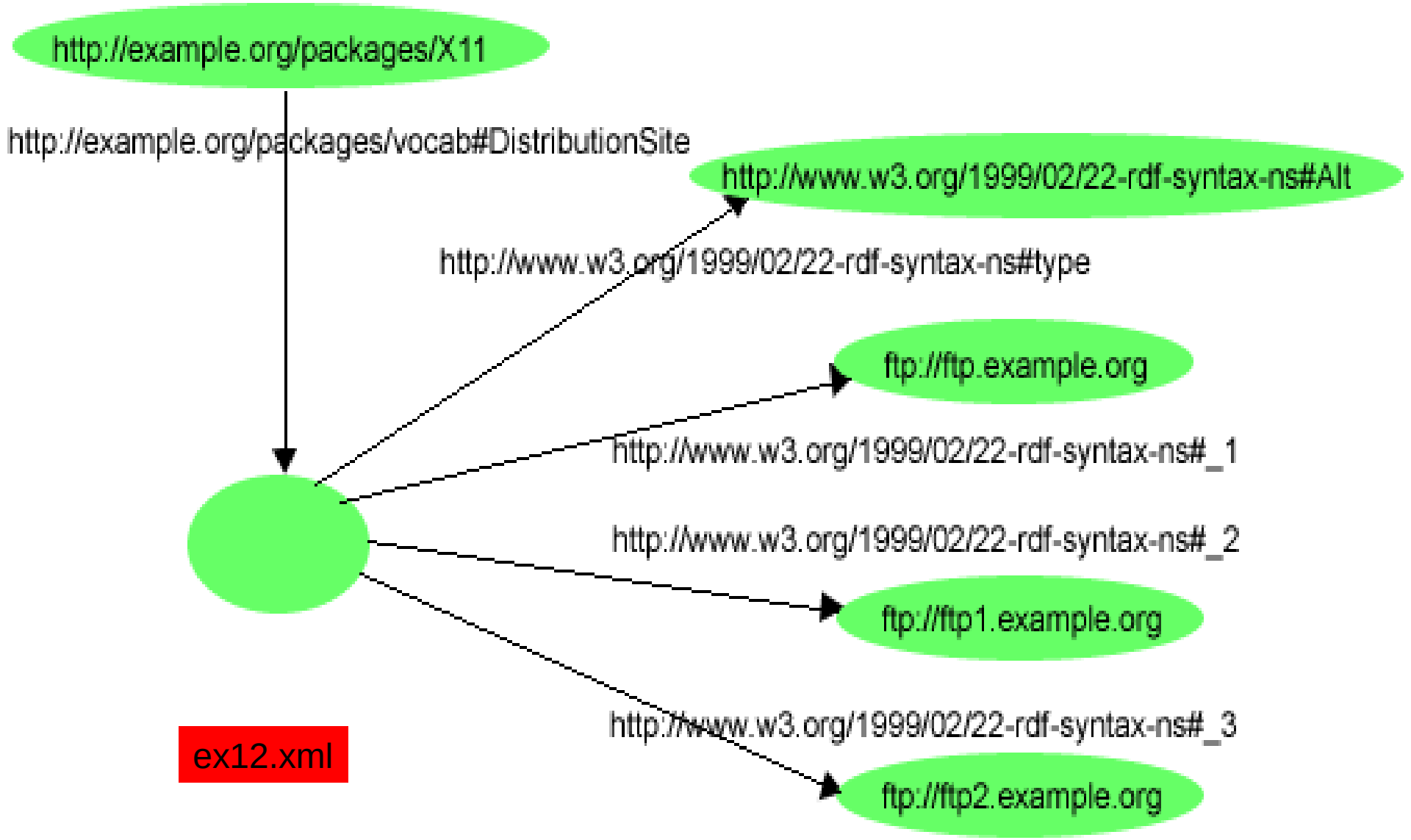
RDF Container Model (II)

- An additional resource that identifies the specific collection
  - A type property to denote the container type
  - Membership properties are named simply "_1", "_2", "_3", etc.
- A common use of containers is as the value of a property

Container Example I



Container Example II



PRINCIPLES OF LINKED DATA

http://linkeddatabook.com/editions/1.0/#

Linked Data Principles

- A set of best practices for publishing and interlinking structured data on the Web
  - Sharing structured data on global scale
- Linked Data principles
  - Use URIs as names for things.
  - Use HTTP URIs, so that people can look up those names.
  - When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL).
  - Include links to other URIs, so that they can discover more things.

Naming Things with URIs

Items, Things, Resources

- To publish data on the Web, the items in a domain of interest must first be identified
  - Things whose properties and relationships will be described in the data
    - ✓Web documents, real-world entities, abstract concepts
  - *Resource* is used to refer to these *things of interest*, which are, in turn, identified by HTTP URIs
- 1-1 principle: URI for a real-world object and URI for documents describing that object

HTTP URIs

- Provide a simple way to create globally unique names in a decentralized fashion
- Serve as a name and as a means of accessing information describing the identified entity

Making URIs De-referenceable

- De-referenceable
  - HTTP clients can look up the URI using the HTTP protocol and retrieve a description of the resource that is identified by the URI
    - ✓ Resources can be classic HTML documents, real-world objects and abstract concepts
- Descriptions of resources are embodied in…
  - HTML: read by humans
  - RDF data: read by machines
- HTTP content negotiation
  - Clients send HTTP headers with each request to indicate what kinds of documents they prefer

Two Strategies for De-referenceable

- 303 URIs (303 redirect)
  - 303 See Other
  - Two HTTPs requests to retrieve a single description of a real-world object
- Hash URIs
  - Base_part#fragment
  - Examples
    - ✓ http://biglynx.co.uk/vocab/sme/
    - ✓ http://biglynx.co.uk/vocab/sme#SmallMediumEnterprise
    - ✓ http://biglynx.co.uk/vocab/sme#Team
  - More data than necessary will be sent from the server

Hash versus 303

- 303 URIs are often used to serve resource descriptions that are part of very large data sets, such as the description of an individual concept from *DBpedia*, an RDF-ized version of Wikipedia
- Hash URIs are often used to identify terms within RDF vocabularies
  - Definitions of RDF vocabularies are usually small
  - It is often convenient for client applications to retrieve the complete vocabulary definition at once, instead of having to look up every term separately

Providing Useful RDF Information

- RDF Data Model
  - Triple: Subject, predicate, object
    - ✓ Two principal types of RDF triples
      - Literal triples
      - RDF links
  - RDF Graph
- RDF Serialization Formats
  - RDF/XML
  - RDFa: embed RDF triples in HTML documents
  - Turtle
  - N-Triples
  - RDF/JSON

Including Links to Other Things

- Set RDF links pointing into other data sources on the Web
  - Glue that connects data islands into a global, interconnected data space
  - Enable applications to discover additional data sources
- Three types of RDF links
  - Relationship Links
  - Identity Links
  - Vocabulary Links

Relationship Links

- Point at related things in other data sources
  - People, places or genes
- Relationship links enable people to point to background information about the place they live, or to bibliographic data about the publications they have written

About: 王建民

An Entity of Type : athlete, from Named Graph : http://dbpedia.org, within Data Space : dbpedia.org

王建民（1980年3月31日－），大聯盟登錄名字 Chien-Ming Wang，是台灣旅美職業棒球選手。2005年王建民成為美國職棒大聯盟紐約洋基隊的固定先發投手後，在2006年和2007年成為洋基隊最多勝投手，曾被「体育画报」評選為美國最受矚目的20位運動員之一，並在2007及2008年入選「時代 (雜誌)」全球最具影響力的百大人物之一，成為華人圈中的焦點人物，台灣報導常暱稱「建仔」、譽為「台灣之光」。目前爲華盛頓國民隊的先發投手。

Property	Value
dbpedia-owl:abstract	• Chien-Ming Wang ist ein taiwanischer Baseballspieler. Der Pitcher Wang spielte von 2005-2009 in der nordamerikanischen Major League Baseball für die New York Yankees, seit 2010 für die Washington Nationals. Außerdem ist er Nationalspieler der taiwanischen Baseball-Nationalmannschaft. Wang ist gegenwärtig einer von vier taiwanischen Baseballspielern in der MLB – neben den Los Angeles Dodgers-Spielern Hong-Chih Kuo, Chin-hui Tsao und Chin-Feng Chen. Im Jahr 2007 wurde Wang vom US-Wochenmagazin Time zu einer der 100 einflussreichsten Personen der Welt gekürt. • Chien-Ming Wang (born March 31, 1980) is a Taiwanese Major League Baseball pitcher for the Washington Nationals. He was initially signed as an amateur free agent by the New York Yankees for the 2000 season, and played for the Staten Island Yankees. He was considered the Yankees ace pitcher for the 2006 and 2007 seasons, after winning 19 games in both years, to finish among the Major League leaders in victories. • Chien-Ming Wang, né le 31 mars 1980 à Tainan, Taïwan, est un lanceur des Ligues majeures de baseball depuis 2005. Il fait partie des Nationals de Washington. • 王 建民（ワン・チェンミン、Chien-Ming Wang, 1980年3月31日 - ）は、台湾台南市出身のプロ野球選手（投手）。現在は、MLBワシントン・ナショナルズに所属している。NHKの野球放送では、日本語の音読みを当てて「おうけんみん」と呼ばれ、新聞や他局の野球放送では、「ワン・チェンミン」と呼ばれている。 • 왕치엔밍은 미국 메이저리그 프로 야구 내셔널리그 동부지구 팀인 워싱턴 내셔널스의 투수이다. 2000년 스테이튼 아일랜드 양키스와 아마추어 자유계약선수로 계약을 맺었다. 2006년과 2007년 양키스 선발투수로 활약하면서 이름을 알리게 되었다. 2009년 시즌 끝으로 방출되어 2010년 2월, 워싱턴 내셔널스로 입단하였다. • 王建民（1980年3月31日－），大聯盟登錄名字 Chien-Ming Wang，是台灣旅美職業棒球選手。2005年王建民成為美國職棒大聯盟紐約洋基隊的固定先發投手後，在2006年和2007年成為洋基隊最多勝投手，曾被「体育画报」評選為美國最受矚目的20位運動員之一，並在2007及2008年入選「時代 (雜誌)」全球最具影響力的百大人物之一，成為華人圈中的焦點人物，台灣報導常暱稱「建仔」、譽為「台灣之光」。目前爲華盛頓國民隊的先發投手。
dbpedia-owl:award	• dbpedia:World_Series • dbpedia:Win–loss_record_(pitching) • dbpedia:American_League
dbpedia-owl:battingSide	• Right
dbpedia-owl:birthDate	• 1980-03-31 (xsd:date)
dbpedia-owl:birthPlace	• dbpedia:Taiwan • dbpedia:Tainan
dbpedia-owl:formerTeam	• dbpedia:New_York_Yankees
dbpedia-owl:number	• 40 (xsd:integer)
dbpedia-owl:position	• Starting pitcher
dbpedia-owl:statisticLabel	• dbpedia:Win–loss_record_(pitching)
dbpedia-owl:statisticValue	• 59.000000 (xsd:float)
dbpedia-owl:team	• dbpedia:Washington_Nationals • dbpedia:New_York_Yankees
dbpedia-owl:throwingSide	• Right
dbpedia-owl:thumbnail	• http://upload.wikimedia.org/wikipedia/commons/thumb/e/ed/MG_3905_Chien-Ming_Wang.jpg/200px-MG_3905_Chien-Ming_Wang.jpg
dbpedia-owl:wikiPageExternalLink	• http://www.youtube.com/watch?v=A7WoaQXqyVc • http://www.franksfieldofdreams.com/feheroes/wangcm.htm
dbpprop:after	• dbpedia:Josh_Beckett
dbpprop:awards	• * Led AL in wins * 2006 Starting Pitcher of the Year Award * World Series champion
dbpprop:bats	• Right

http://dbpedia.org/page/Chien-Ming_Wang

Identity Links

- Point at URI aliases used by other data sources to identify the same real-world object or abstract concept
- Enable clients to retrieve further descriptions about an entity from other data sources
- Identity links have an important social function as they enable different views of the world to be expressed on the Web of Data
- Multiple URIs identifying the same entity
  - http://www.w3.org/2002/07/owl#sameAs

rdfs:label	▪ Wang Chien-ming ▪ Chien-Ming Wang ▪ Chien-Ming Wang ▪ 王建民 ▪ 왕치엔밍 ▪ 王建民
owl:sameAs	▪ http://ja.dbpedia.org/resource/王建民 ▪ http://fr.dbpedia.org/resource/Chien-Ming_Wang ▪ http://de.dbpedia.org/resource/Wang_Chien-ming ▪ http://ko.dbpedia.org/resource/왕치엔밍 ▪ freebase:王建民
wdrs:describedby	▪ http://dbpedia.org/page/Chien-Ming_Wang ▪ http://dbpedia.org/data/Chien-Ming_Wang.n3 ▪ http://dbpedia.org/data/Chien-Ming_Wang.nt ▪ http://dbpedia.org/data/Chien-Ming_Wang.xml ▪ http://dbpedia.org/data/Chien-Ming_Wang.json ▪ http://dbpedia.org/data/Chien-Ming_Wang.ttl ▪ http://dbpedia.org/data/Chien-Ming_Wang.jrdf ▪ http://dbpedia.org/data/Chien-Ming_Wang.jsod ▪ http://dbpedia.org/data/Chien-Ming_Wang.atom
http://www.w3.org/ns/prov#wasDerivedFrom	▪ http://en.wikipedia.org/wiki/Chien-Ming_Wang?oldid=495288273
foaf:depiction	▪ http://upload.wikimedia.org/wikipedia/commons/e/ed/MG_3905_Chien-Ming_Wang.jpg
foaf:givenName	▪ Chien-Ming
foaf:isPrimaryTopicOf	▪ http://en.wikipedia.org/wiki/Chien-Ming_Wang
foaf:name	▪ Chien-Ming Wang
foaf:surname	▪ Wang
is dbpedia-owl:wikiPageDisambiguates of	▪ dbpedia:CMW
is dbpedia-owl:wikiPageRedirects of	▪ dbpedia:Chien_Ming_Wang ▪ dbpedia:建民網 ▪ dbpedia:王建民 ▪ dbpedia:Wang_40 ▪ dbpedia:Wang_Chien-Ming ▪ dbpedia:Chin-Ming_Wang ▪ dbpedia:Chin_Ming_Wong ▪ dbpedia:Chin-Ming_Wong

Identity Links (Cont.)

- Essentiality for using identify links
  - Different opinions
  - Traceability
  - No central points of failure
- Fact vs. Claim
  - Data quality and truth

Vocabulary Links

- Point from data to the definitions of the vocabulary terms that are used to represent the data, as well as from these definitions to the definitions of related terms in other vocabularies
- Vocabulary links make data self-descriptive and enable Linked Data applications to understand and integrate data across vocabularies
- Mixture of distinct terms from different RDF vocabularies that are used by a data source to publish data on the Web

Property
dbpedia-owl:abstract
dbpedia-owl:award
dbpedia-owl:battingSide
dbpedia-owl:birthDate
dbpedia-owl:birthPlace
dbpedia-owl:formerTeam
dbpedia-owl:number
dbpedia-owl:position
dbpedia-owl:statisticLabel
dbpedia-owl:statisticValue
dbpedia-owl:team
dbpedia-owl:throwingSide
dbpedia-owl:thumbnail
dbpedia-owl:wikiPageExternalLink
dbpprop:after
dbpprop:awards
dbpprop:bats
dbpprop:before
dbpprop:birthDate
dbpprop:birthPlace
dbpprop:br
dbpprop:brm

Vocabulary Links (Cont.)

- Advocating the reuse of terms from widely deployed vocabularies
- Making data as self-descriptive as possible
  - A Linked Data application which discovers some data on the Web that is represented using a previously unknown vocabulary should be able to find all meta-information that it requires to translate the data into a representation that it understands and can process
    - ✓ Every vocabulary term links to its own definition – OR –
    - ✓ Publish mappings between terms from different vocabularies in the form of RDF links

Describing Things with RDF

- Triples that describe the resource with literals.
- Triples that describe the resource by linking to other resources (e.g., triples stating the resource's creator, or its type).
- Triples that describe the resource by linking from other resources (i.e., incoming links).
- Triples describing related resources (e.g., the name and maybe affiliation of the resource's creator).
- Triples describing the description itself (i.e., data about the data, such as its provenance, date of collection, or licensing terms).
- Triples about the broader data set of which this description is a part.

Short Summary

- Linked Data provides a more generic, more flexible publishing paradigm which makes it easier for data consumers to discover and integrate data from large numbers of data sources
  - A unifying data model
  - A standardized data access mechanism
  - Hyperlink-based data discovery
  - Self-descriptive data
- Easier for data consumers to discover, access, and integrate data

FOUR RULES, FIVE STARS, AND A PLAN

Four Rules (Principles)

- Use URIs as names for things.
- Use HTTP URIs, so that people can look up those names.
- When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL).
- Include links to other URIs, so that they can discover more things.

Five Stars


LINKED OPEN DATA
★ On the web
★★ Machine-readable data
★★★ Non-proprietary format
★★★★ RDF standards
★★★★★ Linked RDF
IS YOUR DATA 5★?

A Plan

- Define your model
  - URI, resources, relationships
- **Select** or define your metadata terms
  - Terms should be defined using an RDF-based standard (RDF or OWL)
- **Select** or define any controlled vocabularies
  - Each list of terms will be described in an RDF-compatible format (such as SKOS)
- Create links from your data (terms) to related data (terms) on the Web

Choosing and Using Vocabularies to Describe Data

- RDF provides a generic, abstract data model for describing resources using subject, predicate, object triples. However, it does not provide any domain-specific terms for describing classes of things in the world and how they relate to each other.
- Taxonomies, Vocabularies
- Ontologies expressed in SKOS (Simple Knowledge Organization System), RDFS (the RDF Vocabulary Description Language, also known as RDF Schema) and OWL (the Web Ontology Language).

SKOS, RDFS and OWL

- SKOS is a vocabulary for expressing conceptual hierarchies, often referred to as taxonomies, while RDFS and OWL provide vocabularies for describing conceptual models in terms of classes and their properties.
- SKOS, RDFS and OWL provide a continuum of expressivity.
  - SKOS is widely used to represent thesauri, taxonomies, subject heading systems, and topical hierarchies (for instance that mechanics belong to the boarder topic of physics).
  - RDFS and OWL are used in cases where subsumption relationships between terms should be represented (for instance that all athletes are also persons).

Reusing Existing Terms

- The Dublin Core Metadata Initiative (DCMI) Metadata Terms vocabulary
  - defines general metadata attributes such as title, creator, date and subject.
- The Friend-of-a-Friend (FOAF) vocabulary
  - defines terms for describing persons, their activities and their relations to other people and objects.
- The Semantically-Interlinked Online Communities (SIOC)
  - designed for describing aspects of online community sites, such as users, posts and forums.
- The Description of a Project (DOAP) vocabulary
  - defines terms for describing software projects, particularly those that are Open Source.
- The Music Ontology
  - defines terms for describing various aspects related to music, such as artists, albums, tracks, performances and arrangements.
- The Programmes Ontology
  - defines terms for describing programmes such as TV and radio broadcasts.

Reusing Existing Terms

- The Good Relations Ontology
  - defines terms for describing products, services and other aspects relevant to e-commerce applications.
- The Creative Commons (CC) schema
  - defines terms for describing copyright licenses in RDF.
- The Bibliographic Ontology (BIBO)
  - provides concepts and properties for describing citations and bibliographic references (i.e., quotes, books, articles, etc.).
- The OAI Object Reuse and Exchange vocabulary
  - is used by various library and publication data sources to represent resource aggregations such as different editions of a document or its internal structure.
- The Review Vocabulary
  - provides a vocabulary for representing reviews and ratings, as are often applied to products and services.
- The Basic Geo (WGS84) vocabulary
  - defines terms such as lat and long for describing geographically-located things.

Swoogle

- Creator: eBiquity (University of Maryland, Baltimore County)
- Three search modes:
  - **ontology** searches the full text of ontology documents
  - **data** searches actual instance data
  - **term** searches only terms that have been defined as classes or properties

vocab.org

- Creator: Ian Davis
- The terms at vocab.org do not serve a particular application. For the most part, they are generalized vocabularies, each covering a narrow area.
- The vocab.org version of FRBR is currently the most used vocabulary for expressing FRBR in linked data, in part because it was the first expression of FRBR in RDF, but also because it is a relatively simple and therefore easy to understand implementation.

Open Metadata Registry

- Creators: Diane Hillmann, Jon Phipps
- A number of library-related ontologies (and vocabularies) can be found in the Open Metadata Registry (OMR).
- It is currently being used by the Joint Steering Committee for Development of RDA as well as IFLA.
- With all of the terms used in RDA, ISBD, and members of the FR family (FRBR, FRAD, and FRSAD), there is a wealth of terms to reuse in bibliographic metadata.

Subject List and Thesauri

- Non-library and General
  - DBpedia Ontology
  - Freebase Types
  - New York Times Subjects
  - Artificial Intelligence: OpenCyc and UMBEL
    - ✓ OpenCyc
    - ✓ UMBEL
- Library-specific Subject Heading
  - Library of Congress Subject Headings (LCSH)
  - National Agriculture Library (NAL)
  - Thesaurus for Graphic Materials (TGM)

Subject List and Thesauri (Cont.)

- Library Classification Schemes
  - Dewey Summaries
  - Universal Decimal Classification (UDC)
- Non-library-based List
  - Geoname
  - BBC Wildlife Ontology
  - Dublin Core Type Vocabulary
- Library-based Lists
  - MARC Countries
  - MARC Geographic Areas
  - MARC Languages
  - MARC Relators
  - RDA Vocabularies
- Name Authorities
  - Virtual International Authority File (VIAF)
  - New York Times People and Organizations

Subject List and Thesauri (Cont.)

- Preservation
  - LC Preservation Level Role and Preservation Events
    - ✓ Library of Congress Preservation Events
    - ✓ Library of Congress Preservation Level Role

Defining Terms

- In cases where existing vocabularies are not adequate to describe a particular data set, new terms will need to be developed in a dedicated vocabulary.
  - Supplement existing vocabularies rather than reinventing their terms.
  - Only define new terms in a namespace that you control.
  - Use terms from RDFS and OWL to relate new terms to those in existing vocabularies.
  - Apply the Linked Data principles equally rigorously to vocabularies as to data sets – URIs of terms should be dereferenceable.
  - Document each new term with human-friendly labels and comments – use rdfs:label and rdfs:comment.
  - Only define things that matter.

Find More Linked Data

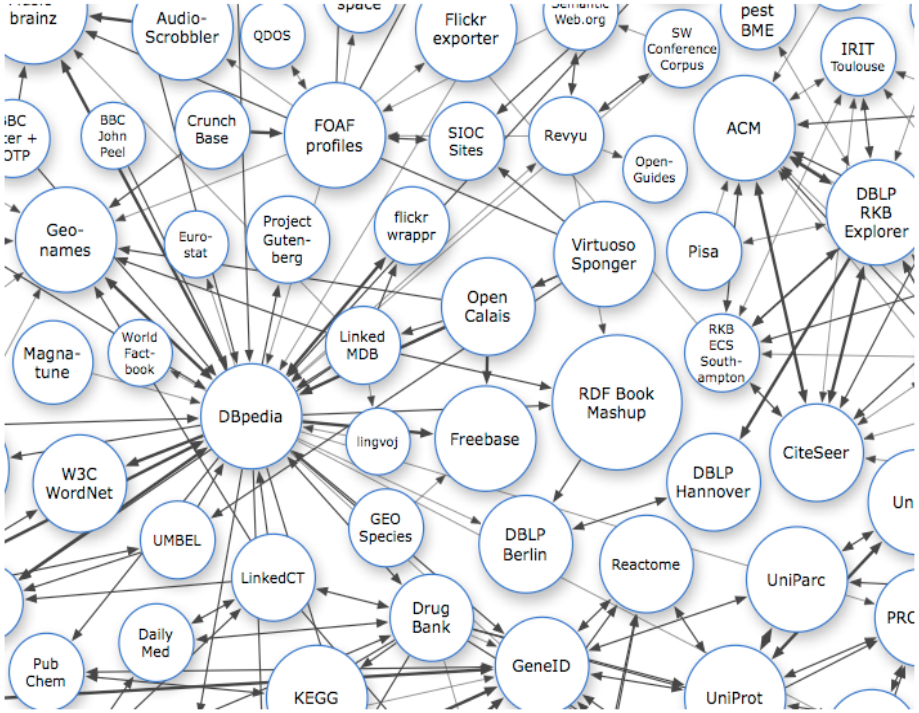
Linked Data - Connect Distributed Data across the Web

- Home
- Guides and Tutorials
- Frequently Asked Questions
- Glossary
- Images and Posters
- Presentations
- Data Sets
- Tools
- Events
- Calls for Papers
- Research
- News and Blogs
- Domains
- See Also

Linked Data

Linked Data is about using the Web to connect related data that wasn't previously linked, or using the Web to lower the barriers to linking data currently linked using other methods. More specifically, Wikipedia defines Linked Data as "a term used to describe a recommended best practice for exposing, sharing, and connecting pieces of data, information, and knowledge on the Semantic Web using URIs and RDF."

This site exists to provide a home for, or pointers to, resources from across the Linked Data community.



Login to edit (authorised users only at present)

Username: *

Password: *

Log in

- Log in using OpenID
- Request new password

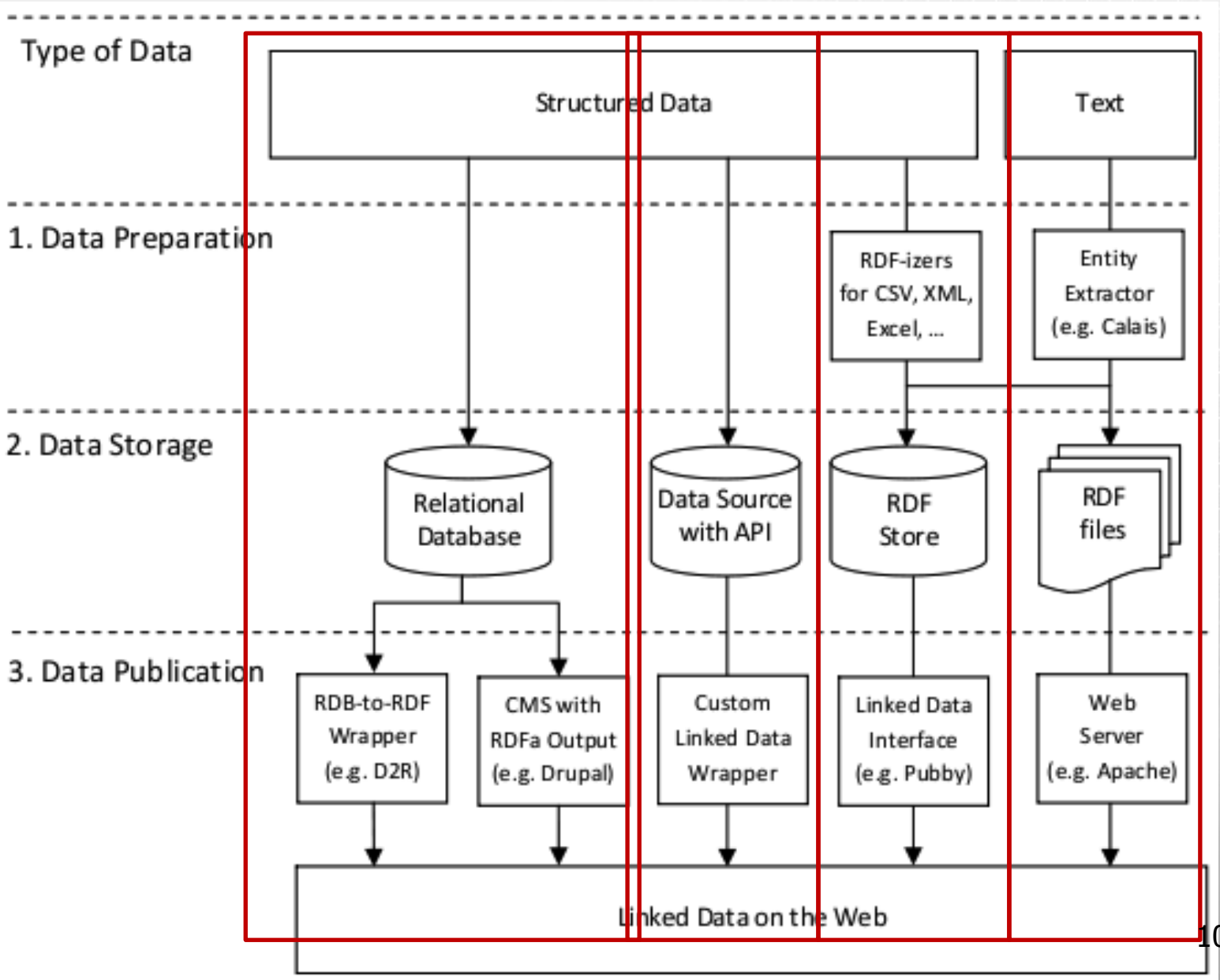
Navigation

- Feed aggregator

Syndicate

IMPLEMENTATION FRAMEWORK

Patterns in a Nutshell

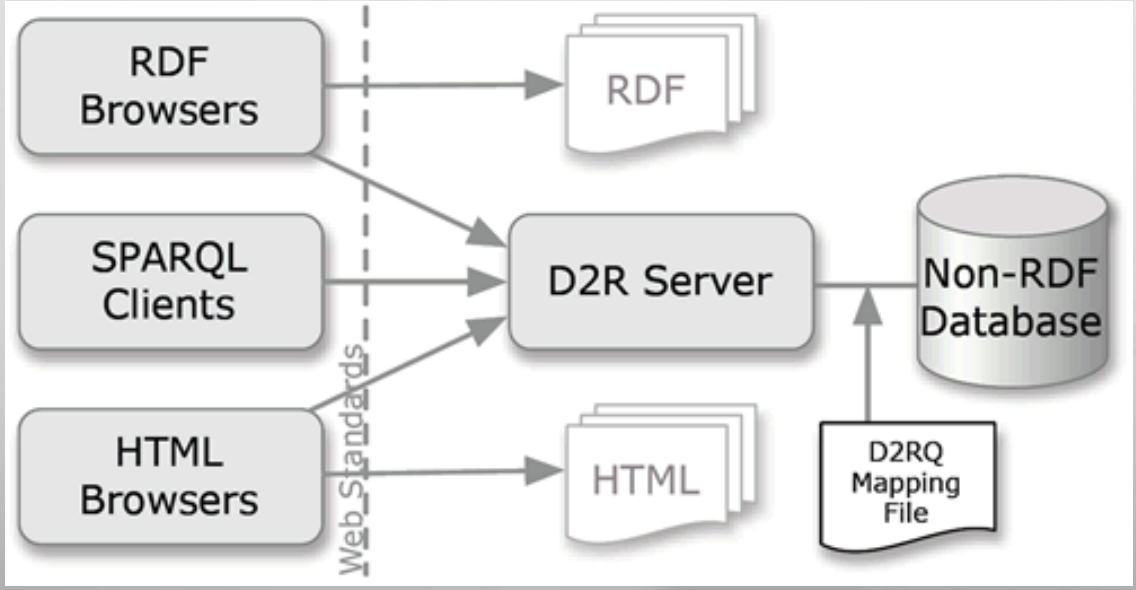


Additional Consideration

- Data Volume: How much data needs to be served?
  - Larger data sets that describe multiple entities should be split into separate files.
- Data Dynamism: How often does the data change?
  - If your data changes frequently it will be preferable to use a storage and management mechanism (such as an RDF store) that enables frequent changes.

Serving Linked Data from RDB

- One widely used tool designed for this purpose is D2R Server
- D2R Server relies on a declarative mapping between the database schema and the target RDF terms, provided by the data publisher



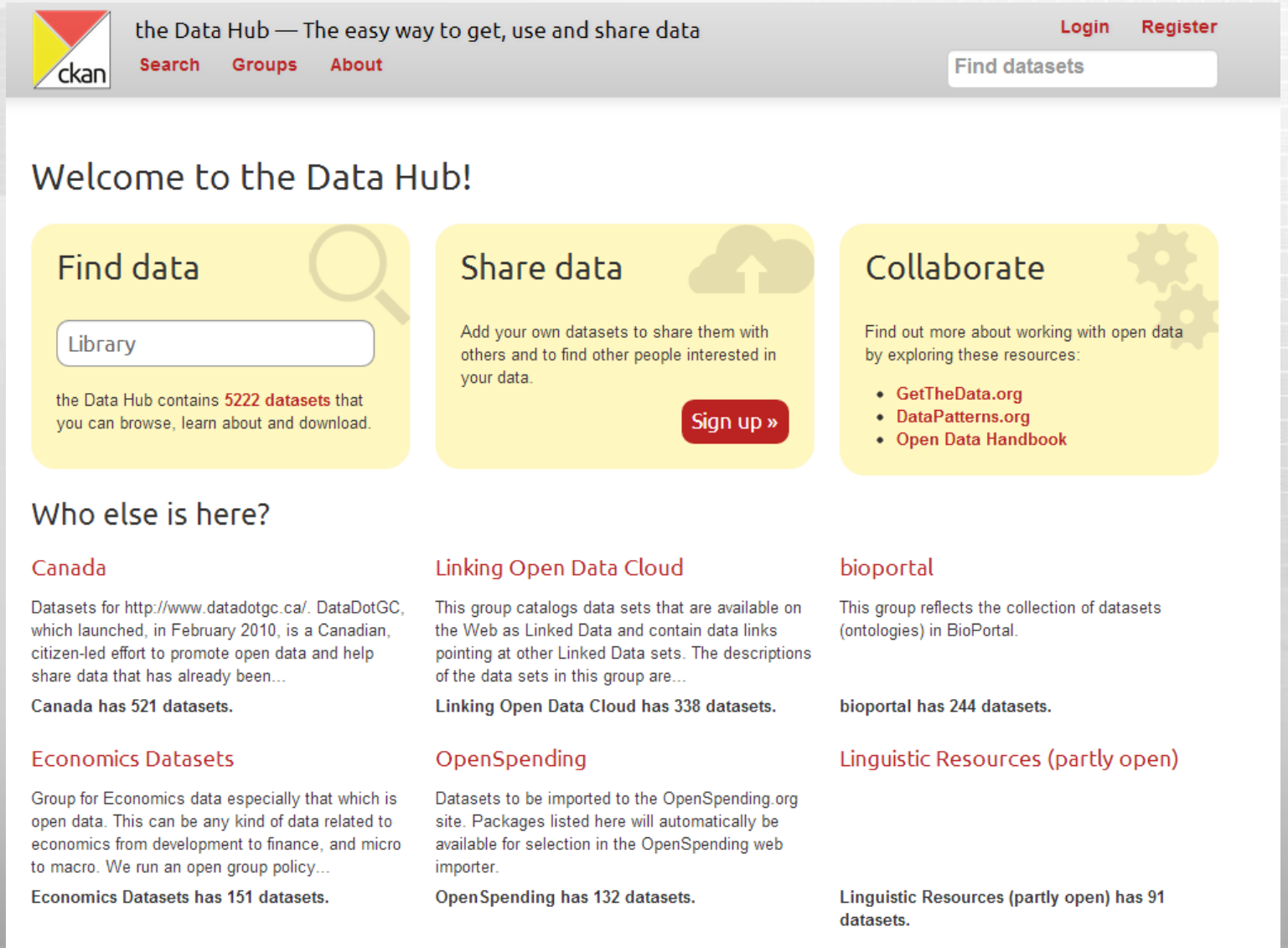
Serving Linked Data from RDB (Cont.)

- Using D2R Server to publish a relational database as Linked Data typically involves the following steps:
  - Download and install the server software as described in the Quick Start section of the D2R Server homepage.
  - Have D2R Server auto-generate a D2RQ mapping from the schema of your database
  - Customize the mapping by replacing auto-generated terms with terms from well-known and publicly accessible RDF vocabularies

Serving Linked Data from RDB (Cont.)

- Using D2R Server to publish a relational database as Linked Data typically involves the following steps:
  - Set RDF links pointing at external data sources
  - Set several RDF links from an existing interlinked data source (ex FOAF profile) to resources within the new data set, to ensure crawlers can discover the data.
  - Add the new data source to the CKAN registry in the group LOD Cloud

CKAN Registry



Serving Linked Data by Wrapping Existing Applications or Web API

- In general, Linked Data wrappers do the following:
  - They assign HTTP URIs to the resources about which the API provides data.
  - When one of these URIs is dereferenced asking for application/rdf+xml, the wrapper rewrites the client's request into a request against the underlying API.
  - The results of the API request are transformed to RDF and sent back to the client.

Linked Data Publishing CheckList

- Does your data set links to other data sets?
- Do you provide provenance metadata?
- Do you provide licensing metadata?
- Do you use terms from widely deployed vocabularies?
- Are the URIs of proprietary vocabulary terms dereferenceable?
- Do you map proprietary vocabulary terms to other vocabularies?
- Do you provide data set-level metadata?
- Do you refer to additional access methods (SPARQL and RDF dump) ?

鏈結(開放)資料

- 鏈結(開放)資料 = Linked (Open) Data, L(O)D
- Linked Data乃是一組最佳實務，用以將結構化資料發布在Web上，並加以串聯
- Linked Data 是實現語意網(Semantic Web)的具體方案
- 美國國會圖書館的一段Linked Open Data短片…呃，希望上完課後你全懂了
- 從語意網談起…