



Linked (Open) Data

Fabrizio Celli – 04/06/2012

OUTLINE

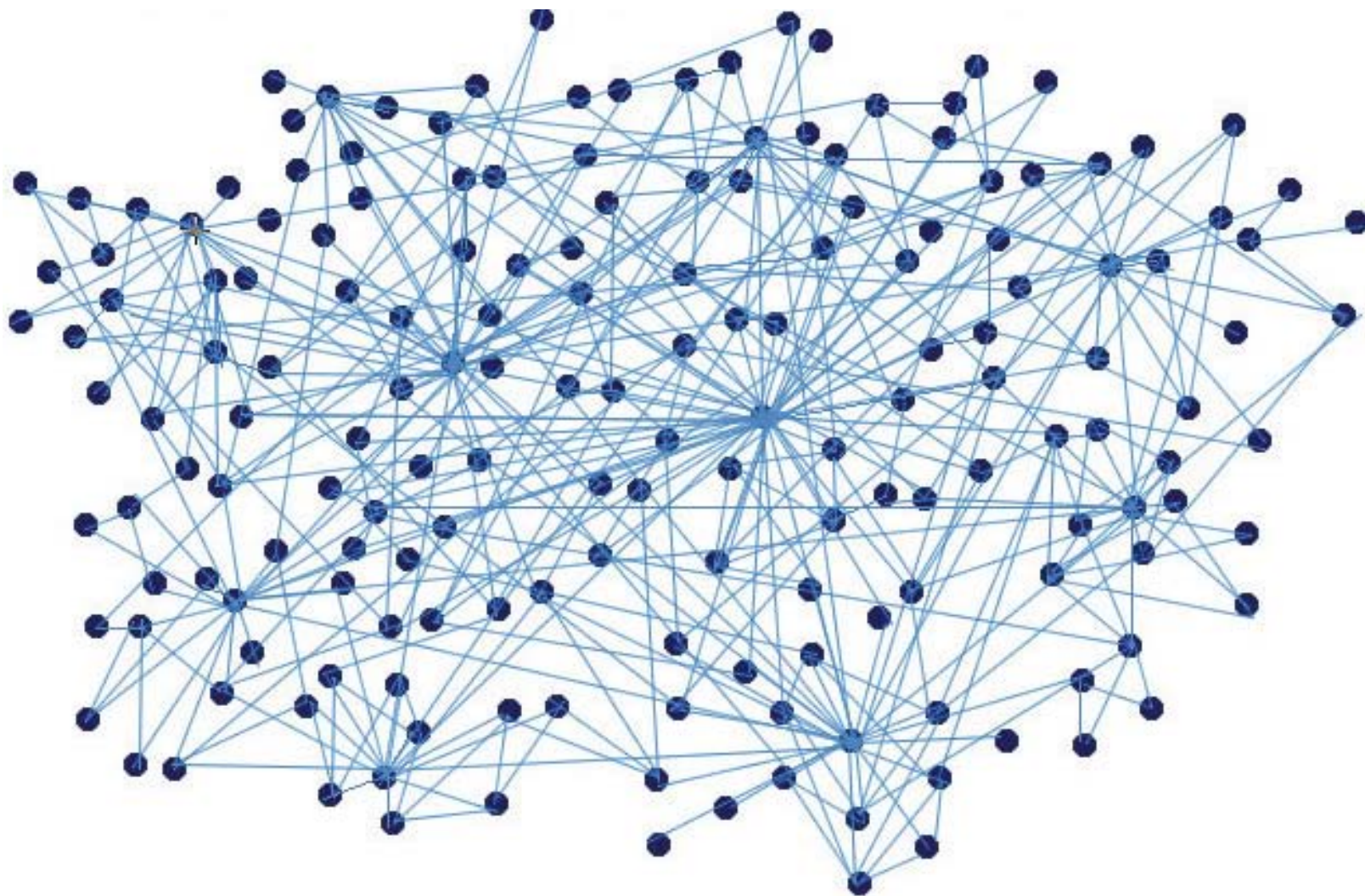
- ▶ Sviluppo del Web
- ▶ Relational Databases
- ▶ Triple e RDF
- ▶ SPARQL
- ▶ Linguaggi RDF
- ▶ Best Practices
- ▶ LOD Cloud
- ▶ OpenAgris (progetto FAO)

La Semantica

- ▶ La semantica è la scienza che studia il significato delle parole
- ▶ Comunicare un significato con una sequenza di simboli:
 - ▶ Luca ama i cani
 - ▶ I cani spaventano Maria
 - ▶ Chi ama i cani?
- ▶ Il Web semantico usa la semantica per rappresentare, combinare e condividere la conoscenza tra comunità di macchine

Web 1.0

- ▶ Il Web è un sistema di **documenti interconnessi** (interlinked) che possono essere acceduti attraverso **Internet**
- ▶ I **Web browsers** usano il protocollo **HTTP** per comunicare con **Web Servers**
- ▶ L'utente accede a pagine **HTML** e usa gli hyperlinks per navigare tra le pagine
- ▶ L'utente può così accedere a qualsiasi **documento** raggiungibile sul Web



Web 2.0

- ▶ Gli autori iniziano ad utilizzare Ajax
- ▶ Le persone **condividono** (share) foto, video e pensieri
- ▶ Le persone **interagiscono** sui social networks
- ▶ **Pubblicano** contenuti nei blogs, usano **tags** per dare significato ai documenti
- ▶ L'utente ha piena libertà di interagire con i siti Web, che diventano dinamici

Il Web Semantico

- ▶ Le pagine Web sono scritte in HTML
- ▶ HTML descrive la struttura delle informazioni (formattazione), cioè la sintassi e non la semantica
- ▶ Se una macchina potesse comprendere il significato dell'informazione, potrebbe aiutare l'utente a trovare ciò di cui ha veramente bisogno
- ▶ Il Web di oggi si basa su documenti interconnessi, il Web semantico su **Cose** e sulle **relazioni** tra Cose

Domande per il Web Semantico

- ▶ Quali sono i registi che hanno studiato alla Università di Berkeley e hanno iniziato la loro carriera come Stuntmen?
- ▶ Quali cantanti hanno iniziato la loro carriera nello stesso anno di Robby Williams?
- ▶ Quali banche di New York hanno prestato soldi al Senatore dello Utah?
- ▶ Quali giocatori sono andati all'Università di Austin e hanno giocato coi Dallas Cowboys come Cornerback? [Tom Landry]



football players at austin dallas cowboys cornerback



Ricerca

Circa 146.000 risultati (0,35 secondi)

SafeSearch

Tutto

Immagini

Maps

Video

Notizie

Shopping

Più contenuti

Roma

Cambia località

Nel Web

Pagine in italiano

Pagine da: Italia

Pagine straniere tradotte

Più strumenti

[Bryant Westbrook - Wikipedia, the free encyclopedia](#)

en.wikipedia.org/wiki/Bryant_Westbrook - Traduci questa pagina

Westbrook **played** college **football** at the University of Texas at **Austin** where he ...
Westbrook later **played** for the **Dallas Cowboys** in 2002 and was cut due to ...

[Miles Austin - Wikipedia, the free encyclopedia](#)

en.wikipedia.org/wiki/Miles_Austin - Traduci questa pagina

In **football**, he **played** wide receiver and defensive back, and as a senior, he won ... "Miles **Austin: Dallas Cowboys'** overnight sensation | Sports News | News for ...

[Miles Austin - Dallas Cowboys Blog - ESPN Dallas](#)

espn.go.com/.../dallas/cowboys/.../miles-austin - Traduci questa pagina

5 days ago – Visit Miles **Austin - Dallas Cowboys** Blog for information, in-depth analysis and ... Most of the **Cowboys** frankly looked like **football players** who hadn't swung a
With the **Cowboys** adding **cornerback** Akwasi Owusu-Ansah on ...

[Miles Austin | Dallas Cowboys | Stats, Pictures, News | NFL Football ...](#)

msn.foxsports.com/nfl/player/...austin/418819 - Traduci questa pagina

player Miles **Austin** nfl news, stats, fantasy info, scouting, awards, game logs, hometown, college, birth ... The **Dallas Cowboys** Like Having a Lot of **Cornerbacks** ...

[Frank Walker | Dallas Cowboys | Stats, Pictures, News | NFL ...](#)

msn.foxsports.com/nfl/player/frank.../140655 - Traduci questa pagina

23 May 2012 – **player** Frank Walker nfl news, stats, fantasy info, scouting, awards, game logs, ... LaMark Armour, Baraka Atkins, Junior Aumavae, Miles **Austin**, Dan Bailey, Alan Ball,

Raw Data

- ▶ Bisogna pubblicare qualcosa che le macchine siano in grado di comprendere
- ▶ Tim Berners-Lee, l'inventore del World Wide Web, suggerì di pubblicare **Raw Data**, in maniera aperta (Open)



- ▶ Bisogna esporre questi dati in un linguaggio che ne evidenzi la semantica e che sia comprensibile alle macchine

Stato attuale dei dati sul Web

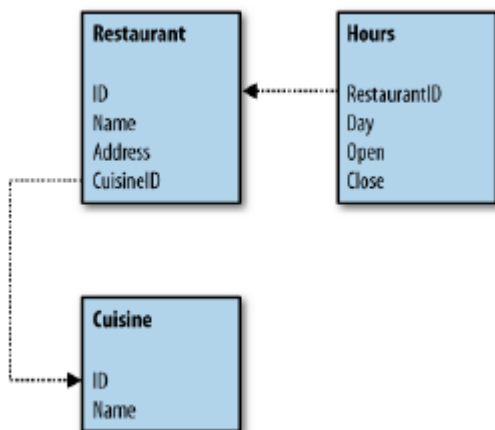
- ▶ Esistono diversi modi di modellare i dati:
 - ▶ Tabelle
 - ▶ Databases relazionali
 - ▶ XML, CSV ...
- ▶ Troppi formati e modelli di dati diversi
- ▶ È difficile integrare i dati e scoprire relazioni tra concetti
- ▶ I dati in diverse sorgenti di dati non sono interconnessi

Relational Databases

- ▶ Strumenti potenti e veloci per memorizzare grandi insiemi di dati
- ▶ Tecnologie mature che provengono da anni di ricerca e ottimizzazioni
- ▶ Il modello è ben definito e il pattern di utilizzo è prevedibile
- ▶ I dati di diverse tabelle sono collegati con foreign keys
- ▶ Difficoltà di evolvere gli schemi: continui **schema migration** verso nuovi schemi identificati
- ▶ Costoso e difficile per schemi molto complessi

Esempio

- ▶ Modelliamo le informazioni di un ristorante (indirizzo, orari e giorni di apertura, tipo di cucina, prezzi)
- ▶ Esempio dal testo «*Programming the semantic Web*, T. Segaran, C. Evans, J. Taylor. Copyright 2009. O'Reilly. ISBN: 978-0-596-15381-6.»



Restaurant				
ID	Name	Address	Price	CuisineID
1	Dell Lllama	Peachtree Rd	\$	1
2	Peking Inn	Lake St	\$\$\$	2

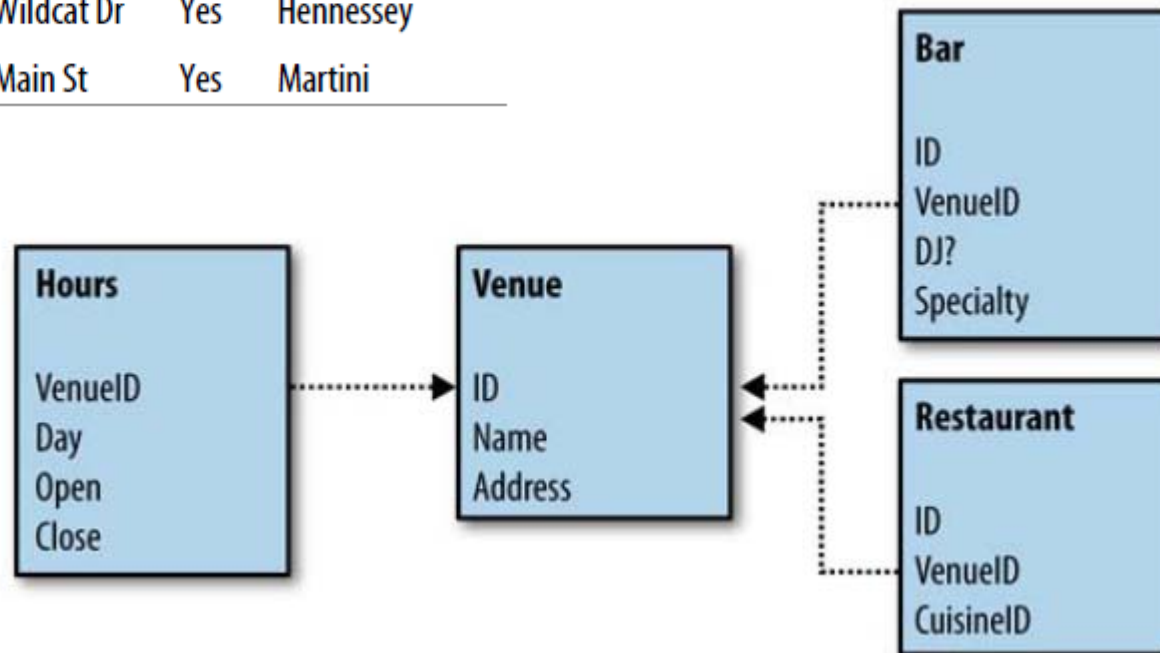
Cuisine		Hours			
ID	Name	RestID	Day	Open	Close
1	Dell	1	Mon	11	16
2	Chinese	1	Tue	11	16
3	Thal	1	Wed	11	16
4	Fast Food	1	Thu	11	19
		1	Fri	11	20
		2	Thu	5	22
		2	Fri	5	23
		2	Sat	5	23

Esempio

- ▶ NB: la semantica dei valori è descritta dallo schema (nomi delle tabelle e relazioni)
- ▶ Assumiamo ora di ricevere un nuovo database di bar con informazioni che non sono nello schema dei ristoranti
- ▶ Un ristorante può avere un bar, ma un bar può avere un DJ e una specialità di drink, ma un nome e un indirizzo in comune col ristorante
- ▶ Per evitare duplicazione dei dati, occorre cambiare lo schema e trasformare i dati secondo il nuovo schema (schema migration), introducendo la tabella *Venue (=sede)*

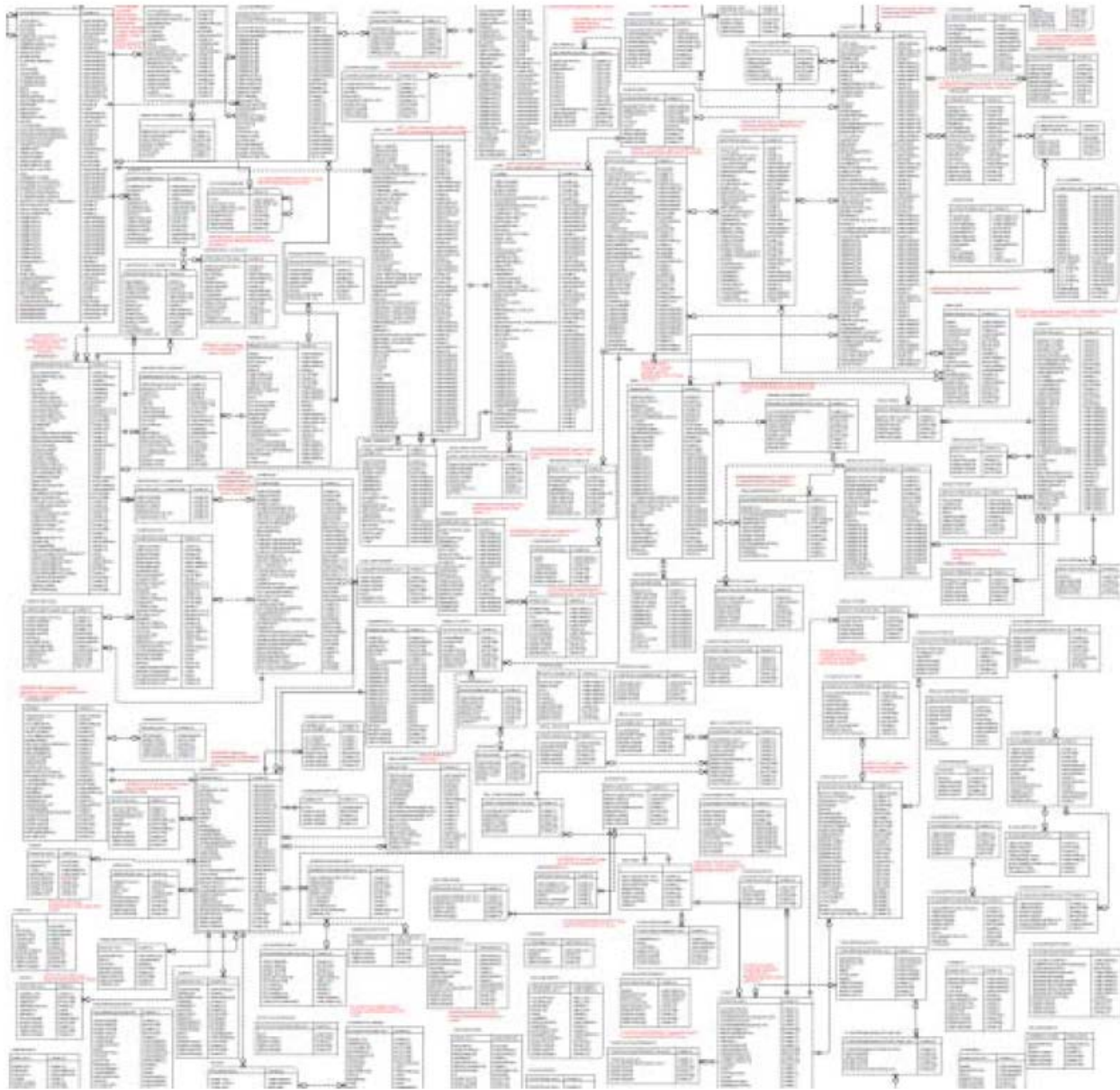
Esempio

Bar	Address	DJ	Specialty drink
The Bitter End	14th Ave	No	Beer
Peking Inn	Lake St	No	Scorpion Bowl
Hammer Time	Wildcat Dr	Yes	Hennessey
Marquis de Salade	Main St	Yes	Martini



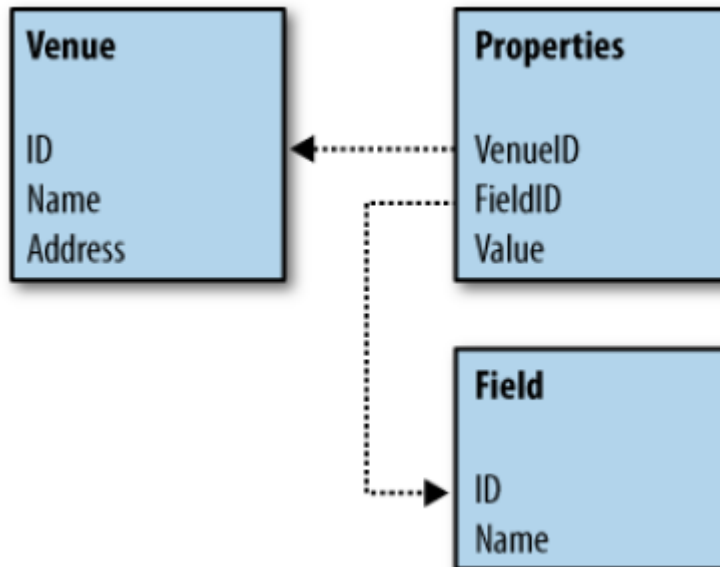
Esempio

- ▶ È facile immaginare che, maturando l'applicazione dei ristoranti, si possono aggiungere altri dettagli (sala per musica live, spazi in affitto per eventi...)
- ▶ Oltre alle migrazioni dovute alle evoluzioni dei dati, un altro problema è la complessità degli schemi quando si ha a che fare con diversi tipi di dati



Esempio più flessibile

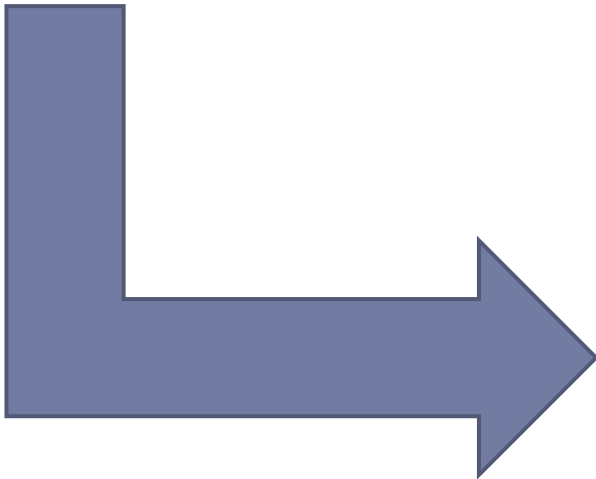
- ▶ Schema che degrada le performance del database ma flessibile



Relazioni Semantiche

- ▶ Estremizzando, possiamo spostare tutto il contenuto di Venue in un'unica tabella, quella delle proprietà
- ▶ Infatti, il nome e l'indirizzo sono semplici proprietà di Venue
- ▶ Facendo il JOIN con la tabella field, resta un'unica tabella di tre colonne

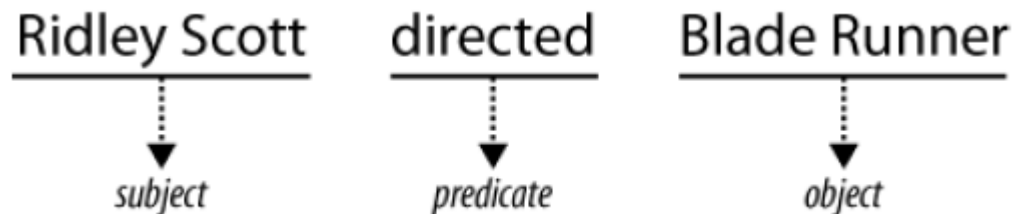
Properties			Field	
VenueID	FieldID	Value	ID	Name
1	1	Deli	1	Cuisine
1	2	\$	2	Price
1	7	Deli Llama	3	Specialty Cocktail
1	8	Peachtree Rd	4	DJ?
2	1	Chinese	5	Live Music?
2	2	\$\$\$	6	Music Genre
2	3	Scorpion Bowl	7	Name
2	4	No	8	Address
2	7	Peking Inn		
2	8	Lake St		
3	5	Yes		
3	6	Jazz		
3	7	Thai Tonic		
3	8	Branch Dr		



Properties		
VenueID	Field	Value
1	Cuisine	Deli
1	Price	\$
1	Name	Deli Llama
1	Address	Peachtree Rd
2	Cuisine	Chinese
2	Price	\$\$\$
2	Specialty Cocktail	Scorpion Bowl
2	DJ?	No
2	Name	Peking Inn
2	Address	Lake St
3	Live Music?	Yes
3	Music Genre	Jazz
3	Name	Thai Tonic
3	Address	Branch Dr

Le triple

- ▶ La tripla è il blocco fondamentale di una rappresentazione semantica
- ▶ Una tripla è composta da un soggetto, un predicato e un oggetto (asserzioni della logica dei predicati)

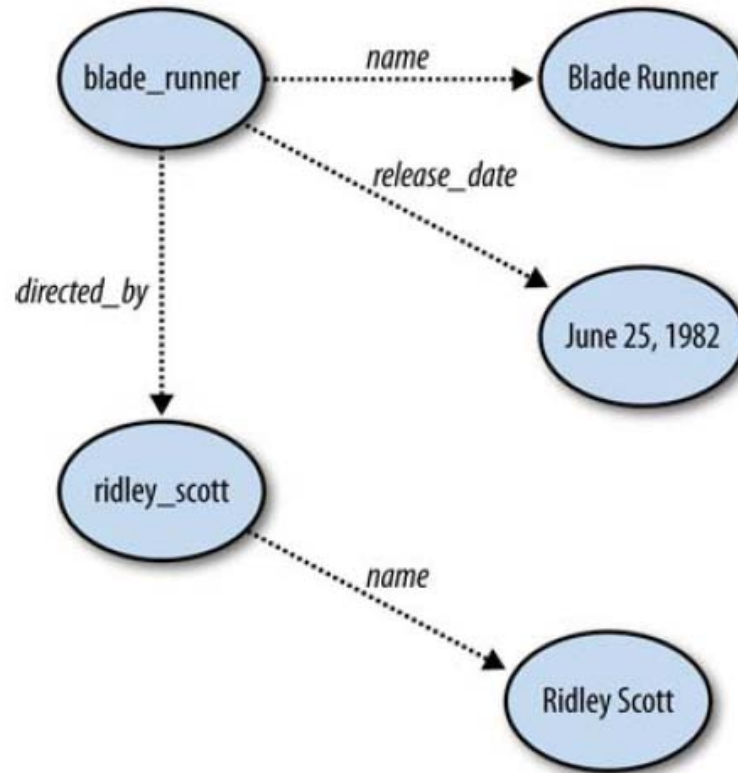


Le triple

- ▶ Il soggetto corrisponde in genere ad un concetto univoco, o entità, che può essere una persona, un luogo, un'idea...
- ▶ Il soggetto deve essere caratterizzato da un identificatore univoco globalmente (URI)
- ▶ I predicati rappresentano proprietà dell'entità a cui sono collegati
- ▶ Gli oggetti possono essere:
 - ▶ Soggetti di altre triple
 - ▶ Valori letterali, come stringhe o numeri

Grafi orientati

- ▶ Le triple possono essere assemblate in catene, formando un grafo orientato



RDF

- ▶ Strutturare dati in grafi è semplice, ma occorre renderli disponibili agli utenti
- ▶ RDF (Resource Description Framework) definisce uno standard per esprimere grafi di dati e condividerli con persone e macchine
- ▶ È un linguaggio per esprimere modelli di dati usando triple

URI e Risorse

- ▶ Ogni nodo del grafo deve avere un identificatore univoco, per poter essere referenziato
- ▶ Usiamo le **URI** (Universal Resource Identifier), nel formato schema:\\identificatore
- ▶ Ogni cosa esistente è concepita come risorsa, identificabile da una URI
- ▶ Il nodo del grafo **non è la cosa** identificata dalla URI, ma la rappresenta
- ▶ **Dereferenziare** una URI significa usarla in un browser per ottenere la descrizione della risorsa

Tim Berners-Lee



- ▶ *All kinds of conceptual things, they have names now that start with HTTP*
- ▶ Usare URIs per identificare le cose
- ▶ Usare URIs HTTP per poterle dereferenziare
- ▶ Fornire informazioni utili quando una URI è dereferenziata, usando formati standard come RDF/XML
- ▶ Collegare la risorsa ad altre URIs esterne, per favorire la scoperta di altre informazioni sul Web

Blank Nodes

- ▶ Se non si conosce la URI di una risorsa o non esiste un identificatore, si possono usare nodi vuoti oppure letterali

Formati di Serializzazione

- ▶ Un formato di serializzazione è una notazione per esprimere le triple RDF
- ▶ N-Triples (o Notation 3): files con estensione `.nt` in cui ogni linea è uno statement contente soggetto, predicato e oggetto
- ▶ N3: usa i namespaces e consente di raggruppare le triple associate allo stesso soggetto
- ▶ RDFa: annotazione di pagine XHTML con dati RDF (es. si usa l'attributo `about` per la URI del soggetto)

Esempi da Wikipedia

```
<http://www.w3.org/ntriples/> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
    <http://xmlns.com/foaf/0.1/Document> .
<http://www.w3.org/ntriples/> <http://purl.org/dc/terms/title> "N-Triples"@en-US .
<http://www.w3.org/ntriples/> <http://xmlns.com/foaf/0.1/maker> _:art .
_:art <http://www.w3.org/1999/02/22-rdf-syntax-ns#><http://xmlns.com/foaf/0.1/Person> .
_:art <http://xmlns.com/foaf/0.1/name> "Art Barstow".
```

@prefix dc: <http://purl.org/dc/elements/1.1/>.

```
<http://en.wikipedia.org/wiki/Tony_Benn>
    dc:title "Tony Benn";
    dc:publisher "Wikipedia".
```

RDF/XML

- ▶ Soluzione proposta dal W3C (consorzio per gli standards Web)
- ▶ Usa XML per esprimere modelli RDF
- ▶ L'elemento radice `<rdf:RDF>` dichiara i namespaces usati
- ▶ Il primo elemento definisce anche il **tipo** della risorsa
- ▶ L'attributo `rdf:about` specifica il soggetto di una risorsa quando esiste una URI

Esempio da AGRIS (FAO)

```
<bibo:Article rdf:about="http://agris.fao.org/aos/records/ES2011A01090">
  <dct:identifier>ES2011A01090</dct:identifier>
  <dct:title xml:lang="es"><![CDATA[Potencial]]></dct:title>
  <dct:creator>
    <foaf:Person>
      <foaf:name><![CDATA[Campo, Laura]]></foaf:name>
    </foaf:Person>
  </dct:creator>
  <dct:issued>2010</dct:issued>
  <dct:subject rdf:resource="http://aims.fao.org/aos/agrovoc/c_24013"/>
  <bibo:abstract xml:lang="es">
    <![CDATA[Los objetivos de este trabajo ...]]></bibo:abstract>
  <bibo:abstract xml:lang="en">
    <![CDATA[The objectives of this study ...]]></bibo:abstract>
  <bibo:language>esp</bibo:language>
</bibo:Article>
```

SPARQL

- ▶ Simple Protocol And RDF Query Language
- ▶ Linguaggio di interrogazione standardizzato per grafi RDF
- ▶ Ricerca, filtro dei risultati, costruzioni di grafi, espressioni regolari
- ▶ Quattro costrutti principali: **SELECT**, **CONSTRUCT**, **ASK**, **DESCRIBE**

Esempio da AGROVOC (FAO)

Data la URI di una risorsa AGROVOC (dizionario di termini usati nell'ambito dell'agricoltura), ottenere il nome della risorsa e le URI della stessa risorsa in altri vocabolari:

```
SELECT ?pl ?em {  
  
  OPTIONAL {<http://aims.fao.org/aos/agrovoc/c\_4039> skos:exactMatch ?em } .  
  
  <http://aims.fao.org/aos/agrovoc/c\_4039> skos:prefLabel ?pl .  
  
  FILTER ( ( lang(?pl)="en" ) )  
  
}
```

PI	em
"Japan"@en	< http://www.fao.org/countryprofiles/geopolitical/resource/Japan >
"Japan"@en	< http://eurovoc.europa.eu/1524 >
"Japan"@en	< http://lod.nal.usda.gov/nalt/35135 >
"Japan"@en	< http://zbw.eu/stw/descriptor/17568-3 >
"Japan"@en	< http://lod.gesis.org/thesoz/concept/10048140 >
"Japan"@en	< http://dewey.info/class/952/ >

Linguaggio FOAF

- ▶ FOAF (Friend of a Friend): usato per rappresentare informazioni sulle persone, come nome, compleanno, sito Web...
- ▶ Descrive relazioni tra persone che si conoscono
- ▶ Consultabile: <http://xmlns.com/foaf/spec/>

FOAF Core	Social Web
<ul style="list-style-type: none">◦ Agent◦ Person◦ name◦ title◦ img◦ depiction (depicts)◦ familyName◦ givenName◦ knows◦ based_near◦ age◦ made (maker)◦ primaryTopic (primaryTopicOf)◦ Project◦ Organization◦ Group◦ member◦ Document◦ Image	<ul style="list-style-type: none">• nick• mbox• homepage• weblog• openid• jabberID• mbox_sha1sum• interest• topic_interest• topic (page)• workplaceHomepage• workInfoHomepage• schoolHomepage• publications• currentProject• pastProject• account• OnlineAccount• accountName• accountServiceHomepage• PersonalProfileDocument• tipjar• sha1• thumbnail• logo

Dati Legacy

- ▶ I dati sul Web non sono stati pubblicati in formato semantico
- ▶ Occorre estrarli, convertirli in RDF, memorizzarli in un *triplestore* e pubblicarli
 - ▶ Si identificano e si leggono i dati dalla sorgente
 - ▶ Si crea uno schema per essi, cioè una serie di namespaces e predicati che useremo per modellare il grafo RDF
 - ▶ Generare l'RDF in uno dei formati di serializzazione
 - ▶ Caricare l'RDF in un triplestore, che creerà il grafo delle triple e consentirà le interrogazioni SPARQL

Triplestore

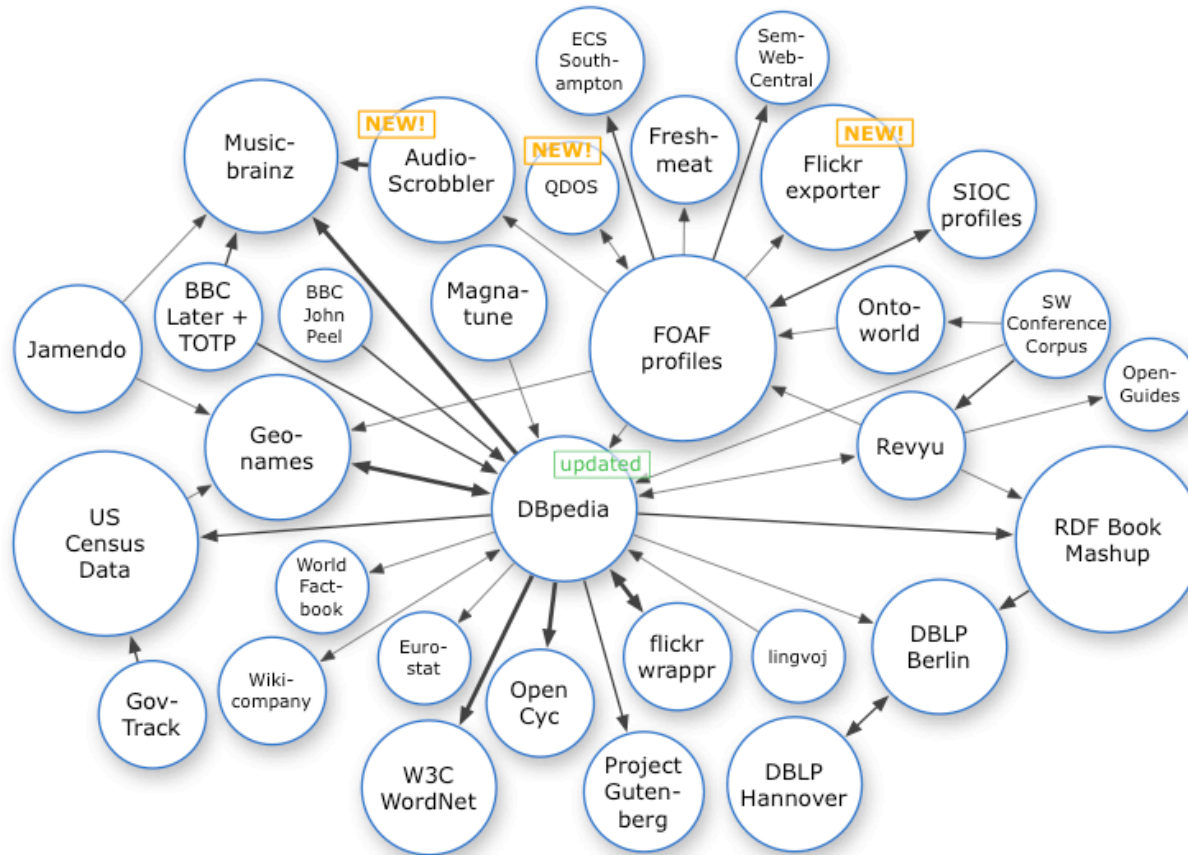
- ▶ È un database per memorizzare triple ed eseguire interrogazioni
- ▶ Può contenere miliardi di triple
- ▶ Sesame (Java, Open), Virtuoso (Commerciale), Allegrograph (Commerciale), OWLIM (Open, Commerciale)

Alcuni riferimenti

- ▶ *Programming the semantic Web*, T. Segaran, C. Evans, J. Taylor. Copyright 2009. O'Reilly. ISBN: 978-0-596-15381-6.
- ▶ <http://www.w3.org/RDF/>
- ▶ <http://www.w3.org/TR/rdf-sparql-query/>
- ▶ <http://www.w3.org/TR/void/>
- ▶ <http://xmlns.com/foaf/spec/>
- ▶ <http://dublincore.org/documents/dcmi-terms/>
- ▶ <http://bibliontology.com/>

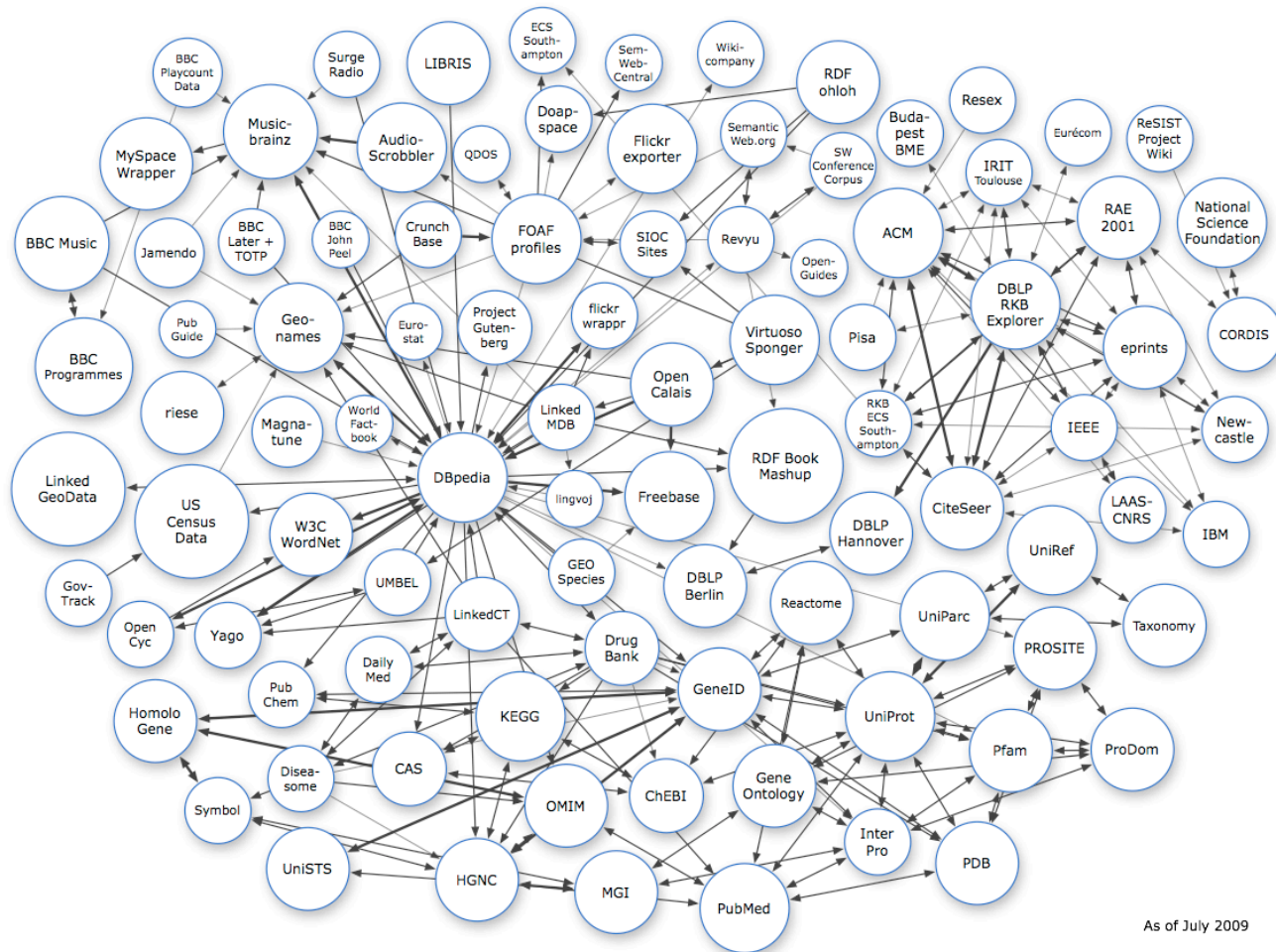
Linked Data Cloud 2008

“Linking Open Data cloud diagram, by Richard Cyganiak and Anja Jentzsch. <http://lod-cloud.net/>”



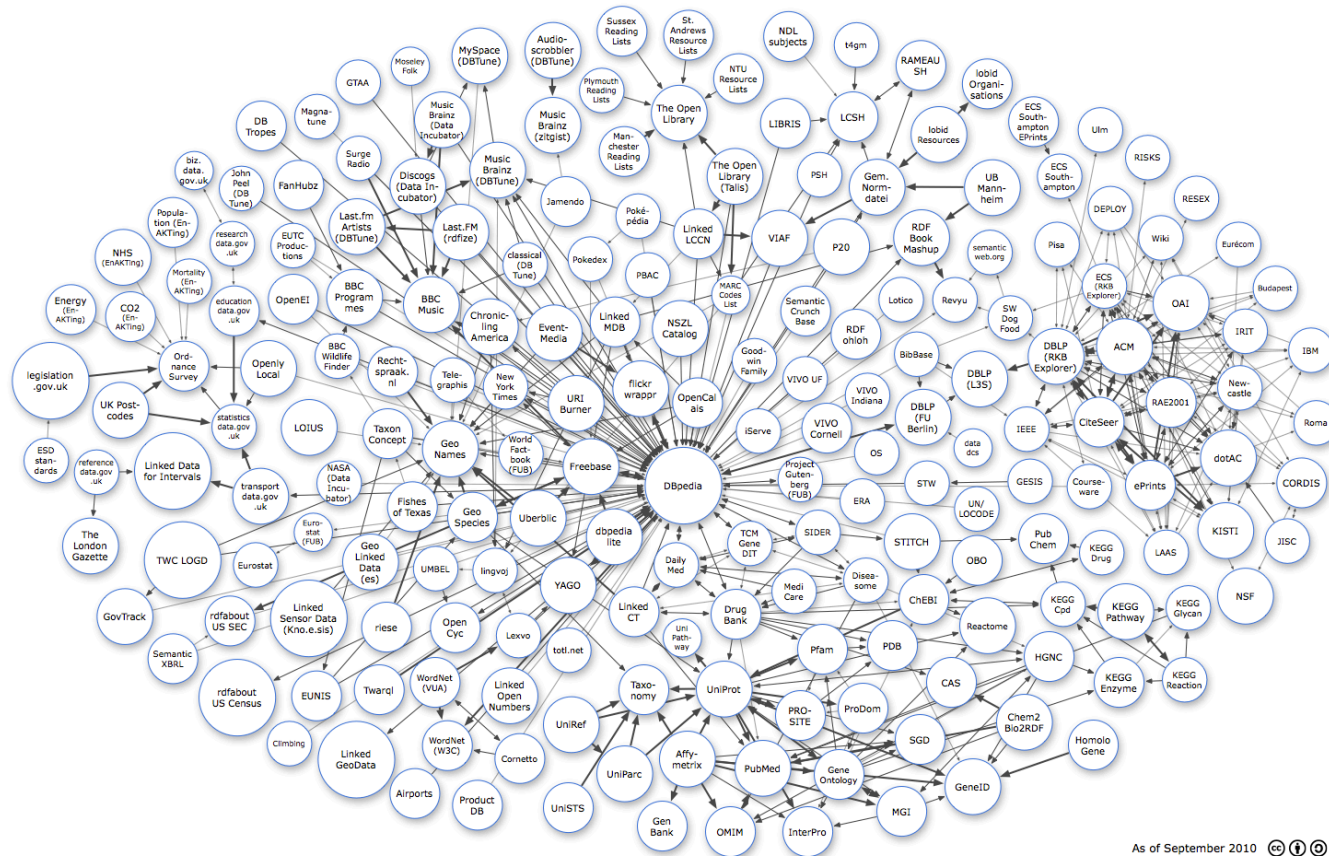
Linked Data Cloud 2009

“Linking Open Data cloud diagram, by Richard Cyganiak and Anja Jentzsch. <http://lod-cloud.net/>”



Linked Data Cloud 2010

“Linking Open Data cloud diagram, by Richard Cyganiak and Anja Jentzsch. <http://lod-cloud.net/>”

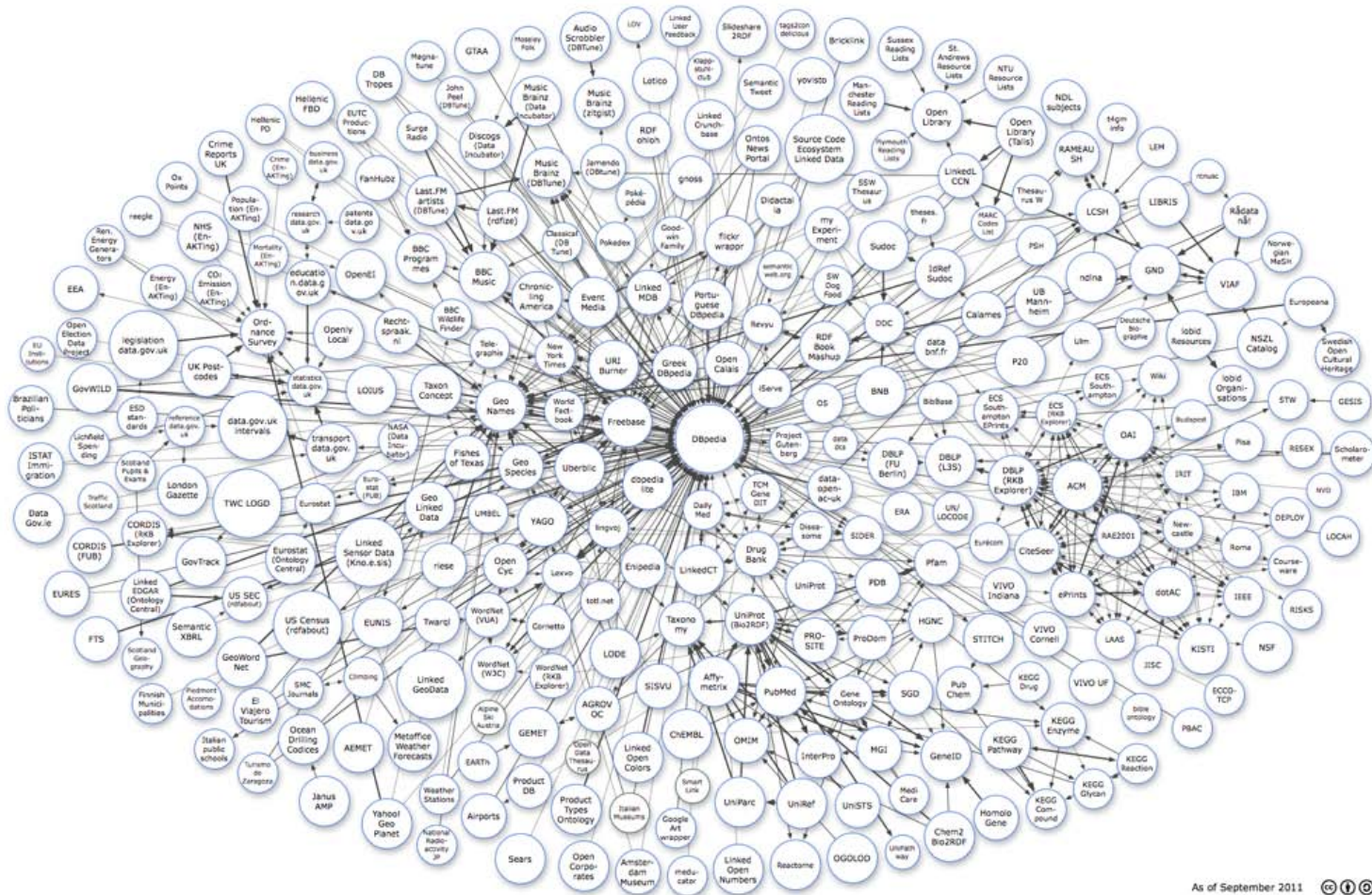



As of September 2010



Linked Data Cloud 2011

“Linking Open Data cloud diagram, by Richard Cyganiak and Anja Jentzsch. <http://lod-cloud>



As of September 2011 

W3C 5 Stars

- ▶ Le 5 regole d'oro per la pubblicazione di dati in RDF
- ▶ <http://www.w3.org/DesignIssues/LinkedData.html>





OpenAgris 
beta

AGRIS

- ▶ AGRIS nasce nel 1974 da un'iniziativa FAO per rendere le informazioni sulla ricerca in ambito agricolo disponibili a tutto il mondo e in maniera gratuita
- ▶ AGRIS è una collezione di 3 milioni di referenze bibliografiche (articoli scientifici, tesi, articoli di giornale...)
- ▶ AGRIS è un database XML sotto un Solr Search Engine
- ▶ AGRIS contiene metadati, cioè descrizioni della risorsa e non la risorsa stessa
- ▶ AGRIS è uno dei più grandi sistemi informativi nel dominio delle scienze agricole
- ▶ <http://agris.fao.org/>



Food and Agriculture
Organization of the
United Nations

for a world without hunger

FAO Home

AGRIS Search

About AGRIS

For Content Providers

Guide

AGRIMetaMaker

F.A.Q.

Login

AGRIS :: International Information System for the Agricultural Sciences and Tec


Search AGRIS - from 1975 to date

http://agris.fao.org?query=&request_locale=

Search

Advanced Search

Search Categories

 Search help

Search

Reset

AGRIS is a global public domain Database with 2830342 structured bibliographical records on agricultural science and technology. 79.78% of records are citations from scientific journals. The bibliographic references contain either links to the full text of the publication or additional information retrieved from related Internet resources.

> [Go to AGRIS search](#)



[Fish pathology](#) (Dec 2009)

Surveillance of salmonid viruses especially targeting infectious salmon anemia virus in Japan

[Kasai, H.](#), [Hokkaido Univ., Hakodate \(Japan\).](#) [Faculty of Fisheries Sciences;](#) [Iwawaki, S.;](#) [Yoshimizu, M.](#)

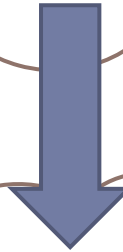
Date of publication	Dec 2009
AGRIS Categories	Aquaculture production and management; Veterinary science and hygiene; Pests of animals
AGROVOC English terms	Salmon; Trout; Viroses; Viruses; Pathogenesis; Infectious diseases; Anaemia; Haematopoietic necrosis virus; Disease surveys; Japan
AGROVOC French terms	Saumon; Truite; Virose; Virus; Pathogenese; Maladie infectieuse; Anemie; Haematopoietic necrosis virus; Enquete pathologique; Japon
AGROVOC Spanish terms	Salmon; Trucha; Virosis; Virus; Patogenesis; Enfermedades infecciosas; Anemia; Virus necrosis hematopoyetica; Encuestas sanitarias; Japon
Language	English
Notes	1 tab. 1 fig. 26 ref.
Type	Summary
Journal Title	Fish pathology
Vol. No.	v. 44(4) p. 182-184
Abstract (English)	Infectious salmon anemia (ISA) is a virus disease of Atlantic salmon <i>Salmo salar</i> in Europe and the Americas, but it has not been isolated in Far East Asia. In this study, we conducted virus isolation with ASK and ASE cells targeting ISA virus (ISAV) from a total of 5,967 fish belonging to eight salmonid species in Japan from 2005 to 2007. ISAV was not isolated from any fish examined but infectious hematopoietic necrosis virus was isolated from 116 fish belonging to three species, while infectious pancreatic necrosis virus was found in 14 fish from three species. It was considered that Japan is still free from ISAV.

AGRIS: Problemi

Dairy Herd Management (Jan 1977)
How stable is your vacuum level? [Milking, dairy cows]

Date of publication	Jan 1977
AGRIS Categories	Animal production
Language	English
Journal Title	Dairy Herd Management
Vol. No.	v. 14(1) p. 10, 14-15

Non c'è il collegamento al full-text della risorsa, non ci sono gli autori...



CARENZA DI INFORMAZIONI!

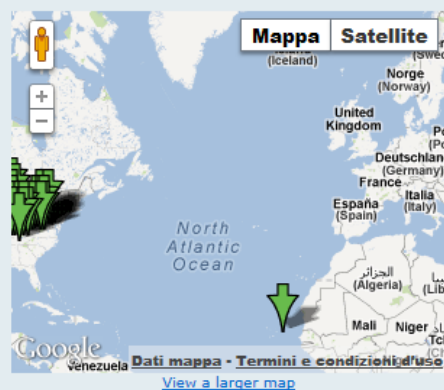


Impossibilità di connettersi ad altre sorgenti di dati. L'utente è interessato al full-text.

OpenAgris

- ▶ OpenAgris è un'applicazione Web interamente basata su RDF
- ▶ Aggrega informazioni da diverse sorgenti di dati per espandere la conoscenza di AGRIS su un argomento o una pubblicazione
- ▶ OpenAgris è un ambiente LOD che combina vari datasets interconnessi, producendo un risultato significativo
- ▶ La traduzione in RDF ha prodotto circa 80 milioni di triple
- ▶ Usando il dizionario Agrovoc come backbone, OpenAgris si connette a molti altri dataset esistenti (DBPedia, WorldBank, Geopolitical Ontology...)

viruses distribution map. Data from [Global Biodiversity Information Facility \(GBIF\)](#)



Statistics from FAO Country Profiles and IFPRI

Japan

POPULATION TOTAL (2010): 126 536 000 [\[SOURCE\]](#)

Gross domestic product

GDP TOTAL (2009): 5 068 996 (millions of US dollars) [\[SOURCE\]](#)

Human Development Index
(0 - 1 / low - high)

HDI TOTAL (2010): 0.884 (Calculated) [\[SOURCE\]](#)

FROM: [FAO Country Profiles](#)

Surveillance of salmonid viruses especially targeting infectious salmon anemia virus in Japan [dec2009]

<http://agris.fao.org/aos/records/JP2010001379>

*Kasai, H., Hokkaido Univ., Hakodate (Japan). Faculty of Fisheries Sciences
Iwawaki, S.
Yoshimizu, M.*

Infectious salmon anemia (ISA) is a virus disease of Atlantic salmon *Salmo salar* in Europe and the Americas, but it has not been isolated in Far East Asia. In this study, we conducted virus isolation with ASK and ASE cells targeting ISA virus (ISAV) from a total of 5,967 fish belonging to eight salmonid species in Japan from 2005 to 2007. ISAV was not isolated from any fish examined but infectious hematopoietic necrosis virus was isolated from 116 fish belonging to three species, while infectious pancreatic necrosis virus was found in 14 fish from three species. It was considered that Japan is still free from ISAV.

Agrovoc Keywords:

- [Anaemia](#)
- [Trout](#)
- [infectious diseases](#)
- [Viroses](#)
- [Disease surveys](#)
- [viruses](#)
- [pathogenesis](#)
- [Salmon](#)
- [Haematopoietic necrosis virus](#)
- [Pests of animals](#)
- [Japan](#)

Data from DBpedia:

▶	Trout
▶	Pathogenesis
▶	Salmon
▶	Japan

Data from Google™

▼ About the Title

[Surveillance of Salmonid Viruses Especially Surveillance of Salmonid Viruses Especially Targeting Infectious Salmon Anemia Virus in Japan](#). Author(s): Kasai, H; Iwawaki, S; Yoshimizu, M. Journal Title: ...
[discover-decouvirr.cisti-icist.nrc-cnrc.gc.ca](#)

From FAO serials dataset

▼ **Fish Pathology (Journal)**

http://aims.fao.org/serials/c_40a1aaa9

ISSN: 0388-788X

FREQUENCY: Quarterly

START DATE: 1966

▼ Agris articles from the same journal:

- Mortality in pond-cultured shrimp *Penaeus monodon* in the Philippines associated with *Vibrio harveyi* and white spot syndrome virus
- Effectiveness of fasting against red sea bream iridoviral disease in red sea bream
- Field trials of a vaccine with water-soluble adjuvant for bacterial coldwater disease in ayu *Plecoglossus altivelis*
- Histological detection of aquatic fungi by Uvitex 2B, a fluorescent dye

Cosa c'è dietro?

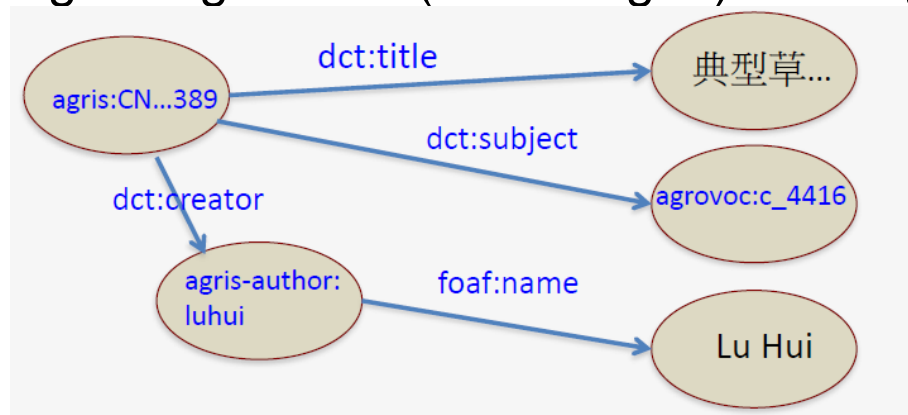
- ▶ Traduzione del repository XML in RDF
- ▶ Scelta dei vocabolari da utilizzare: usare vocabolari diffusi aiuta le macchine a comprendere il significato dei dati e a collegare diversi datasets
- ▶ Creazione di nuovi vocabolari quando mancano le proprietà per modellare una determinata relazione
- ▶ Pulizia e normalizzazione dei dati
- ▶ Applicazione di tags Agrovoc ai records per aumentare il significato semantico e favorire le connessioni
- ▶ Scelta dei datasets a cui interconnettersi e generazione del mapping verso quei datasets (si usa Agrovoc come backbone)

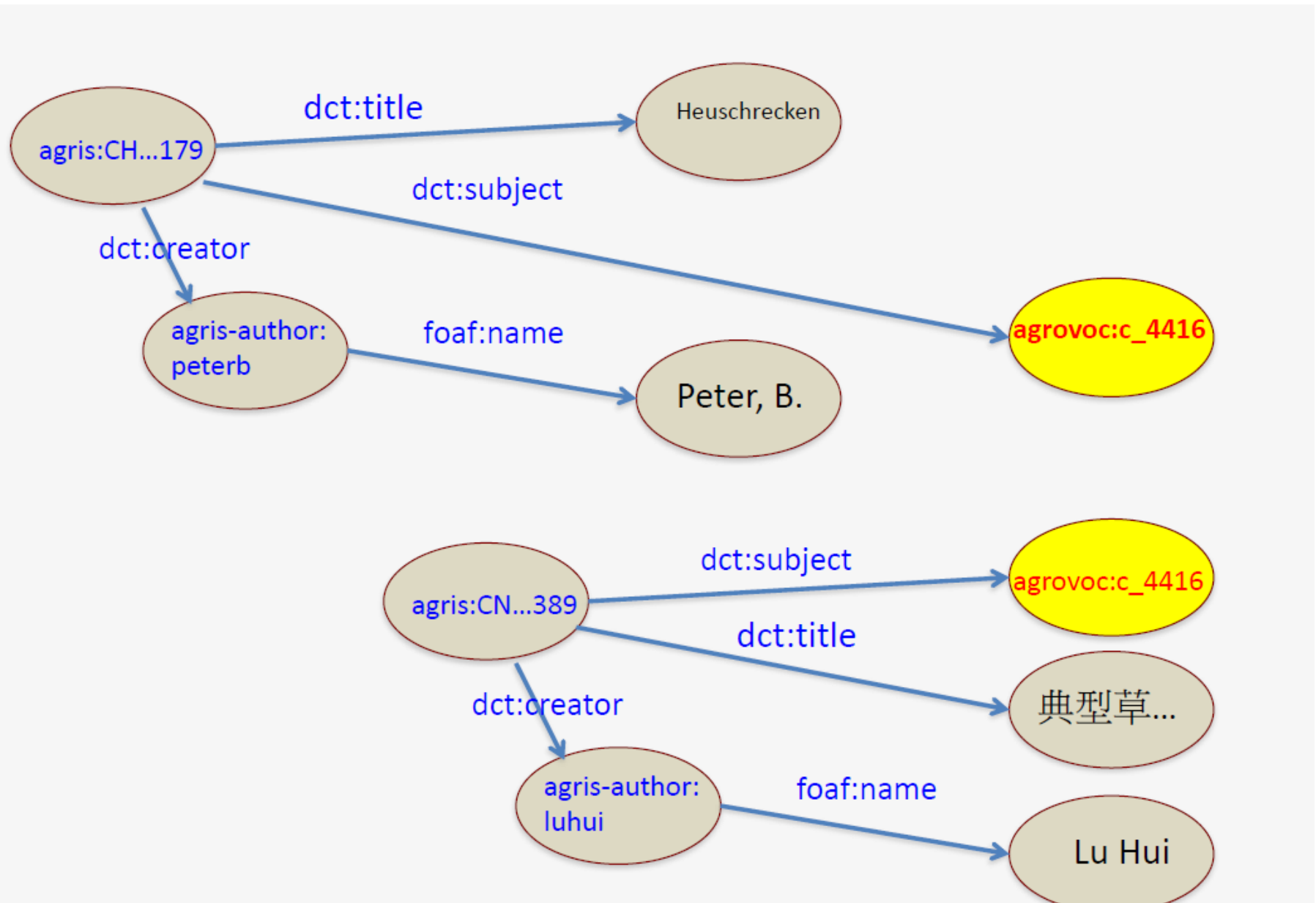
Esempio: AGRIS record (AGRIS AP XML)

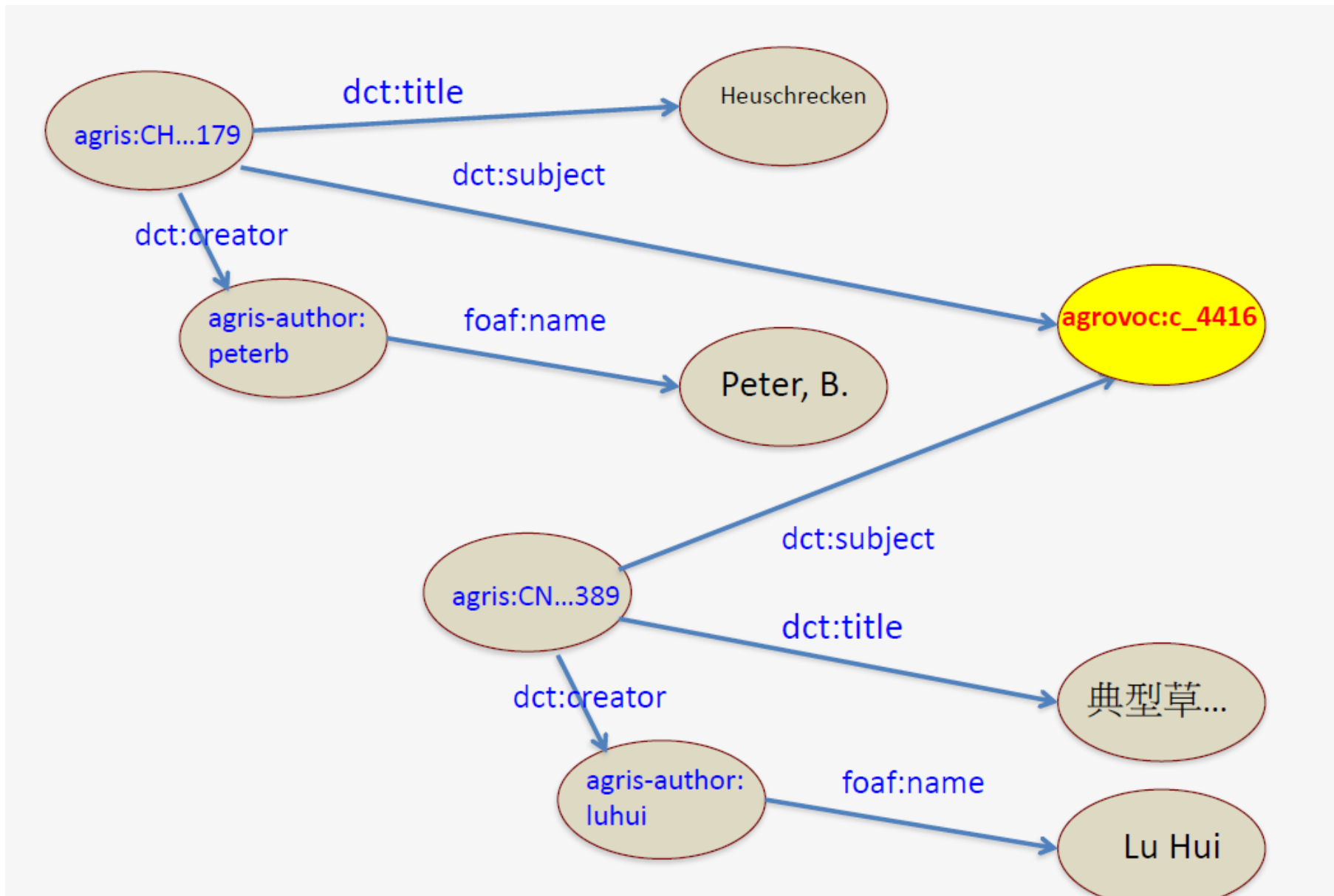
```
<ags:resource ags:ARN="CN2009002389">
  <dc:title xml:lang="en">Study on the mortality and competition of ...</dc:title>
  <dc:title xml:lang="Zh">
    <dcterms:alternative>典型草原三种蝗虫种群死亡率和竞争的研究</dcterms:alternative>
  </dc:title>
  <dc:creator>
    <ags:creatorPersonal>Lu Hui</ags:creatorPersonal>
    <ags:creatorPersonal>Han Jianguo</ags:creatorPersonal>
  </dc:creator>
  <dc:date><dcterms:dateIssued>Sep. 2008</dcterms:dateIssued></dc:date>
  <dc:subject>
    <ags:subjectThesaurus scheme="ags:AGROVOC" xml:lang="en">ACRIDIDAE</ags:subjectThesaurus>
  </dc:subject>
  <dc:description>
    <dcterms:abstract xml:lang="Zh">通过田间...</dcterms:abstract>
    <dcterms:abstract xml:lang="En">The three ...</dcterms:abstract>
  </dc:description>
  <dc:language scheme="ags:ISO639-1">Zh</dc:language>
  <ags:citation>
    <ags:citationTitle xml:lang="Zh">Acta Agrestia Sinica</ags:citationTitle>
    <ags:citationIdentifier scheme="ags:ISSN">1007-0435</ags:citationIdentifier>
  </ags:citation>
</ags:resource>
```


Traduzione in RDF

- ▶ Assegnazione di URI alla risorsa (la URI deve essere unica e permanente nel tempo):
<http://agris.fao.org/aos/records/CN2009002389>
- ▶ Scelta dei vocabolari
- ▶ Assegnazione URIs ai termini Agrovoc che determinano gli argomenti della referenza
- ▶ Assegnazione URIs ai giornali che hanno articoli in AGRIS
- ▶ Futuro: disambiguare gli autori (ora stringhe) e assegnare URIs









Acrididae

at AGROVOC Thesaurus

http://aims.fao.org/aos/agrovoc/c_4416

Property	Value
altLabel	<ul style="list-style-type: none">■ Criquet (fr)■ آکریدیده (fa)
broader	<ul style="list-style-type: none">■ <http://aims.fao.org/aos/agrovoc/c_5433>
closeMatch	<ul style="list-style-type: none">■ <http://dbpedia.org/resource/Acrididae>
exactMatch	<ul style="list-style-type: none">■ <http://lod.nal.usda.gov/nalt/3016>
<i>date of creation</i>	<ul style="list-style-type: none">■ 1981-01-09
<i>date of last update</i>	<ul style="list-style-type: none">■ 2007-04-20
hasMember	<ul style="list-style-type: none">■ <http://aims.fao.org/aos/agrovoc/c_31924>
hasTaxonomicLevel	<ul style="list-style-type: none">■ <http://aims.fao.org/aos/agrovoc/c_330937>
isThemeOf	<ul style="list-style-type: none">■ <http://aims.fao.org/aos/agrovoc/c_49974>
narrower	<ul style="list-style-type: none">■ <http://aims.fao.org/aos/agrovoc/c_1210>■ <http://aims.fao.org/aos/agrovoc/c_2349>■ <http://aims.fao.org/aos/agrovoc/c_26587>■ <http://aims.fao.org/aos/agrovoc/c_31925>■ <http://aims.fao.org/aos/agrovoc/c_31926>■ <http://aims.fao.org/aos/agrovoc/c_32308>■ <http://aims.fao.org/aos/agrovoc/c_32309>■ <http://aims.fao.org/aos/agrovoc/c_33397>■ <http://aims.fao.org/aos/agrovoc/c_33742>
prefLabel	<ul style="list-style-type: none">■ ACRIDIDAE (de)■ Acrididae (lo)■ Acrididae (hu)■ Acrididae (sk)■ Acrididae (cs)■ Acrididae (th)■ Acrididae (es)■ Acrididae (en)■ Acrididae (fr)

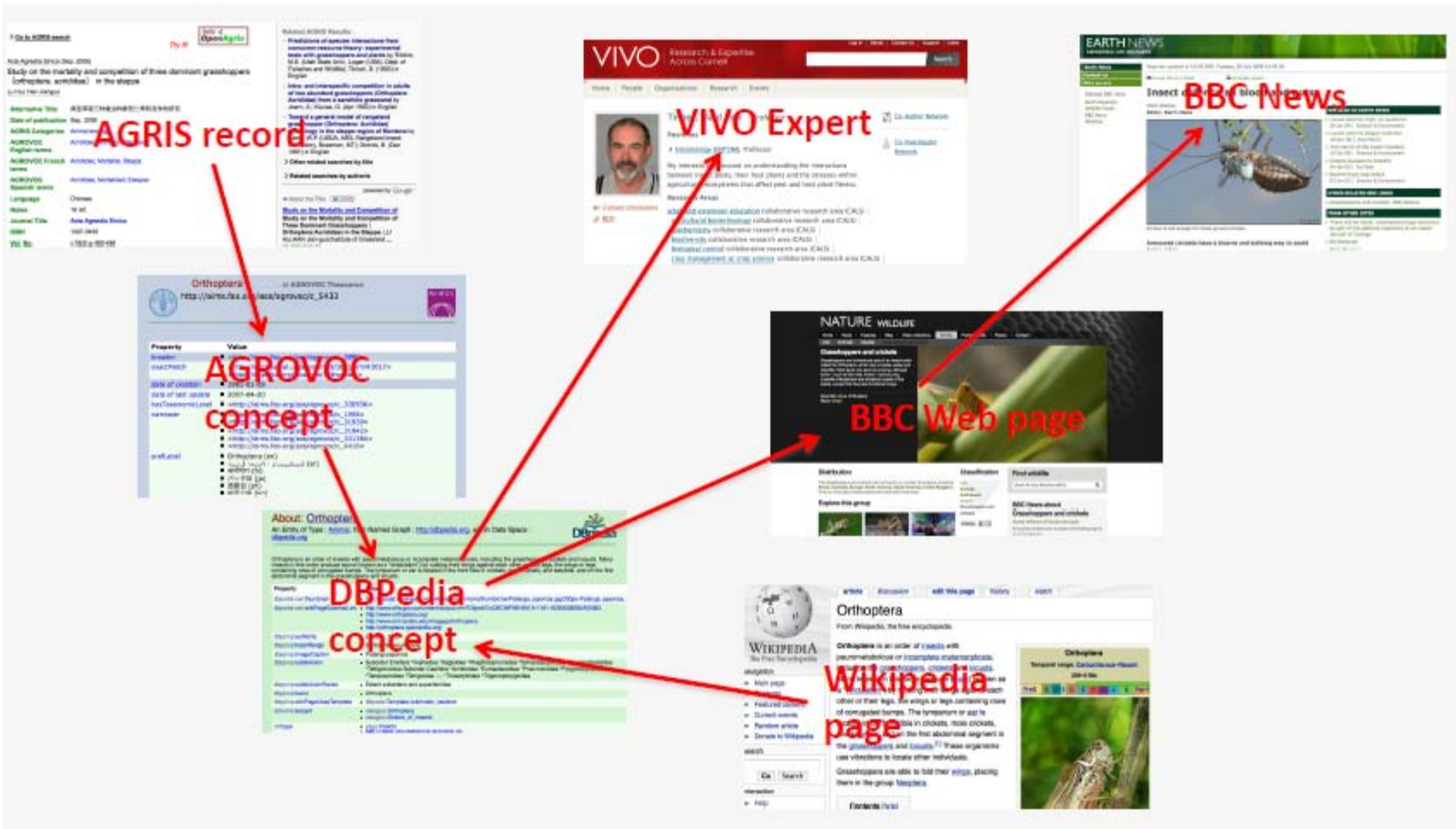
About: [Acrididae](#)

An Entity of Type : [eukaryote](#), from Named Graph : <http://dbpedia.org>, within Data Space : [dbpedia.org](#)

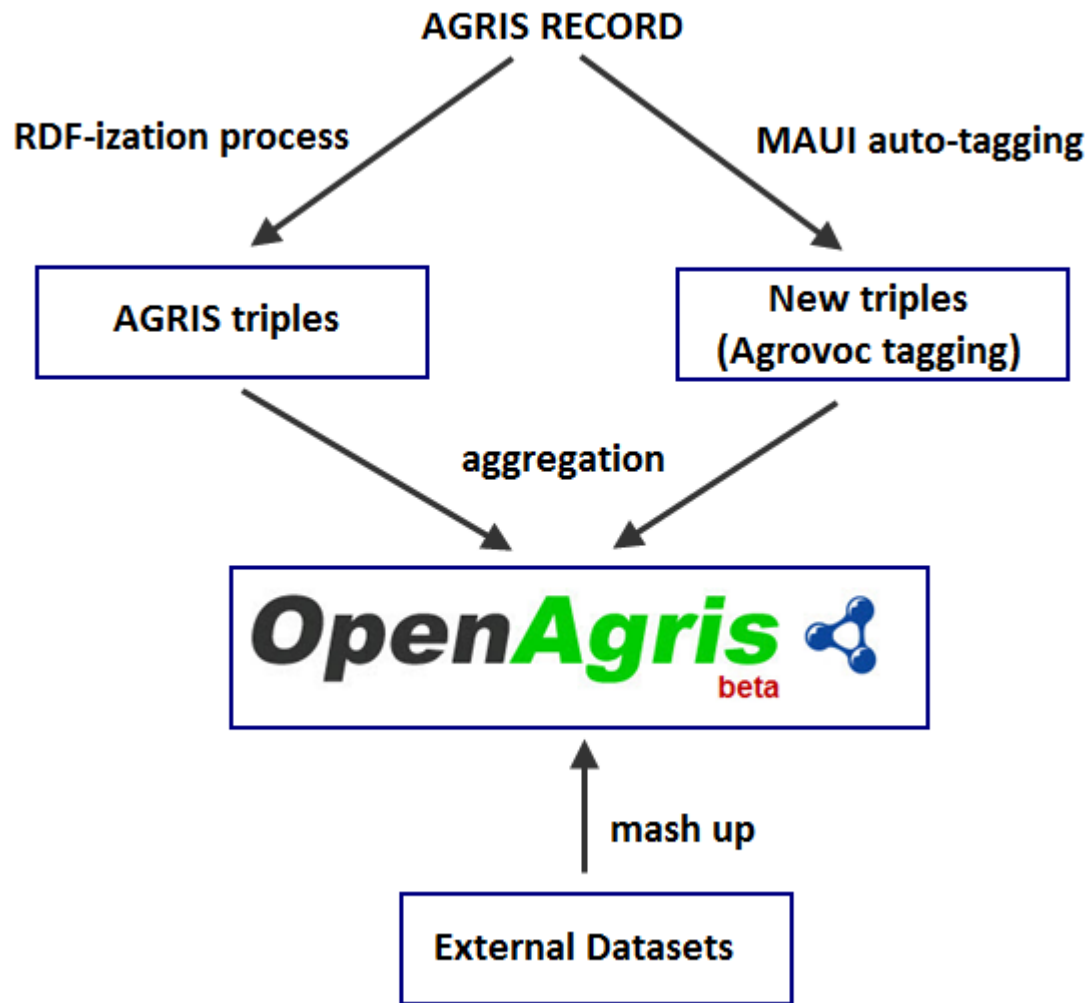


Gli Acrididi sono una famiglia di insetti ortotteri del sottordine Celiferi.

Property	Value
dbpedia-owl:abstract	<ul style="list-style-type: none">▪ The Acrididae are the predominant family of grasshoppers, comprising some 10,000 of the 11,000 species of the entire suborder Caelifera. The Acrididae are best known because all locusts (swarming grasshoppers) are of the Acrididae. The subfamily Oedipodinae is sometimes classified as a distinct family Oedipodidae in the superfamily Acridoidea. Acrididae grasshoppers are characterized by antennae relatively short and stout, and tympana on the side of the first abdominal segment. The name "Acrididae" is derived from Latin acer, meaning sharp, bitter, or caustic. http://wordinfo.info/words/index/info/view_unit/12/?letter=a&page=1&spage=1&s=acrid▪ Los acrididos (Acrididae), conocidos popularmente como langostas o saltamontes, son una familia de insectos ortópteros caracterizados por su gran facilidad para migrar de un sitio a otro y, en determinadas circunstancias, reproducirse muy rápidamente llegando a formar devastadoras plagas capaces de acabar con la vegetación de grandes extensiones de terreno. Generalmente no vuelan mucho a pesar de que disponen de alas. Sólo cuando se juntan grandes grupos de individuos de la misma especie liberan las feromonas apropiadas para activar la conducta migratoria y un crecimiento mayor de las alas y de esa manera poder dispersarse por otros territorios, evitando la competencia intraespecífica por el alimento.▪ Gli Acrididi sono una famiglia di insetti ortotteri del sottordine Celiferi.▪ Les Acrididae sont une famille d'orthoptères caelifères.
dbpedia-owl:class	<ul style="list-style-type: none">▪ dbpedia:Insect
dbpedia-owl:family	<ul style="list-style-type: none">▪ dbpedia:Acridoidea
dbpedia-owl:kingdom	<ul style="list-style-type: none">▪ dbpedia:Animal
dbpedia-owl:order	<ul style="list-style-type: none">▪ dbpedia:Grasshopper▪ dbpedia:Orthoptera
dbpedia-owl:phylum	<ul style="list-style-type: none">▪ dbpedia:Arthropod
dbpedia-owl:synonym	<ul style="list-style-type: none">▪ Acrididae MacLeay 1821
dbpedia-owl:thumbnail	<ul style="list-style-type: none">▪ http://upload.wikimedia.org/wikipedia/commons/thumb/c/c8/Coryphistes_ruricola.jpg/200px-Coryphistes_ruricola.jpg
dbpedia-owl:wikiPageExternalLink	<ul style="list-style-type: none">▪ http://osf2x.orthoptera.org/HomePage.aspx▪ http://wordinfo.info/words/index/info/view_unit/12/?letter=a&page=1&spage=1&s=acrid▪ http://osf2x.orthoptera.org/Common/basic/Taxa.aspx?TaxonNameID=3207
dbpprop:classis	<ul style="list-style-type: none">▪ Insecta
dbpprop:familia	<ul style="list-style-type: none">▪ Acrididae
dbpprop:familiaAuthority	<ul style="list-style-type: none">▪ MacLeay, 1819
dbpprop:imageCaption	<ul style="list-style-type: none">▪ Bark Mimicking Grasshopper, <i>Coryphistes ruricola</i>
dbpprop:imageWidth	<ul style="list-style-type: none">▪ 240 (xsd:integer)
dbpprop:name	<ul style="list-style-type: none">▪ Acrididae



Tagging Automatico



Tagging Automatico

- ▶ Il tagging automatico consente di aggiungere parole chiave prese dal dizionario Agrovoc alla referenza bibliografica
- ▶ Ciò consente di aggiungere semantica al record e aumenta la possibilità di interconnettersi a datasets esterni
- ▶ Si usa MAUI 1.2 «trained» con Agrovoc e alcuni documenti indicizzati dalla FAO stessa (<http://code.google.com/p/maui-indexer/>)
- ▶ Possibile solo quando AGRIS contiene il link al full-text per la risorsa: il full-text deve essere scaricato e analizzato
- ▶ Se non c'è tale link, si possono usare le API di Google per trovarlo sul Web

Sviluppi futuri

- ▶ Web Crawler per estrarre automaticamente triple dal Web (da siti *trusted*), annotare le risorse con Agrovoc e memorizzarle nel triplestore
- ▶ Previste un miliardo di triple
- ▶ Problemi di performances: studi sull'uso del GRID computing per eseguire il crawling del Web
- ▶ Studi sull'ottimizzazione e la parallelizzazione delle queries SPARQL per ottenere risposte in tempo reale
- ▶ Studi sul **reasoning** in RDF per scoprire relazioni nascoste

Alcuni riferimenti

- ▶ <http://agris.fao.org/>
- ▶ <http://aims.fao.org/openagris>
- ▶ <http://aims.fao.org/agrovoc/lod>
- ▶ <http://aims.fao.org/standards/lode-bd>