

# Linked Data and RDF

COMP60421

Sean Bechhofer

sean.bechhofer@manchester.ac.uk

# Building a Semantic Web

- **Annotation**
  - Associating metadata with resources
- **Integration**
  - Integrating information sources
- **Inference**
  - Reasoning over the information we have.
  - Could be light-weight (taxonomy)
  - Could be heavy-weight (logic-style)
- **Interoperation** and **Sharing** are key goals

## Linked Data\*

- Linked Data or the Data Web is about using the Web to connect related data that wasn't previously linked.
- The intention is that we move from a web of *documents* to a web of *data*
  - The Web as database
- The Linked Data approach builds heavily on RDF.

\*Linked data slides based on material from Ian Davis and Tom Heath:  
<http://www.slideshare.net/iandavis/30-minute-guide-to-rdf-and-linked-data>

## Database tables

isbn	title	author	publisherId	pages
0743267478	Q&A	Vikas Swarup	1435	336
014029466X	The Rotters' Club	Jonathan Coe	1546	416
...	...	...	...	...
..	...	...	...	...

## Rows represent “things”

isbn	title	author	publisherId	pages
0743267478	Q&A	Vikas Swarup	1435	336
014029466X	The Rotters' Club	Jonathan Coe	1546	416
...	...	...	...	...
..	...	...	...	...

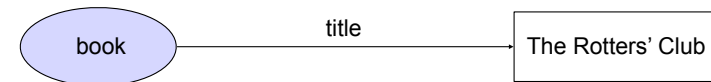
## Columns represent “properties”

isbn	title	author	publisherId	pages
0743267478	Q&A	Vikas Swarup	1435	336
014029466X	The Rotters' Club	Jonathan Coe	1546	416
...	...	...	...	...
..	...	...	...	...

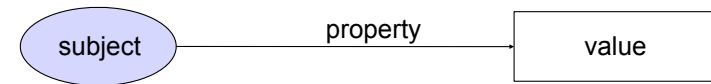
## Intersections represent properties of things

isbn	title	author	publisherId	pages
0743267478	Q&A	Vikas Swarup	1435	336
014029466X	The Rotters' Club	Jonathan Coe	1546	416
...	...	...	...	...
..	...	...	...	...

## Graphical Representation



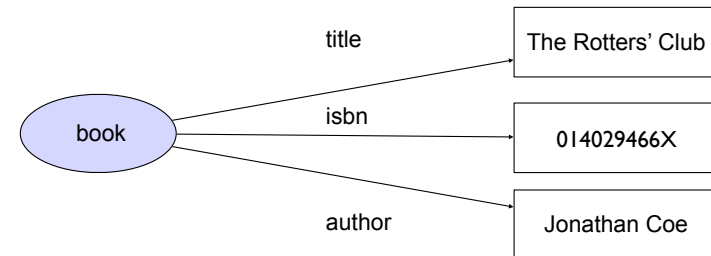
more generally:



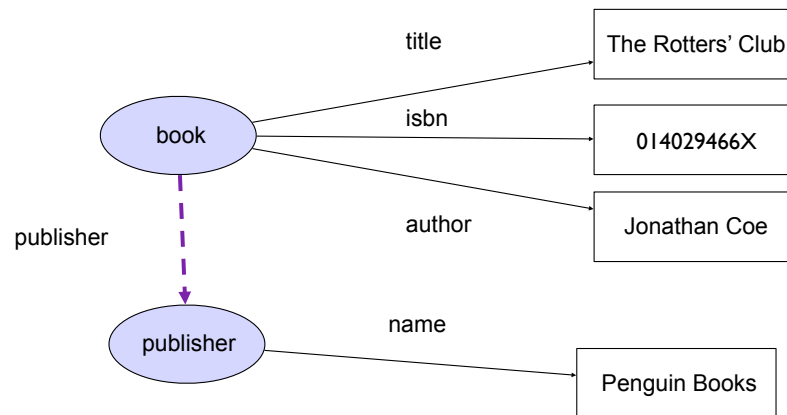
## Selecting multiple properties

isbn	title	author	publisherId	pages
0743267478	Q&A	Vikas Swarup	1435	336
014029466X	The Rotters' Club	Jonathan Coe	1546	416
...	...	...	...	...
..	...	...	...	...

## Multiple properties graphically



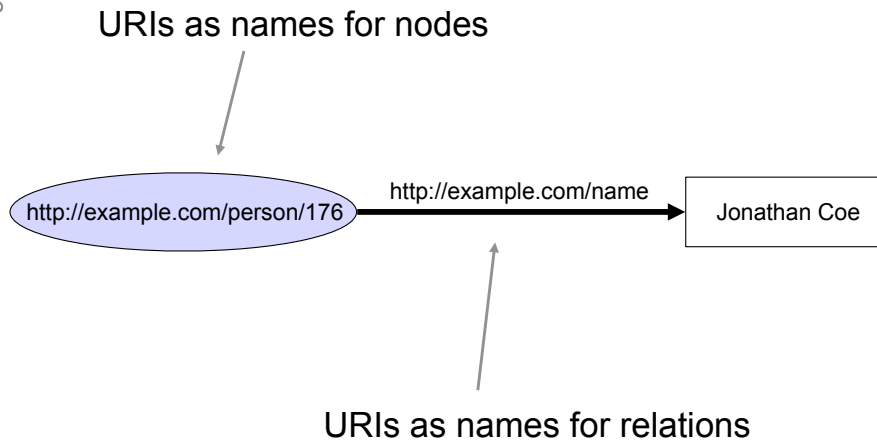
## Relations between “things”



## Identification

- We need to be able to identify things globally and uniquely.
- URIs provide this capability
- Key to Linked Data is the use of URIs, specifically http:// URIs.

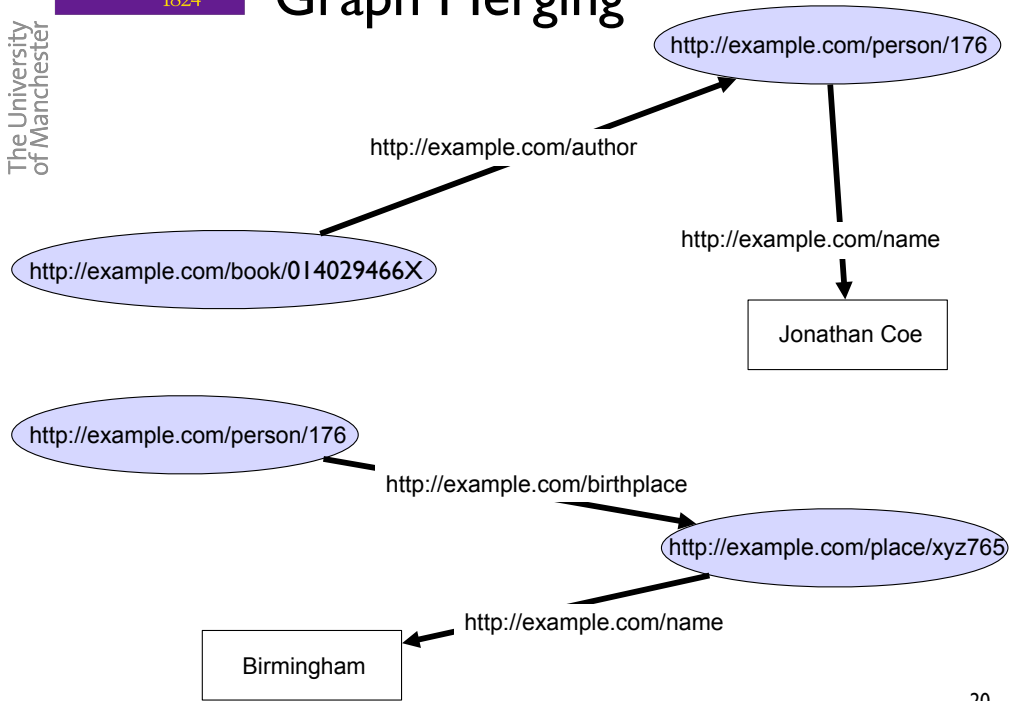
## URIs in graphs



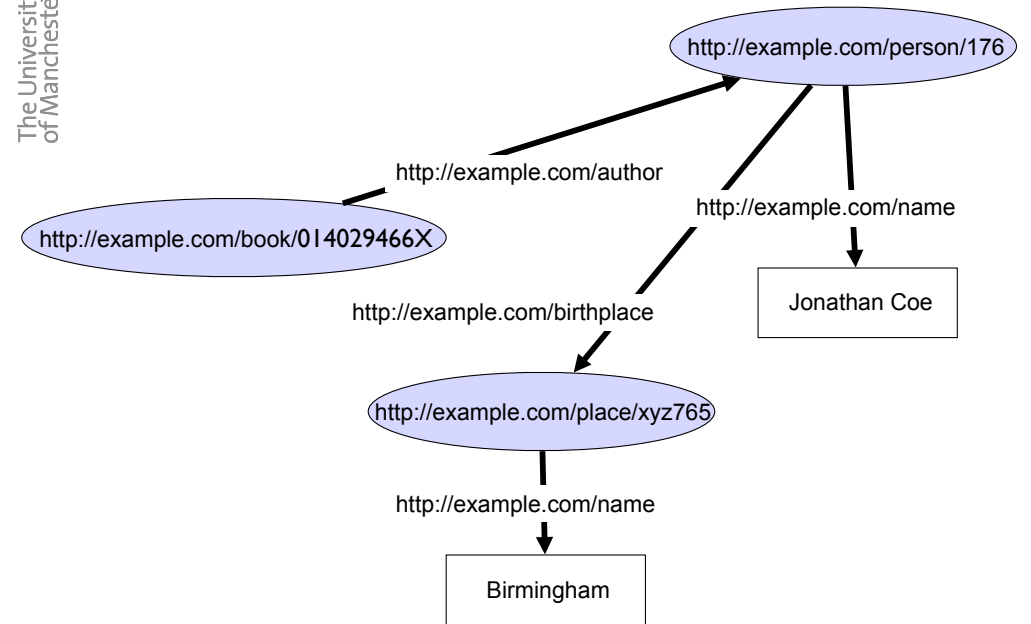
## URIs and naming

- URIs identify the things we are describing.
- If two people create data using the same URI, the assumption is that they are describing the same thing.
- Merging/integrating data then becomes easy
  - Although introduces issues of URI control.

## Graph Merging



## Graph Merging





## URIs are active

- URIs can be more than just names -- they can be dereferenced, and information can be retrieved.
- In particular, we can lookup the URIs in a graph and potentially retrieve more information about the URI.
- “Follow your nose” navigation
- Information should be returned in appropriate, machine readable formats (e.g. another graph)

## Linked Data Principles

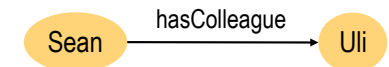
1. Use URIs as names for things
  2. Use http URIs so that those names can be dereferenced.
  3. When a URI is looked up, provide useful information
  4. Include statements that link to other URIs so that more information can be discovered.
- Common infrastructure facilitates construction of applications.
    - Largely browsers up to now....
  - Other guidelines relating to connecting documents with the data that describes them.
    - Use of content negotiation to supply “appropriate” representations
    - Use of microformats/RDFa to publish data

# RDF

- RDF stands for Resource Description Framework
- It is a W3C Recommendation
  - <http://www.w3.org/RDF>
- RDF is a graphical formalism ( + concrete syntax)
  - for representing metadata
  - for describing the semantics of information in a machine-accessible way
- Provides a simple data model based on triples.
- Allows us to represent relationships between things.

# The RDF Data Model

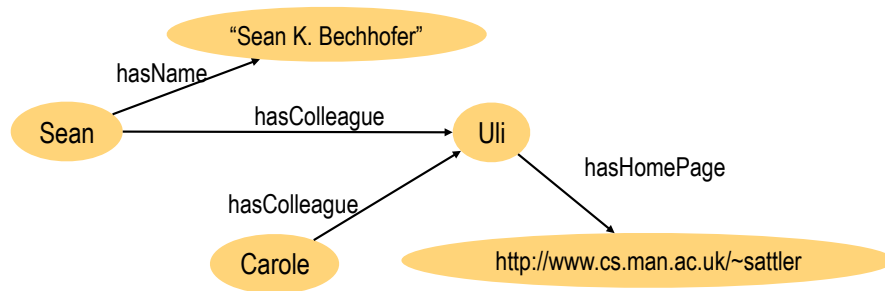
- Statements are <subject, predicate, object> triples:
  - <Sean, hasColleague, Uli>
- Can be represented as a graph:



- Statements describe properties of resources
- A resource is any object that can be pointed to by a URI
- Properties themselves are also resources (URIs)

## Linking Statements

- The subject of one statement can be the object of another
- Such collections of statements form a directed, labeled graph



- Note that the object of a triple can also be a “literal” (a string)

## RDF Syntax

- RDF has an XML syntax that has a specific meaning:
- Every **Description** element describes a resource
- Every attribute or nested element inside a **Description** is a **property** of that **Resource**
- We can refer to resources by URIs

```

<Description about="some.uri/person/sean_bechhofer">
  <hasColleague resource="some.uri/person/uli_sattler"/>
  <hasName rdf:datatype="xsd:string">Sean K. Bechhofer</hasName>
</Description>
<Description about="some.uri/person/uli_sattler">
  <o:hasHomePage>http://www.cs.mam.ac.uk/~sattler</o:hasHomePage>
</Description>
<Description about="some.uri/person/carole_goble">
  <o:hasColleague resource="some.uri/person/uli_sattler"/>
</Description>
  
```

## What does RDF give us?

- A mechanism for **annotating** data and resources.
- Single (simple) data model.
- Syntactic consistency between names (URIs).
- Low level **integration** of data.
- The Linked Data/Web of Data approach.

## Querying RDF: SPARQL

- RDF provides us with a way of representing information as a graph
- SPARQL allows us to query this information  
<http://www.w3.org/TR/sparql11-overview/>
- Provides a query language and the description of a protocol for interacting with SPARQL “endpoints” via HTTP

```
PREFIX etree:<http://etree.linkedmusic.org/vocab/>
PREFIX mo:<http://purl.org/ontology/mo/>
PREFIX event:<http://purl.org/NET/c4dm/event.owl#>
PREFIX skos:<http://www.w3.org/2004/02/skos/core#>
PREFIX timeline:<http://purl.org/NET/c4dm/timeline.owl#>
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>

SELECT DISTINCT ?artist WHERE
{
  ?art rdf:type mo:MusicArtist.
  ?art skos:prefLabel ?artist.
}
```

## RDF(S): RDF Schema

- RDF gives a formalism for meta data annotation, and a way to write it down in XML, but it doesn't give any special meaning to vocabulary such as `subClassOf` or `type`
  - Interpretation is an **arbitrary** binary relation
- RDF Schema extends RDF with a **schema vocabulary** that allows you to define basic vocabulary terms and the relations between those terms
  - `Class`, `Property`
  - `type`, `subClassOf`
  - `range`, `domain`

## RDF(S)

- These terms are the RDF Schema building blocks (constructors) used to create vocabularies:
  - `<Person, type, Class>`
  - `<hasColleague, type, Property>`
  - `<Professor, subClassOf, Person>`
  - `<Carole, type, Professor>`
  - `<hasColleague, range, Person>`
  - `<hasColleague, domain, Person>`
- Semantics gives “extra meaning” to particular RDF predicates and resources
  - specifies how terms should be interpreted

## What does RDF(S) give us?

- Ability to use simple schema/vocabularies when describing our resources.
- **Consistent** vocabulary use and **sharing**.
- Basic **inference**
  
- Note that RDF is a data model. There are many ways of serialising this data:
  - RDF/XML
  - Turtle
  - N3
  - json-ld

## Problems with RDF(S)

- RDF(S) is **too weak** to describe resources in sufficient detail
  - No **localised range and domain** constraints
    - Can't say that the range of **hasChild** is **Person** when applied to **Persons** and **Elephant** when applied to **Elephants**
  - No **existence/cardinality** constraints
    - Can't say that all *instances* of **Person** have a **mother** that is also a **Person**, or that **Persons** have exactly 2 **parents**
  - No **transitive, inverse** or **symmetrical** properties
    - Can't say that **isPartOf** is a **transitive** property, that **hasPart** is the **inverse** of **isPartOf** or that **touches** is **symmetrical**
- Difficult to provide **reasoning support**
  - No "native" reasoners for non-standard semantics
  - May be possible to reason via FO axiomatisation

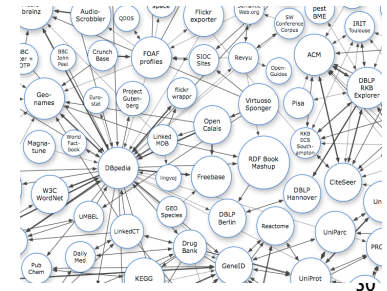
## OWL

- OWL: Web Ontology Language
- Extends existing Web standards
  - Such as XML, RDF, RDFS
- Is (hopefully) easy to understand and use
  - Based on familiar KR idioms
- Of “adequate” expressive power
- Formally specified
  - Possible to provide automated reasoning support
- But you already know all this...



## Linked Data Benefits

- Separation of data from formatting and presentational aspects
- Self-describing data. Applications encountering unfamiliar vocabularies can dereference and access definitions
- Simplified data access via HTTP and RDF
  - Heterogeneity of Web APIs
- Open
  - Applications not implemented against fixed set of data sources.



## RDF/Linked/Open Data

- dbpedia
  - RDFised version of wikipedia
  - Scraping structured information from info-boxes.
  - Quality?
- Government Data
  - <https://data.gov.uk/organogram/cabinet-office>
- Open Data Institute
- BBC
- GeoNames
  - Geographical data
  - Lat/long, postal codes etc.
- LCSH
  - SKOS

37

## dbpedia

The screenshot shows a web browser displaying the dbpedia.org page for Manchester. The page title is "About: Manchester" and it is identified as an entity of type "City status in the United Kingdom". A table lists properties and their values:

Property	Value
dbpedia-owl:PopulatedPlace/areaTotal	• 115.65
dbpedia-owl:PopulatedPlace/populationDensity	4313.0
dbpedia-owl:abstract	Manchester is a city and metropolitan borough in Greater Manchester, England. In the 2011 UK Census, its population was recorded as 498,800. Manchester lies within one of the UK's largest metropolitan areas, the metropolitan county of Greater Manchester, which has an estimated population of 2.6 million. The demonym of Manchester is Mancunian and symbols include the Manchester bee.

Below the table, there is a detailed text description in English and German. The English text describes Manchester's location, population, and history, mentioning its role as the first industrialized city and its status as a major financial and cultural center. The German text provides a similar overview in German.

32



MANCHESTER 1824 wikidata

Manchester - Wikidata

Secure https://www.wikidata.org/wiki/Q18125

English Not logged in Talk Contributions Create account Log in

Read View history Search Wikidata

Item Discussion

Manchester (Q18125)

major city in Greater Manchester, England, UK

Wikipedia (125 entries) edit

Language	Label	Description	Also known as
English	Manchester	major city in Greater Manchester, England, UK	
Portuguese	Manchester	No description defined	
Zulu	No label defined	No description defined	
Polish	Manchester	miało w Anglii	

All entered languages

Statements

instance of city 0 references edit add reference add value

part of Greater Manchester 1 reference edit add value

image Manchester Montage.png 1 reference edit add value

inception 1301 Gregorian 1 reference edit add value

country United Kingdom start time 6 December 1922 Gregorian edit

MANCHESTER 1824 BBC

BBC Music - Mogwai

www.bbc.co.uk/music/artists/d700b3f5-45af-4d02-95ed-57d301bda93d

Sign in News Sport Weather iPlayer TV Radio More Search the BBC

MUSIC

HOME SHOWCASE REVIEWS GENRES

Mogwai

Formed 1996.

PLAYED MOST ON BBC RADIO 6music

Share This Page 1 so far

Share facebook twitter

BBC Music Showcase

BBC MUSIC SHOWCASE

Watch and listen to exclusive music clips

Latest Tracks Played On The BBC

Whita Noise  
BBC 6 MUSIC | GIDEON COE 24/11/2011

Hunted By A Freak  
BBC 6 MUSIC | CHRIS HAWKINS 12/11/2011

Mexican Grand Prix  
BBC 6 MUSIC | CHRIS HAWKINS AND YOU LEFT A GIG EARLY... WHY?

Hound Of Winter  
BBC 6 MUSIC | GIDEON COE 24/08/2011

San Pedro  
BBC RADIO 1 | MICK GRIMSHAW WEDNESDAY: MILES KANE SPENDS THE NIGHT WITH EDITH

Biography

Mogwai are a Scottish post-rock band, formed in 1995 in Glasgow. The band consists of Stuart Braithwaite (guitar, vocals), John Cummings (guitar, vocals).

**THINGS** BETA

**Mogwai**  
Scottish post-rock band  
[http://www.bbc.co.uk/things/b46ff538-c31e-49bf-ae1-0e0a98a3b379#id](http://www.bbc.co.uk/things/b46ff538-c31e-49bf-ae1-0e0a98a3b379)

[wikipedia image](#) Mogwai () are a Scottish post-rock band, formed in 1995 in Glasgow. The band consists of Stuart Braithwaite (guitar, vocals), Barry Burns (guitar, piano, synthesizer, vocals), Dominic Aitchison (bass guitar), and Martin Bulloch (drums). The band typically compose lengthy guitar-based instrumental pieces that feature dynamic contrast, melodic bass guitar lines, and heavy use of distortion and effects. The band were for several years signed to Glasgow label Chemical Underground, and have been distributed by different labels such as Matador in the US and Play It Again Sam in the UK, but now use their own label Rock Action Records in the UK, and Sub Pop in North America. The band were frequently

**Properties**

**rdfs:type**  
 core:Thing  
 isgoin:TagConcept

**core:disambiguationHint**  
 Scottish post-rock band

**core:label**  
 en-gb: Mogwai

**core:preferredLabel**  
 Mogwai

**core:primaryTopicOf**  
<http://www.bbc.co.uk/music/artists/d700b3f5-45af-4d02-95ed-57d301bda93e>  
<http://www.imdb.com/name/nm1912841/>  
<http://www.mogwai.co.uk/>  
<https://en.wikipedia.org/wiki/Mogwai>

**core:sameAs**  
 ddpedia:Mogwai  
<http://musicbrainz.org/artist/d700b3f5-45af-4d02-95ed-57d301bda93e#>  
<http://www.wikidata.org/entity/Q645980>

[View Turtle \(.ttl\)](#)

35

**LIBRARY OF CONGRESS** ASK A LIBRARIAN DIGITAL COLLECTIONS LIBRARY CATALOGS [GO](#) [DONOR](#)

The Library of Congress > Authorities & Vocabularies > LC Subject Headings

**Furniture**

From Library of Congress Subject Headings

**Details** Visualization Suggest Terminology

**Furniture**

**URT(s)**  
 > <http://id.loc.gov/authorities/subjects/sh85052522>  
 > <info:lc/authorities/sh85052522>  
 > <http://id.loc.gov/authorities/sh85052522#concept>

**Instance Of**  
 > [MADS/RDF Topic](#)  
 > [MADS/RDF Authority](#)  
 > [SKOS Concept c4](#)

**Scheme Membership(s)**  
 > [Library of Congress Subject Headings](#)

**Collection Membership(s)**  
 > [LCSH Collection - Authorized Headings](#)  
 > [LCSH Collection - General Collection](#)  
 > [LCSH Collection - Term Permitted to be Indirectly Subdivided Geographically](#)

**Variants**  
 > [Wood furniture](#)  
 > [Wooden furniture](#)

**Broader Terms**  
 > [Decoration and ornament](#)  
 > [Decorative arts](#)  
 > [House furnishings](#)

**Narrower Terms**  
 > [Architect-designed furniture](#)  
 > [Artist-designed furniture](#)  
 > [Bamboo furniture](#)  
 > [Beds \(furniture\)](#)  
 > [Bedroom furniture](#)

36

## Modelling

Live Music Archive Linked Data

This server provides access to Linked Data that describes the audio held in the Internet Archive's *Live Music Archive* (also sometimes known as "etree"). The metadata from etree has been converted to RDF and is exposed through a SPARQL endpoint along with browsable pages.

The dataset contains information describing over 100,000 performances by 4,000 artists including 1,600,000 individual tracks, each of which may be available in a number of formats.

**Sample Resources**

- The artist *Mogwai*: <http://etree.linkedmusic.org/artist/422aacc0-4aac-012f-19e9-00254bd44c28>
  - Mogwai Live at the Forum on 16-10-1999: <http://etree.linkedmusic.org/performance/mogwai1999-10-16.flac16>
    - The Forum venue (as played by Mogwai on 16-10-1999): <http://etree.linkedmusic.org/venue/mogwai1999-10-16.flac16>
- The artist *Calexico*: <http://etree.linkedmusic.org/artist/4229f010-4aac-012f-19e9-00254bd44c28>
  - Calexico Live at the Serenadenhof on 11-08-2009: <http://etree.linkedmusic.org/performance/calexico2009-08-11.flac>
- The artist *Andrew Bird*: <http://etree.linkedmusic.org/artist/422c57d0-4aac-012f-19e9-00254bd44c28>
- VoID metadata describing the dataset: <http://etree.linkedmusic.org/etree>

**Vocabularies**

The dataset makes use of a number of vocabularies including:

- The Music Ontology: <http://purl.org/ontology/mo/>
- The Similarity Ontology: <http://purl.org/ontology/similarity/>
- The Event Ontology: <http://purl.org/NET/oddm/event.owl>

**Links**

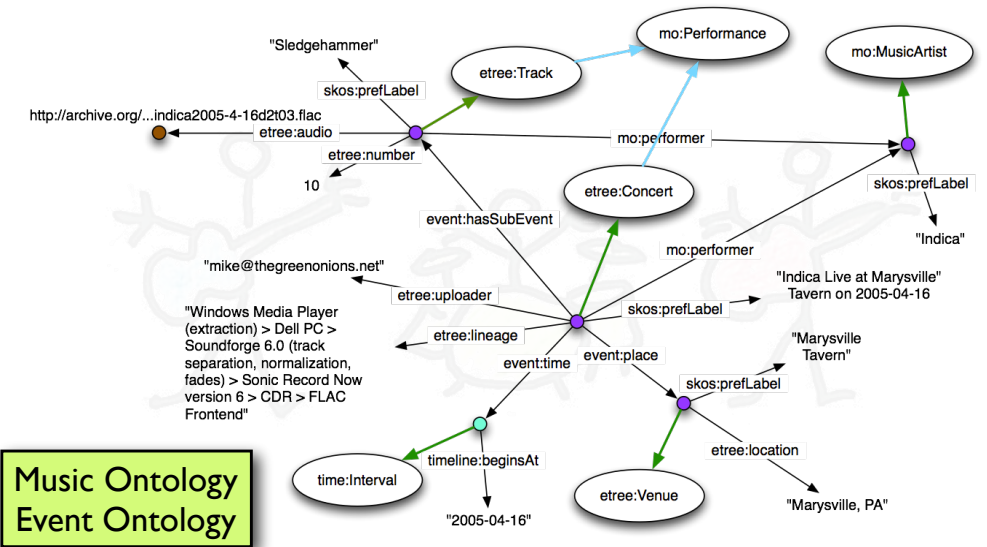
Artists are linked to entries in *MusicBrainz*, while locations of performances are linked to entries in *Geonames* and *last.fm* using a number of *methods*. Any such alignments are represented explicitly as *similarities*.

**SPARQL Endpoint**

A SPARQL endpoint for the data is available at <http://etree.linkedmusic.org/sparql>. Some sample queries are shown below. Note that as this is currently an experimental service, query timeouts are limited.

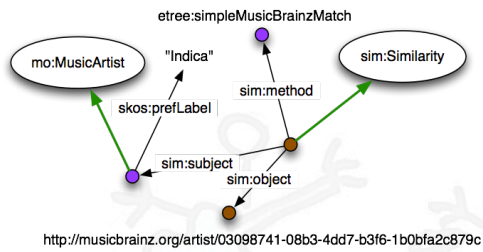
- All artists in the collection.
- Artists in the collection with mappings to MusicBrainz.
- Artists and the locations of recorded performances.

37

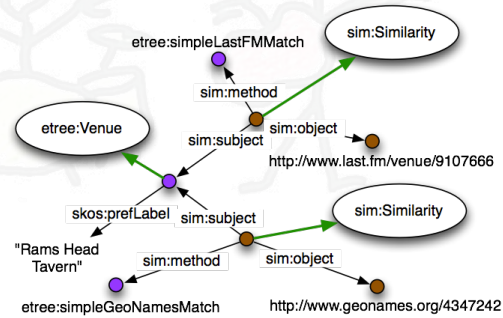


38

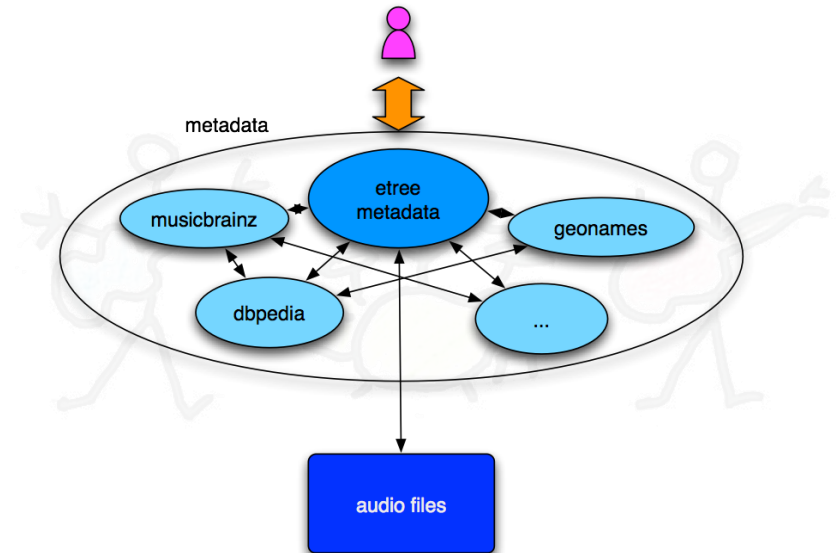
# Similarities



Similarity Ontology



# Big Picture



## LD in Use

- Five Stars of Open Linked Data
  - <http://inkdroid.org/journal/2010/06/04/the-5-stars-of-open-linked-data/>
  - ★ Make data available
  - ★ Make it available as structured data
  - ★ Use non-proprietary formats
  - ★ Use URLs to identify things
  - ★ Link your data
- Costs and Benefits
  - <http://lab.linkeddata.deri.ie/2010/star-scheme-by-example/>

## Issues with Linked Data

- Identity and co-reference
  - Management of identities
  - How do we handle the fact that different URIs may be used to refer to the same things?
  - Use of owl:sameAs may be too strong (can result in all information, including annotations, metadata etc.) being merged.
- Visualisation
  - Big Fat Graph
- Versioning
  - Version information in URLs?
  - Versioning at architectural level (Memento)
  - How does versioning play with a “follow your nose paradigm”?
- Querying
  - Distributed query across data sets
  - LD applications tend to use an “extract, transform, load” approach.

## Issues with Linked Data

- A focus on *data*
  - Vocabularies used to facilitate integration
  - Little *deep* semantics.
  - “Big O vs little o”
    - Role of SKOS and RDF(S)
- Scalability
- A focus on mechanisms for *data publication* rather than *consumption*
  - Lots of work on “recipes”, mangling relation sources into RDF etc.
  - What do you actually *do* with the stuff?
  - End user applications
    - Smart cities
  - *Build it and they will come....???*

## Do you need them all?

1. Use URIs as names for things
2. Use http URIs so that those names can be dereferenced.
3. When a URI is looked up, provide useful information
4. Include statements that link to other URIs so that more information can be discovered.