

Providing **Linked Data**

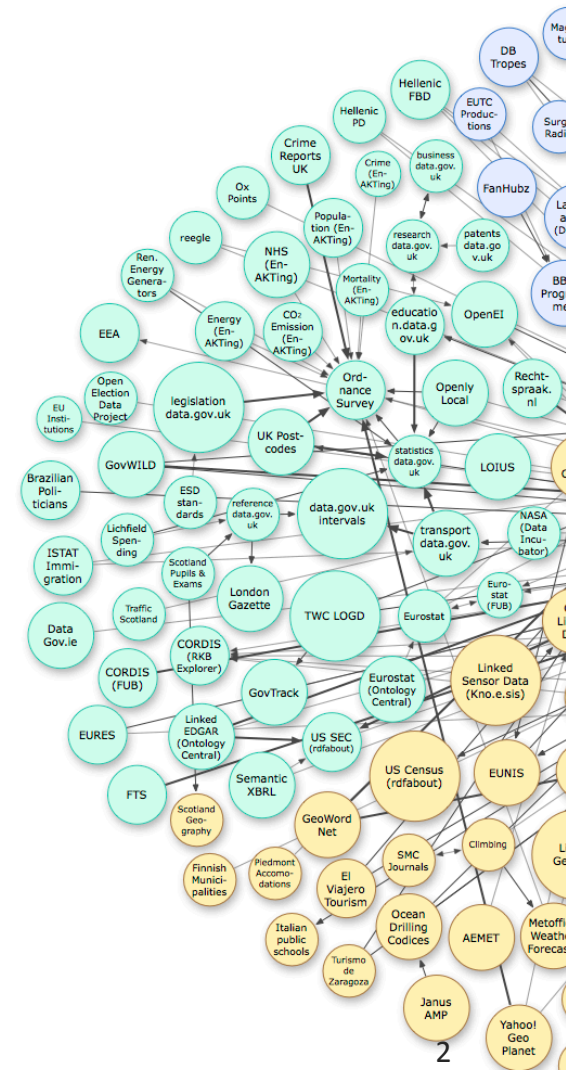
Presented by:
Maribel Acosta



The Open University



LINKED DATA LIFECYCLE



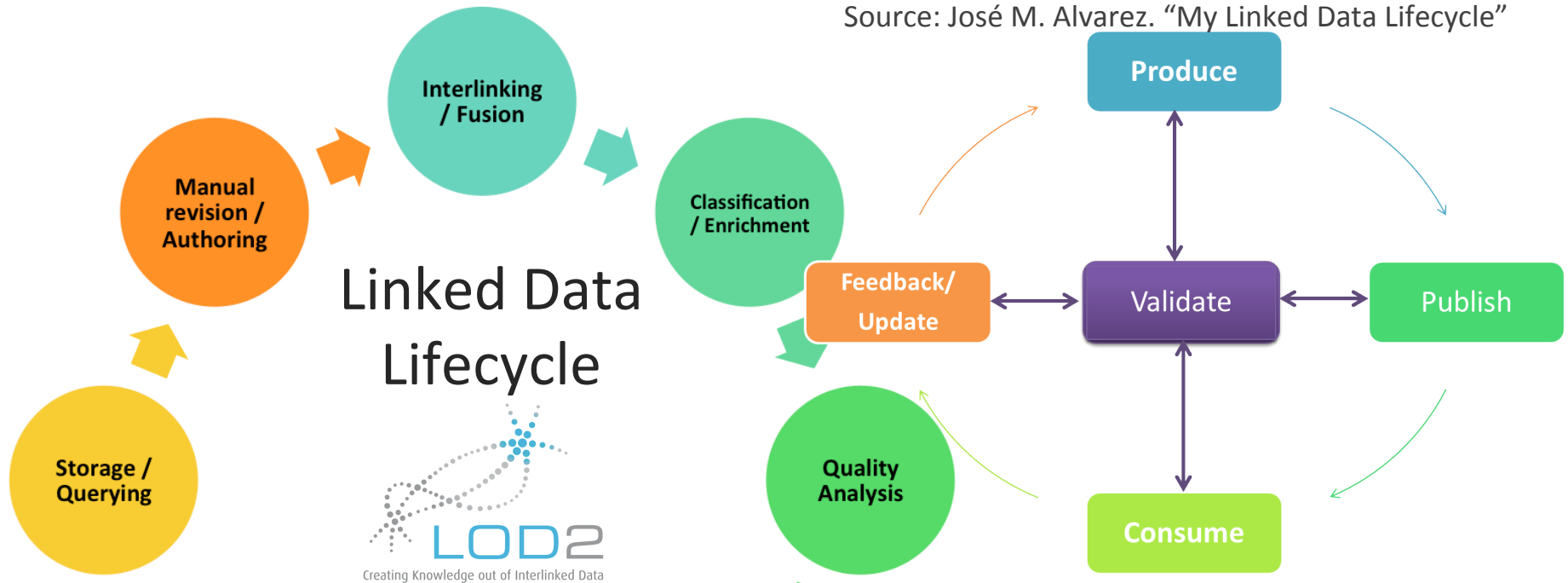
Linked Data Principles

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

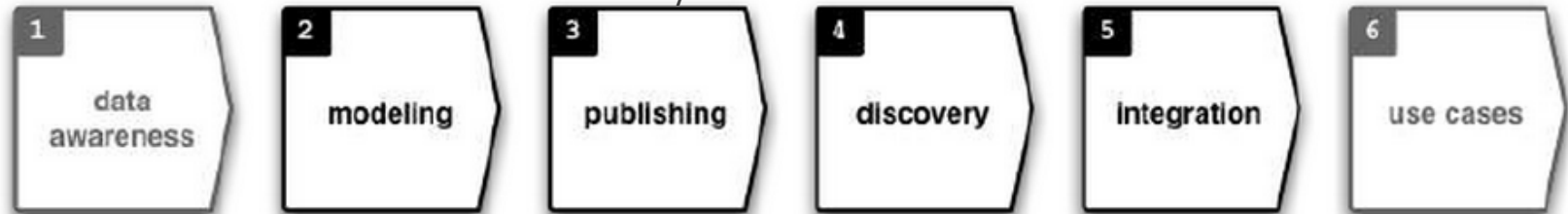
Linked Data Lifecycle

Source: Sören Auer. "The Semantic Data Web" (slides)

Source: José M. Alvarez. "My Linked Data Lifecycle"



Source: Michael Hausenblas. "Linked Data lifecycle"



Core Tasks for Providing Linked Data

Based on the proposed LD lifecycles and the LD principles, we can identify 3 main tasks for providing LD:

- ① **Creating:** includes data extraction, creation of HTTP URIs, and vocabulary selection. (LD principles 1 & 2)
- ② **Interlinking:** involves the creation of (RDF) links to external data sets. (LD principle 4)
- ③ **Publishing:** consists of creating the metadata and making the data set accessible. (LD principle 3)

Agenda

- 1. Creating Linked Data**
- 2. Interlinking Linked Data**
- 3. Publishing Linked Data**
- 4. Linked Data publishing checklist**

CREATING LINKED DATA

Linked Data principles 1 & 2

Extracting the Data

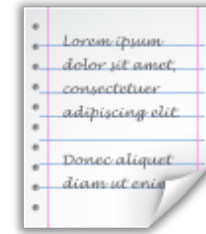
- The data of interest may be stored in a wide range of **formats**:



Spreadsheets
or tabular data



Databases



Text

- Several tools support the process of **mining data** from different repositories, for example:



Naming Things: URIs

- All the *things* or distinct entities within the data must be **named**
 - According to the Linked Data principles, the standard mechanism to name entities is the **URI**
 - Designing **Cool URIs**:
 - Leave out information about the data regarding to: author, technologies, status, access mechanisms, ...
 - **Simplicity**: short, mnemonic URIs
 - **Stability**: maintain the URIs as long as possible
 - **Manageability**: issue the URIs in a way that you can manage
- Source: <http://www.w3.org/TR/cooluris/>

Selecting Vocabularies

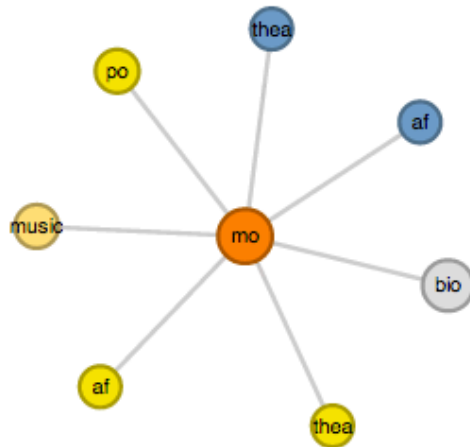
- Vocabularies model the **concepts** and the **relationship** between them in a knowledge domain
- Terms from well-known vocabularies should be **reused** wherever possible
- **New terms** should be define only if you can not find required terms in existing vocabularies
- A large number of **vocabularies** in **RDF** are openly available, e.g., **Linked Open Vocabularies (LOV)**

Selecting Vocabularies (3)

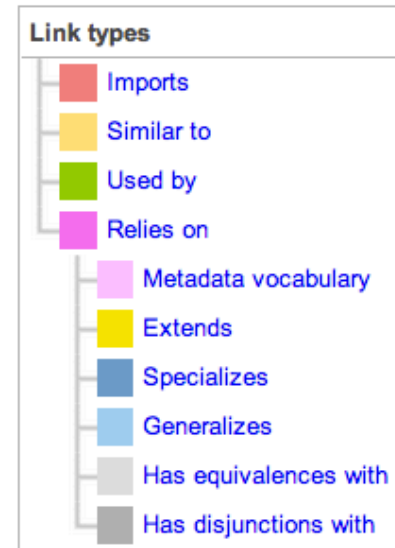
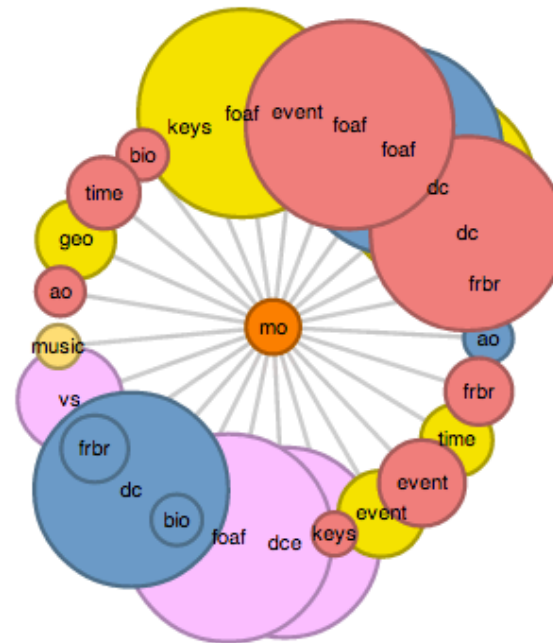
Linked Open Vocabularies: Analyzing MusicOntology

Vocabulary links:

Vocabularies referencing "mo" (7)



Vocabularies referenced by "mo" (25)



Source: http://lov.okfn.org/dataset/lov/details/vocabulary_mo.html



Selecting Vocabularies (4)

Other lists of well-known vocabularies are maintained by:

- W3C SWEO Linking Open Data community project
<http://www.w3.org/wiki/TaskForces/CommunityProjects/LinkingOpenData/CommonVocabularies>
- Library Linked Data Incubator Group: Vocabularies in the library domain
<http://www.w3.org/2005/Incubator/lld/XGR-lld-vocabdataset-20111025>

Creating Linked Data (Quick Recipe)

Steps:

- **Source data** (database, text, ...)
- Names for things (instances): **URIs**
- **Vocabularies**
 - A new concept should be created only when it can not be found in existing vocabularies



RDF dataset

INTERLINKING LINKED DATA

Linked Data principle 4

Interlinking Data Sets

- It's one of the Linked Data principles!

4. Include links to other URIs, so that users can **discover** more things.

- Involves the creation of RDF links between two different RDF data sets:
 - Links at instance level (rdfs:seeAlso, owl:sameAs)
 - Links at schema level (RDFS subclass/subproperty, OWL equivalent class/property, **SKOS mapping properties**)
- Appropriate links are detected via **link discovery**

Challenges for link discovery

- Linked Data sets are **heterogeneous** in terms of vocabularies, formats and data representation
- Large range of knowledge **domains**
- **Scalability**: LD is composed of a large number of data sets and RDF triples, hence it is not possible to compare every possible entity pair

Source: Robert Isele. “LOD2 Webinar Series: Silk”



Challenges for link discovery

- An instance of link discovery is the **entity resolution** problem: *deciding whether two entities correspond to same object in the real world*
 - **Name ambiguities:** typos, misspellings, different languages, homonyms
 - **Structural ambiguities:** same concepts/entities with different structures. Requires the application of ontology and schema matching techniques

Interlinking Data Sets (4)

RDF data sets
can be interlinked:

Manually

- Involves the manual exploration of LD data sets and their RDF resources to identify linking targets
- May not be feasible when the number of entities within the data set is very large

(Semi-)Automatically

- Using tools that perform link discovery based on linkage rules, for example: Silk, Limes and xCurator

owl:sameAs & rdfs:seeAlso (1)

- **owl:sameAs**
 - Creates links between individuals
 - States that two URIs refer to the same individuals
- **rdfs:seeAlso**
 - States that a resource may provide additional information about the subject resource
- Links in **MusicBrainz**:
 - owl:sameAs is used for music artists
 - rdfs:seeAlso is used for albums



owl:sameAs & rdfs:seeAlso (2)

Examples of owl:sameAs and rdfs:seeAlso

```
@prefix owl: <http://www.w3.org/2002/07/owl#>
```

```
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
```

```
@prefix dbpedia: <http://dbpedia.org/resource/>
```

```
<http://musicbrainz.org/artist/b10bbbfc-cf9e-42e0-be17-e2c3e1d2600d>  
owl:sameAs dbpedia:The_Beatles.
```

```
<http://musicbrainz.org/release/f8bf9e45-a1d1-42ab-8f3c-637b1762e5f7>  
rdfs:seeAlso dbpedia:All_you_need_is_love_(album).
```

SKOS

- Simple Knowledge Organization System
 - <http://www.w3.org/TR/skos-reference/>
- Data model for knowledge organization systems (thesauri, classification scheme, taxonomies)
- SKOS data is expressed as RDF triples
- Allows the creation of RDF links between different data sets with the usage of **mapping properties**

SKOS: Mapping Properties

These properties are used to link SKOS concepts (particularly instances) in different schemas:

- **skos:closeMatch**: links two concepts that are sufficiently similar (sometimes can be used interchangeably)
- **skos:exactMatch**: indicates that the two concepts can be used interchangeably.
 - Axiom: It is a **transitive** property
- **skos:relatedMatch**: states an associative mapping link between two concepts

SKOS: Mapping Properties (2)

Example of SKOS exact match

```
@prefix skos: <http://www.w3.org/2004/02/skos/core#>  
@prefix mo: <http://purl.org/ontology/mo/>  
@prefix dbpedia-ont: <http://dbpedia.org/ontology/>  
@prefix schema: <http://schema.org/>
```

```
mo:MusicArtist skos:exactMatch dbpedia-ont:MusicalArtist.
```

```
mo:MusicGroup skos:exactMatch dbpedia-ont:Band.
```

```
mo:MusicGroup skos:exactMatch schema:MusicGroup.
```


SKOS: Mapping Properties (3)

Example of SKOS close match

```
@prefix skos: <http://www.w3.org/2004/02/skos/core#>  
@prefix mo: <http://purl.org/ontology/mo/>  
@prefix dbpedia-ont: <http://dbpedia.org/ontology/>  
@prefix schema: <http://schema.org/>
```

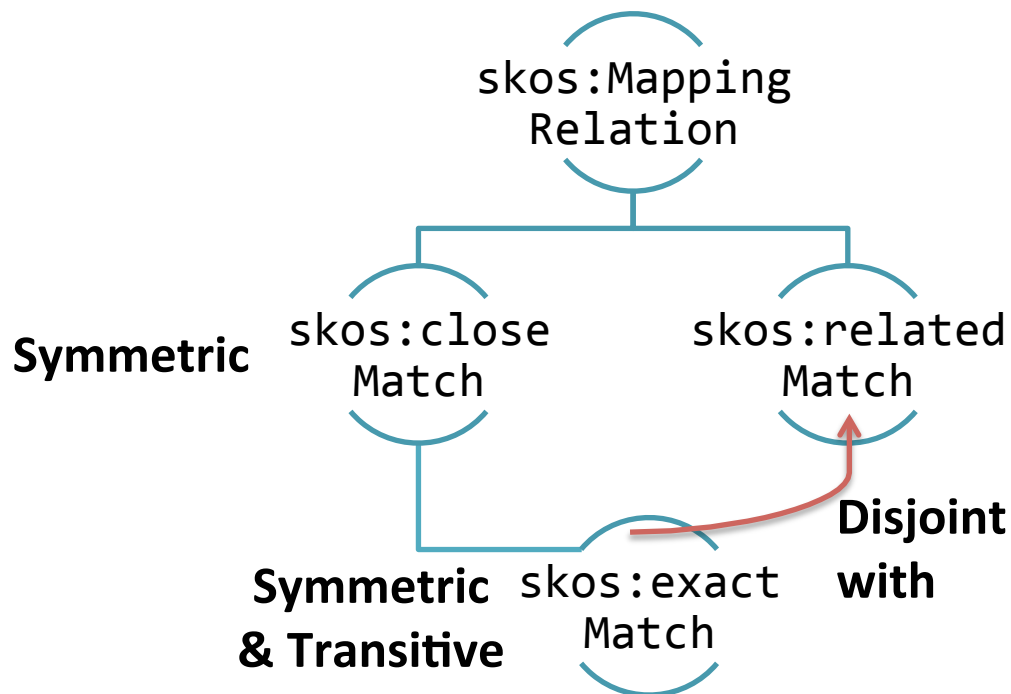
```
mo:SignalGroup skos:closeMatch schema:MusicAlbum.
```

```
mo:SignalGroup skos:closeMatch dbpedia-ont:Album.
```

Integrity conditions

- Guarantee consistency and avoid contradictions in the relationships between SKOS concepts

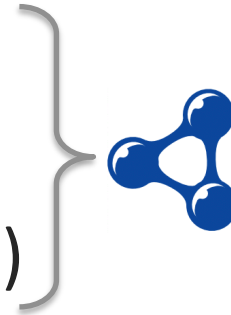
Partial Mapping Relation diagram with integrity conditions



Interlinking Linked Data (Quick Recipe)

Steps:

- Create **links** at instance level
- (Create **links** at schema level)



Interlinked RDF
dataset

PUBLISHING LINKED DATA

Linked Data principle 3

Publishing Linked Data

Once the RDF data set has been created and interlinked, the publishing process involves the following tasks:

1. **Metadata** creation for describing the data set
2. Making the data set **accessible**
3. Exposing the data set in Linked Data **repositories**
4. **Validating** the data set

Describing RDF Data Sets

- Consists of providing (machine-readable) **metadata** of RDF data sets which can be processed by engines
- This information allows for:
 - Efficient and effective search of data sets
 - Selection of appropriate data sets (for consumption or interlinking)
 - Get general statistics of the data sets

Describing RDF Data Sets (2)

- The common language for describing RDF data sets is **VOID (Vocabulary of Interlinked Data sets)**
- Defines an RDF data set with the predicate `void:Dataset`
- Covers 4 types of metadata:
 - General metadata
 - Structural metadata
 - Descriptions of linksets
 - Access metadata

VoID: General Metadata

- General metadata is used by users to identify appropriate data sets.
- Specifies information about description of the data set, contact person/organization, the license of the data set, data subject and some technical features.
- VoID (re)uses predicates from the Dublin Core Metadata¹ and FOAF² vocabularies.

¹ <http://dublincore.org/documents/2010/10/11/dcmi-terms/>

² <http://xmlns.com/foaf/spec/>

General Information

Source: <http://www.w3.org/TR/void/#metadata>

Contains information about the creation of the data set

Predicate	Range	Description
<code>dcterms:title</code>	Literal	Name of the data set.
<code>dcterms:description</code>	Literal	Description of the data set.
<code>dcterms:source</code>	RDF resource	Source from which the data set was derived.
<code>dcterms:creator</code>	RDF resource	Primarily responsible of creating the data set.
<code>dcterms:date</code>	<code>xsd:date</code>	Time associated with an event in the life-cycle of the resource.
<code>dcterms:created</code>	<code>xsd:date</code>	Date of creation of the data set.
<code>dcterms:issued</code>	<code>xsd:date</code>	Date of publication of the data set.
<code>dcterms:modified</code>	<code>xsd:date</code>	Date on which the data set was changed.
<code>foaf:homepage</code>	Literal	Name of the data set.
<code>dcterms:publisher</code>	RDF resource	Entity responsible for making the data set available.
<code>dcterms:contributor</code>	RDF resource	Entity responsible for making contributions to the data set.

Other Information

- **License of the data set:** specifies the usage conditions of the data. The license can be pointed with the property `dcterms:license`
- **Category of the data set:** to specify the topics or domains covered by the data set, the property `dcterms:subject` can be used
- **Technical features:** the property `void:feature` can be used to express technical properties of the data (e.g. RDF serialization formats)

VoID: Structural Metadata

- Provides high-level information about the **internal structure** of the data set
- This metadata is useful when exploring or **querying** the data set
- Includes information about resources, vocabularies used in the data set, statistics and examples of resources in the data set

Information about resources

- **Example resources:** allow users to get an impression of the kind of resources included in the data set. Examples can be shown with the property `void:exampleResource`

```
:MusicBrainz a void:Dataset;  
  void:exampleResource  
    <http://musicbrainz.org/artist/b10bbbfc-cf9e-42e0-be17-e2c3e1d2600d> .
```

- **Pattern for resource URIs:** the `void:uriSpace` property can be used to state that all the entity URIs in a data set start with a given string

```
:MusicBrainz a void:Dataset;  
  void:uriSpace "http://musicbrainz.org/" .
```

Vocabularies used in the data set

- The `void:vocabulary` property identifies the vocabulary or ontology that is used in a data set
- Typically, only the most relevant vocabularies are listed

```
:MusicBrainz a void:Dataset;  
    void:vocabulary <http://purl.org/ontology/mo/> .
```

- This property can only be used for entire vocabularies. It **cannot** be used to express that a subset of the vocabulary occurs in the data set.

Statistics about a data set

Express numeric statistics about a data set:

Predicate	Range	Description
<code>void:triples</code>	Number	Total number of triples contained in the data set.
<code>void:entities</code>	Number	Total number of entities that are described in the data set. An entity must have a URI, and match the <code>void:uriRegexPattern</code>
<code>void:classes</code>	Number	Total number of distinct classes in the data set.
<code>void:properties</code>	Number	Total number of distinct properties in the data set.
<code>void:distinctSubjects</code>	Number	Total number of distinct subjects in the data set.
<code>void:distinctObjects</code>	Number	Total number of distinct objects in the data set.
<code>void:documents</code>	Number	Total number of documents, in case that the data set is published as a set of individual documents.

Source: <http://www.w3.org/TR/void/#metadata>



Partitioned data sets

- The `void:subset` property provides description of *parts* of a data set

```
:MusicBrainz a void:Dataset;  
  void:subset :MusicBrainzGroupMembers .
```

- Data sets can be partitioned based on **classes** or **properties**:
 - `void:classPartition` contains only instances of a particular class
 - `void:propertyPartition` contains only triples with a particular predicate

```
:MusicBrainzGroupMembers a void:Dataset;  
  void:classPartition [ void:class mo:MusicArtist .] ;  
  void:propertyPartition [ void:property mo:member .] .
```

VoID: Describing Linksets

- **Linkset:** collection of RDF links between two RDF data sets

```
@PREFIX void:<http://rdfs.org/ns/void#>  
@PREFIX owl:<http://www.w3.org/2002/07/owl#>
```

```
:DS1 a void:Dataset .  
:DS2 a void:Dataset .  
:DS1 void:subset :LS1 .  
:LS1 a void:Linkset;  
void:linkPredicate  
  owl:sameAs;  
void:target :DS1, :DS2 .
```

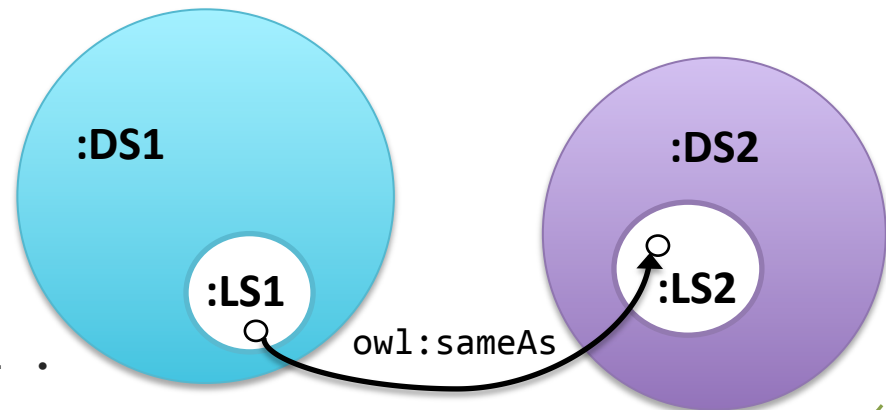


Image based on <http://semanticweb.org/wiki/File:Void-linkset-conceptual.png>

VoID: Describing Linksets (2)

Example

```
@PREFIX void:<http://rdfs.org/ns/void#>
@PREFIX owl:<http://www.w3.org/2002/07/owl#>

:MusicBrainz a void:Dataset .
:DBpedia a void:Dataset .

:MusicBrainz void:classPartition :MBArtists .
:MBArtists void:class mo:MusicArtist .

:MBArtists a void:Linkset;
  void:linkPredicate owl:sameAs;
  void:target :MusicBrainz, :DBpedia .
```

VoID: Access Metadata

The access metadata describes the methods of accessing the actual RDF data set

Method	Predicate	Description
URI look up endpoint	<code>void:uriLookupEndpoint</code>	Specifies the URI of a service for accessing the data set (different from the SPARQL protocol)
Root resource	<code>void:rootResource</code>	URI of the top concepts (only for data sets structured as trees)
SPARQL endpoint	<code>void:sparqlEndpoint</code>	Provides access to the data set via the SPARQL protocol.*
RDF data dumps	<code>void:dataDump</code>	Specifies the location of the dump file. If the data set is split into multiple files, then several values of this property are provided.

* This assumes that the default graph of the SPARQL endpoint contains the data set. VoID cannot express that a data set is contained in a specific named graph. This can be specified with SPARQL 1.1. Service Description



Providing Access to the Data Set

The data set can be accessed via different mechanisms:


Dereferencing
HTTP URIs

RDFa

SPARQL
endpoint

RDF
dump

Dereferencing HTTP URIs

- Allows for easily **exploring** certain resources contained in the data set
- What to return for a URI?  CH 1
 - **Immediate description:** triples where the URI is the subject.
 - **Backlinks:** triples where the URI is the object.
 - **Related descriptions:** information of interest in typical usage scenarios.
 - **Metadata:** information as author and licensing information.
 - **Syntax:** RDF descriptions as RDF/XML and human-readable formats.
- Applications (e.g. LD browsers) render the retrieved information so it can be perceived by a user.

Source: *How to Publish Linked Data on The Web* - Chris Bizer, Richard Cyganiak, Tom Heath.

Dereferencing HTTP URIs (2)

Example: Dereferencing http://dbpedia.org/resource/The_Beatles

 dbpedia.org/page/The_Beatles

About: [The Beatles](#)

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



The Beatles were an English rock band formed in Liverpool in 1960 and one of the most commercially successful and critically acclaimed acts in the history of popular music. The group's best-known lineup consisted of John Lennon (rhythm guitar, vocals), Paul McCartney (bass guitar, vocals), George Harrison (lead guitar, vocals) and Ringo Starr (drums, vocals).

Property	Value
dbpedia-owl:abstract	<ul style="list-style-type: none">The Beatles foren un grup de música procedent de Liverpool, Anglaterra, format per John Lennon, Paul McCartney, George Harrison i Ringo Starr. Són un dels èxits i vendes de discos de la història de la música popular, i, a més, han aconseguit el reconeixement dels crítics. Les seves aportacions innovadores a la cultura van tenir una gran transcendència en els anys seixanta, que es perllongà d'una manera o altra en els anys posteriors. The Beatles són una de les bandes que existeixen avui en dia i aclamades per la crítica de la història de la música popular. The Beatles foren el grup amb més vendes de tot el segle XX. Al Regne Unit van vendre més de 214 milions de discos i singles i àlbums que van arribar al número u. El seu èxit es va repetir també en molts altres països: EMI assegurà que el 1985 el grup havia venut més milions de discos que els Beatles a principi dels anys 60, van ser els iniciadors del fenomen beat. Partint de la base rítmica del rock, les seves composicions foren una veritable troballa. Amb influències del rhythm and blues, i del blues progressiu, el seu estil va seguir una evolució constant. Des de mitjan dels anys 60, amb l'arribada de l'èxit decantar cap a un eclecticisme que incloïa una forta influència oriental i que va donar com a resultat algunes de les obres més significatives de la psicodèlia i del blues progressiu. Les seves cançons van reflectir els problemes d'un cert sector de la joventut de l'època, que pretenia restar al marge de l'anomenada societat de consum, i els aspectes més retrògrads de la societat occidental, alhora que reclamava nous valors tant estètics i artístics com espirituals i socials relacionats, per exemple, amb l'antimilitarisme. La seva dissolució oficial va tenir lloc l'any 1970. Els seus integrants van seguir les respectives carreres musicals en solitari, tot i que van ser molt productius. L'any 1995, amb motiu de l'edició del primer volum de la trilogia Beatles Anthology, va tenir lloc una reunió virtual del grup, amb l'ajuda d'unes cintes que enregistrades i que van ser la base de dues noves cançons: Free as a Bird i Real Love.The Beatles byla anglická rock'n'rollová kapela z Liverpoolu. Jejími členy byli John Lennon (zpěv, doprovodná kytara), Paul McCartney (zpěv, basová kytara), George Harrison (zpěv) a Ringo Starr (bicí, příležitostně zpěv). Patří mezi komerčně nejúspěšnější a kriticky nejuznávanější kapely v historii populární hudby. Hudba Beatles je jedním z nejvlivnějších žánrů 20. století. Když renomovaný hudební časopis Rolling Stone v roce 2003 sestavil žebříček nejlepších alb v dějinách populární hudby, Beatles se do první desítky umístili za Sgt. Pepper's Lonely Hearts Club Band. V roce 1995 se umístili na první místo v seznamu nejprodávanějších alb v historii. V roce 1995 RIAA, která je na základě prodeje singlů a alb v Americe označila jako nejvíce prodávanou kapelu v historii, nejprodávanějším hudebním uskupením v historii Velké Británie vydali Beatles více než 40 různých singlů, alb a desek, které se vyšplhaly na první místo hitparád. Tento komerční úspěch se opakoval v mnoha dalších zemích. V roce 1985 společnost EMI odhadla, že do roku 1985 prodali přes miliardu desek a kazet po celém světě. V roce 2004 umístil časopis Rolling Stone Beatles na první místo v seznamu nejvlivnějších kapel v historii. Podle časopisu pomohla jejich novátorská hudba a kulturní vliv charakterizovat 60. let a jejich vliv na populární kulturu je patrný ještě dnes. Beatles přinesli invazi britské hudby do Spojených států. Přestože měl jejich počáteční hudební styl kořeny v rock and rollu 50. let a domácím skiffllu, vyzkoušela kapela i prvky psychedelického rocku. Svým oblečením, styly a výroky určovali trendy své doby, zatímco jejich vzrůstající zájem o společenské otázky ovlivnil spíše generaci 60. let. The Beatles waren eine britische Rockband in den 1960er Jahren. Mit mehr als 600 Millionen – nach Schätzungen ihrer Plattenfirma EMI sogar mehr als eine Milliarde – verkauften sie zu den kommerziell erfolgreichsten Bands der Musikgeschichte. Die musikalischen Ursprünge der Band begannen Ende der 1950er Jahre, ihre erste große Durchbruch schaffte die Gruppe im Jahr 1963 mit der Single I Want to Hold Your Hand. Aufgrund ihres damals neuartigen Musikstils und ihrer Popularität entwickelten sich die Beatles schnell zu einer der populärsten Bands. Den Höhepunkt ihrer Karriere erreichten die Beatles zwischen 1964 und 1968, als sie die Spitze der Hitparaden anführten. Im Jahr 1970 trennten sich die Wege der vier Band-Mitglieder aufgrund interner Spannungen. Die Musiker verfolgten danach eigene musikalische Wege. The Beatles were an English rock band formed in Liverpool in 1960 and one of the most commercially successful and critically acclaimed acts in the history of popular music. The group's best-known lineup consisted of John Lennon (rhythm guitar, vocals), Paul McCartney (bass guitar, vocals), George Harrison (lead guitar, vocals) and Ringo Starr (drums, vocals). In the mid-1950s rock and roll, the group later worked in many genres ranging from pop ballads to psychedelic rock, often incorporating classical and other elements in their music. Their popularity first emerged as "Beatlemania"; as their songwriting grew in sophistication, by the late 1960s they came to be perceived by many fans and cultural commentators as the epitome of the ideals shared by the era's sociocultural revolutions. As a five-piece line-up of Lennon, McCartney and Harrison on guitar and vocals, with Stuart Sutcliffe (bass) and Paul McCartney (drums) built their reputation playing clubs in Liverpool and Hamburg over a three-year period from 1960. Sutcliffe left the group in 1961, and McCartney was replaced by Starr in 1962.



RDFa

- RDFa = “RDF in attributes”
- Extension to HTML5 for **embedding** RDF within HTML pages:
 - The **HTML** is processed by the browser, the (human) consumer doesn’t see the RDF data
 - The **RDF triples** within the page are consumed by APIs to extract the (semi-)structured data
- It is considered as the **bridge** between the Web of Data and the Web of Documents
- It is a complete **serialization of RDF**

RDFa: Attributes

Attribute role	Attribute	Description
Syntax	prefix	List of prefix-name IRIs pairs
	vocab	IRI that specifies the vocabulary where the concept is defined
Subject	about	Specifies the subject of the relationship
Predicate	property	Express the relationship between the subject and the value
	rel	Defines a relation between the subject and a URL
	rev	Express reverse relationships between two resources
Resource	href	Specifies an object URI for the rel and rev attributes
	resource	Same as href (used when href is not present)
	src	Specifies the subject of a relationship
Literal	datatype	Express the datatype of the object of the property attribute
	content	Supply machine-readable content for a literal
	xml:lang, lang	Specifies the language of the literal
Macro	typeof	Indicate the RDF type(s) to associate with a subject
	inlist	An object is added to the list of a predicate.

RDFa: Example

Extracting RDF from HTML

HTML (+RDFa):

```
<div class="artistheader"  
  about="http://musicbrainz.org/artist/b10bbbfccf9e-42e0-be17-e2c3e1d2600d#_"  
  typeof="http://purl.org/ontology/mo/MusicGroup">  
  ...  
</div>
```

RDF:

```
<http://musicbrainz.org/artist/b10bbbfccf9e-42e0-be17-e2c3e1d2600d#_>
```


RDFa: Example

Extracting RDF from HTML

HTML (+RDFa):

```
<div class="artistheader"  
  about="http://musicbrainz.org/artist/b10bbbfc-cf9e-42e0-be17-e2c3e1d2600d#_"  
  typeof="http://purl.org/ontology/mo/MusicGroup">  
  ...  
</div>
```

RDF:

```
<http://musicbrainz.org/artist/b10bbbfc-cf9e-42e0-be17-e2c3e1d2600d#_>  
  <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
```

RDFa: Example

Extracting RDF from HTML

HTML (+RDFa):

```
<div class="artistheader"  
  about="http://musicbrainz.org/artist/b10bbbfc-cf9e-42e0-be17-e2c3e1d2600d#_"  
  typeof="http://purl.org/ontology/mo/MusicGroup">  
  ...  
</div>
```


RDF:

```
<http://musicbrainz.org/artist/b10bbbfc-cf9e-42e0-be17-e2c3e1d2600d#_>  
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>  
<http://purl.org/ontology/mo/MusicGroup>.
```

RDFa: Example (2)

Extracting RDF from MusicBrainz.org

<http://musicbrainz.org/artist/b10bbbfc-cf9e-42e0-be17-e2c3e1d2600d>

 **The Beatles**
~ Group

Overview Releases Recordings Works Relationships Aliases Tags Details Edit

Annotation

The official name of "The White Album" is "The Beatles"; please do not add incorrectly titled releases to the database!

Please do not re-add individual mono remaster releases to their individual release groups. The mono remasters were never released separately, and are all already in the database as 13 mediums of the release the name "The Beatles in Mono".

Annotation last modified on 2012-10-09 07:44 UTC.

Wikipedia

The Beatles were an English rock band formed in Liverpool in 1960. They became the most commercially successful and critically acclaimed act in the rock music era. The group's best-known lineup consisted of John Lennon, Paul McCartney, George Harrison, and Ringo Starr. Rooted in skiffle and 1950s rock and roll, the Beatles later utilized several genres, ranging from pop ballads to psychedelic rock, often incorporating classical and other elements in innovative ways. In the early 1960s, their enormous popularity first emerged as "Beatlemania", but as their songwriting grew in sophistication, they came to be perceived by many fans and cultural observers as an embodiment of the ideals shared by the era's sociocultural revolutions.

[Continue reading at Wikipedia...](#)

Discography

Page 1 of 10

Album

Year	Title	Artist	Rating	Releases
1963	Please Please Me	The Beatles	★★★★★	10
1963	With The Beatles	The Beatles	★★★★★	10
1964	Introducing... The Beatles	The Beatles	★★★★★	3
1964	Meet The Beatles!	The Beatles	★★★★★	1

Artist information

Sort name: Beatles, The

Type: Group

Founded: 1957

Dissolved: 1970-04-10 (13 years)

Country: United Kingdom

Rating

★★★★★ (see all ratings)

Tags

pop, rock, british, british invasion, liverpool, more...

Editing

Log in to edit


External links

 [Official homepage](#)

 [Allmusic](#)

 [BBC Music](#)


 [Discogs](#)

 [IMDb](#)

 [LyricWiki](#)


 [@thebeatles](#)

 [MusicMoz](#)

 [thebeatles](#)

 [thebeatles](#)

 [The Beatles](#)

 [VIAF: 141205608](#)

 [en: The Beatles](#)



RDFa: Example (2)

Extracting RDF from MusicBrainz.org



W3C RDFa Distiller and Parser

This distiller corresponds to the [RDFa 1.0](#) specification. In 2012, W3C has published an updated version of that specification, called [RDFa Core 1.1](#). A [new distiller](#), processing RDFa 1.1 content, has been implemented which supersedes this one. Note that the new distiller can also process RDFa 1.0 content (there are some minor incompatibilities) if the XHTML+RDFa file uses the right (RDFa 1.0) DTD and/or the @version attribute. Users are advised to migrate to RDFa 1.1 in general, including the RDFa 1.1 distiller.

If you intend to use this service regularly on large scale, [consider downloading the package](#) and use it locally. Storing a (conceptually) “cached” version of the generated RDF, instead of referring to the live service, might also be an alternative to consider in trying to avoid overloading this server...

Distill by URI Distill by File Upload Distill by Direct Text Input

URI of HTML or SVG File:

Output Format:

▶ More Options

Go!

Source: <http://www.w3.org/2007/08/pyRdfa/>



RDFa: Example (2)

Extracting RDF from MusicBrainz.org

<http://www.w3.org/2007/08/pyRdfa/extract?uri=http%3A%2F%2Fmusicbrainz.org%2Fartist%2Fb10bbbfc-cf9e-42e0-be17-e2c3e1d2600d&format=nt>

```
<http://musicbrainz.org/release-group/f4571294-0bef-3f55-93bb-03alca101986#_> <http://www.w3.org/2000/01/rdf-schema#label> "Reel Music".
<http://musicbrainz.org/release-group/f4571294-0bef-3f55-93bb-03alca101986#_> <http://purl.org/dc/terms/title> "Reel Music".
<http://musicbrainz.org/release-group/f4571294-0bef-3f55-93bb-03alca101986#_> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://purl.org/ontology/mo/SignalGroup>.
<http://musicbrainz.org/release-group/e4752326-e5de-3376-a82a-394a2786fc66#_> <http://www.w3.org/2000/01/rdf-schema#label> "Have You Heard the Word?".
<http://musicbrainz.org/release-group/e4752326-e5de-3376-a82a-394a2786fc66#_> <http://purl.org/dc/terms/title> "Have You Heard the Word?".
<http://musicbrainz.org/release-group/e4752326-e5de-3376-a82a-394a2786fc66#_> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://purl.org/ontology/mo/SignalGroup>.
_vSuVqTRu6 <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://vocab.org/bio/0.1/Death>.
_vSuVqTRu6 <http://vocab.org/bio/0.1/date> "1970-04-10"^^<http://www.w3.org/2001/XMLSchema#date>.
<http://musicbrainz.org/release-group/0a4d1024-3463-3c74-8500-ea011bd5d523#_> <http://www.w3.org/2000/01/rdf-schema#label> "Yesterday\u2026 and Today".
<http://musicbrainz.org/release-group/0a4d1024-3463-3c74-8500-ea011bd5d523#_> <http://purl.org/dc/terms/title> "Yesterday\u2026 and Today".
<http://musicbrainz.org/release-group/0a4d1024-3463-3c74-8500-ea011bd5d523#_> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://purl.org/ontology/mo/SignalGroup>.
```




Watch the EUCLID screencast: <http://vimeo.com/euclidproject>

RDF Dump

- An RDF dump refers to a file which contains (part of) a data set specified in an RDF serialization
- The data set can be split into **several RDF dumps**
- A list of available data sets available as RDF dumps can be found at:
 - <http://www.w3.org/wiki/DataSetRDFDumps>

SPARQL Endpoint

- The **SPARQL endpoint** refers to the URI of the listener of the SPARQL protocol service, which handles requests for SPARQL protocol operations
- The user submits **SPARQL queries** to the SPARQL endpoint in order to retrieve only a desired subset of the RDF data set 
- List of available SPARQL endpoints:
 - <http://www.w3.org/wiki/SparqlEndpoints>
 - <http://sparql.es.okfn.org/>



AVAILABILITY
Up / Down

PERFORMANCE
Cold / Warm

INTEROPERABILITY
SPARQL 1.0 / 1.1

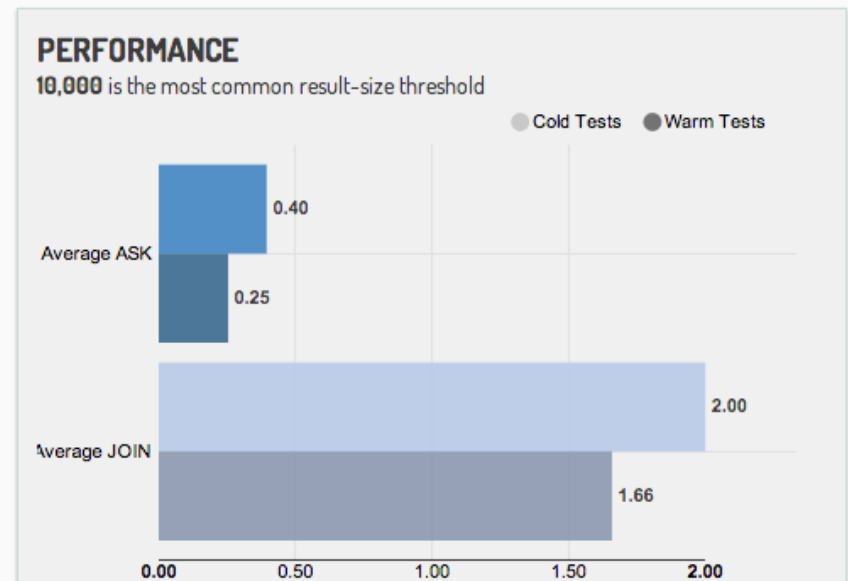
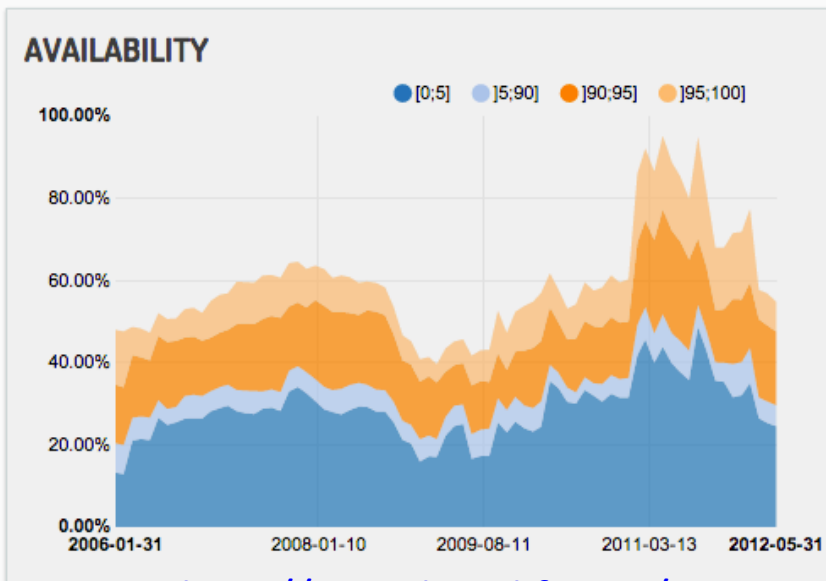
DISCOVERABILITY
VoID, Server

508 endpoints

SPARQL ENDPOINTS STATUS > DATAHUB.IO

Last update: Tue Sep 02 2014 09:17:30 GMT+0000 (UTC)

► Description:



<http://sparqls.okfn.org/>



Using Linked Data Catalogs

- Data catalogs, markets or repositories are platforms dedicated to **provide visibility** to a wide range of data sets from different domains
- Allow **data consumers** to easily find and use the data
- Usually the catalogs offer relevant **metadata** about the creation of the data set

How to publish an RDF data set into a catalog?

Create your own data catalog

Recommended for big organizations/institutions aiming at providing a large number of data sets

Use a data management system, for example:



Upload your data set into an existing catalog

Allows data consumers to easily find new data sets

Common LD catalogs are:

- [the Data Hub](#)
- The Linking Open Data Cloud

Validating Data Sets

There are different ways to validate the published RDF data set:

Accessibility

- **Vapour** - Performs two types of tests: without content negotiation and requesting RDF/XML content
<http://validator.linkeddata.org/vapour>
- **URI Debugger** - Retrieves the HTTP responses of accessing a URI
<http://linkeddata.informatik.hu-berlin.de/uridbg/>
- **RDF Triple-Checker** – Dereferences namespaces associated with the resources used in the document
<http://graphite.ecs.soton.ac.uk/checker/>

Parsing & Syntax

- **W3C RDF/XML Validation Service** – Evaluates the syntax of RDF/XML documents and displays the RDF triples in it
<http://validator.linkeddata.org/vapour>
- **W3C Markup Validation Service** – Checks syntactic correctness for web documents with RDFa markup
<http://validator.w3.org/>

General validators

- **RDF:ALERTS** – Validates syntax, undefined resources, datatype and other types of errors
<http://swse.deri.org/RDFAlerts/>



Validating Data Sets (2)

Example: Validating URIs with Vapour

VAPOUR a Linked Data validator

Validate by URI

URI:
(example: <http://dbpedia.org/resource/Asturias>)

▼ **More options**

- Tests if the RDF responses contain meaningful data (takes some time)
- Test if there are HTML description of the resource
- Test mixed accept headers and q values (takes some time) ⓘ

Expected default response (without content negotiation)

User Agent:

Source: <http://idi.fundacionctic.org/vapour>

Validating Data Sets (3)

Example: Validating URIs with Vapour

Vapour Report

All tests passed!

Summary:

Test requirement	Passed tests
Dereferencing resource URI (requesting RDF/XML)	3/3
Dereferencing resource URI (without content negotiation)	2/2

Source: <http://idi.fundacionctic.org/vapour>



Publishing Linked Data (Quick Recipe)

Steps:

- Create dataset **metadata**
- Provide **access mechanisms**:
 - RDF dump
 - SPARQL endpoint
- **(Expose the dataset in a catalog)**
- **Validate** everything!



Published
RDF dataset

PROVIDING LINKED DATA: CHECKLIST

Providing Linked Data: Checklist (1)

Creating Linked Data

- All the relevant entities/concepts were effectively extracted from the raw data ?
- Are all the created URIs dereferenceable?
- Are you reusing terms from widely accepted vocabularies?

Providing Linked Data: Checklist (2)

Interlinking Linked Data

- Is the data set linked to other RDF data sets?
- Are the created vocabulary terms linked to other vocabularies?

Providing Linked Data: Checklist (3)

Publishing Linked Data

- Do you provide data set metadata?
- Do you provide information about licensing?
- Do you provide additional access methods?
- Is the data set available in LD catalogs?
- Did the data set pass the validation tests?

Summary

In this chapter we studied:

- The Linked Data **lifecycle**:
 - 3 core tasks: creating, interlinking and publishing
- **Creation** of Linked Data:
 - Extracting relevant data, using URIs to name entities and selecting vocabularies and expressing the data using the RDF data model
- **Interlinking** Linked Data:
 - Challenges of link discovery, using Silk to create links between two data sets and using SKOS links
- **Publishing** Linked Data:
 - Creation of data set metadata; publishing the data set via RDF dumps, SPARQL endpoints or RDFa; uploading the data set to a LD catalog

For exercises, quiz and further material visit our website:



<http://www.euclid-project.eu>

eBook



amazonkindle



ePUB



iBooks

Course



iTunes U

Other channels:



@euclid_project



euclidproject



euclidproject