Practical Linked Data for MIR Researchers

What, why, and how.

ISMIR 2011, Miami, 24th October 2011.

Kevin Page,

David De Roure,

Oxford e-Research Centre, University of Oxford.

For updates, examples, and resources, visit:

http://ismir2011.linkedmusic.org/

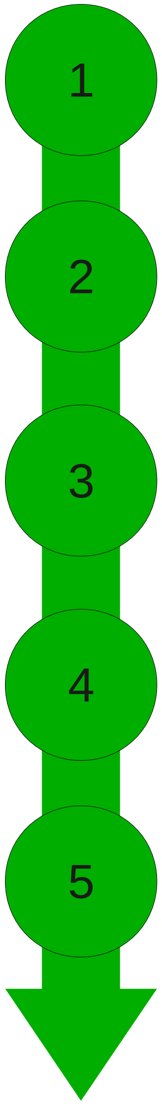
kevin.page@oerc.ox.ac.uk

So what is this linked data business,
and why should I care?

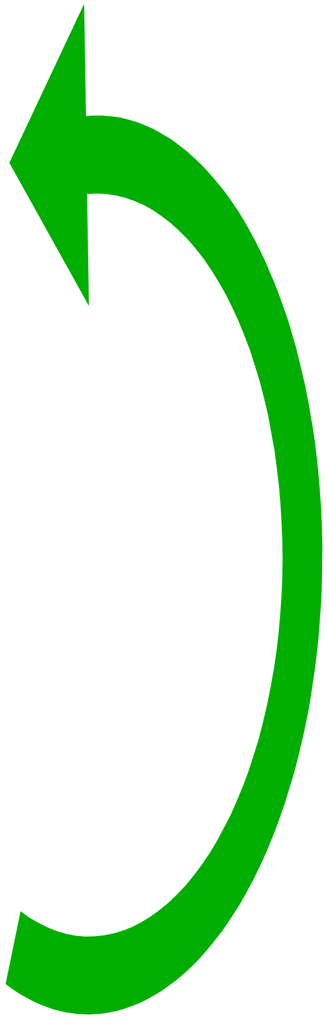
Motivation

http://www.flickr.com/photos/emmablackburn/4817493089/

As an MIR researcher you...



- Building upon previous algorithms and output data...
- Take an input data set
- Develop, combine, or finesse an algorithm or process
- Produce results and output data
- ...which can be used, combined (and improved?)



It's about making your output useful to others...

...and building upon the output of others to make your own work better

...or, at least, easier

http://www.flickr.com/photos/virginiazuluaga/3600587236/

Tutorial Outline

- Context and overview of the Semantic Web
- An example system: Introduction
- Technologies (RDF, Turtle, Linked Data and SPARQL)
  - *Hands on session*
- Technologies (Ontologies, RDFS, OWL, Web Architecture)
- An example system: Implementation (part 1)
  - *Hands on session*
- An example system: Implementation (part 2)
- More on ontologies
- Linked Data: projects, tools, opportunities

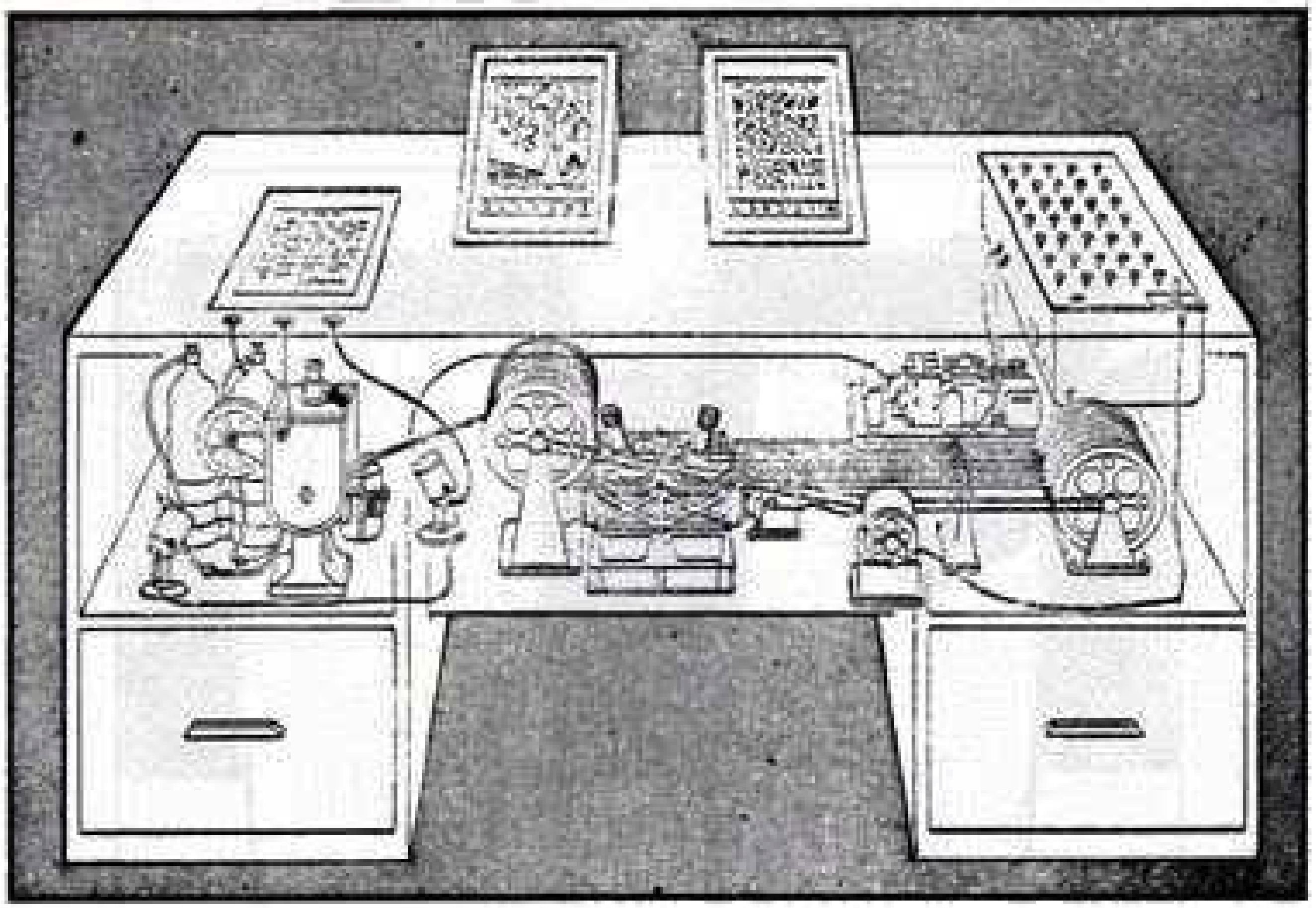
Tutorial examples

- We've tried to keep our examples simple and generic
- Rather than covering everything, we'd like you to leave us today with a good understanding of how the basics work
  - and through examples that step through “by hand”, get a feel for what's going on within
- Our example system provides a framework to understand how Linked Data could interact with an MIR system, but we won't cover all the specialisms the MIR community encompasses

A brief history of the Semantic Web

http://www.flickr.com/photos/pshab/428577252/

Vanevar Bush and the memex



Memex in the form of a desk would instantly bring files and material on any subject to the operator's fingertips. Slanting translucent viewing screens magnify supermicrofilm filed by code numbers. At left is a mechanism which automatically photographs longhand notes, pictures and letters, then files them in the desk for future reference (*LIFE* 19(11), p. 123).

http://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/3881/

Doug Engelbart

“He envisioned intellectual workers sitting at display 'working stations', flying through information space, harnessing their collective intellectual capacity to solve important problems together in much more powerful ways. Harnessing collective intellect, facilitated by interactive computers, became his life's mission at a time when computers were viewed as number crunching tools.”

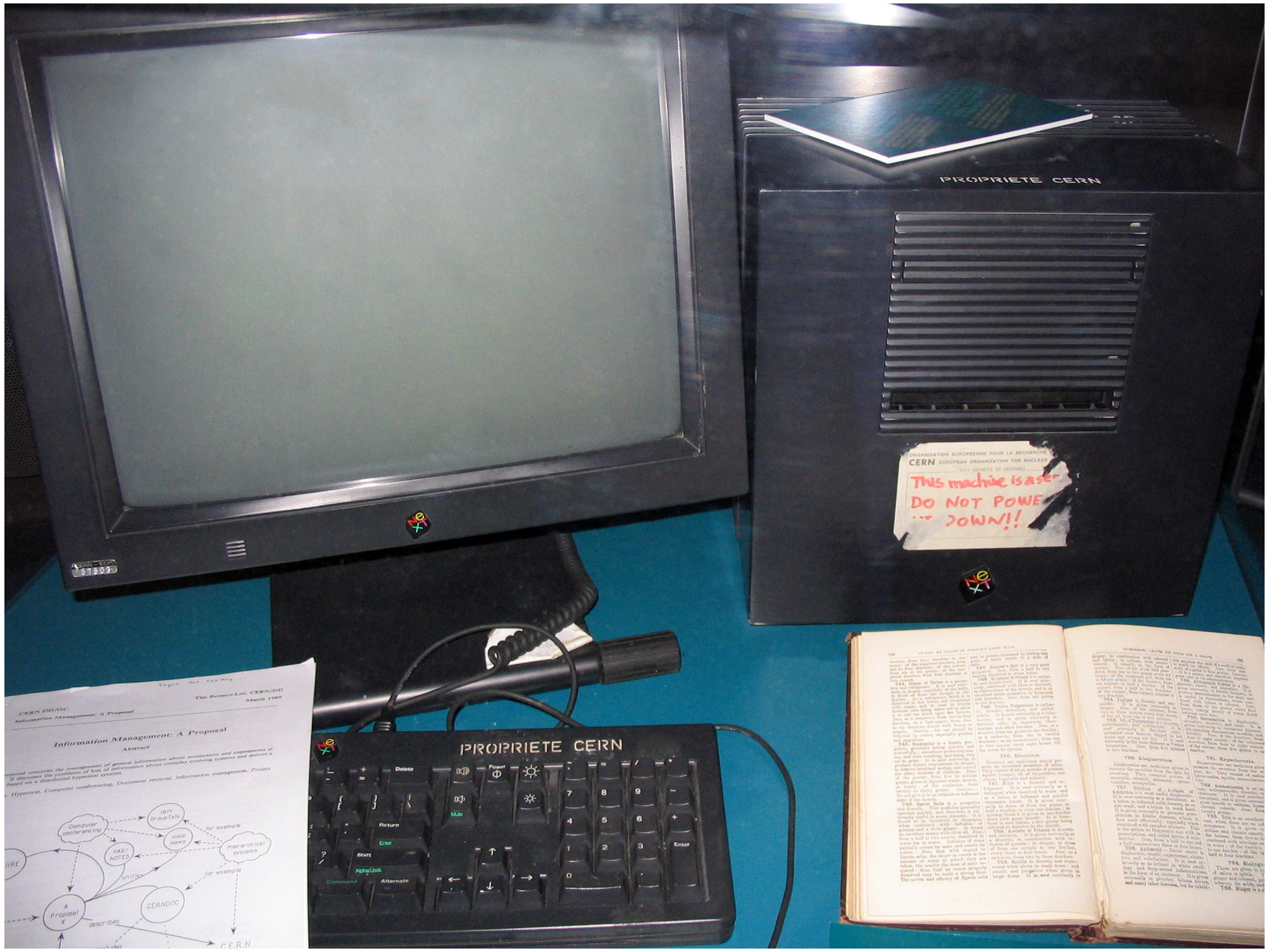
http://en.wikipedia.org/wiki/Douglas_Engelbart

Ted Nelson and hypertext

AN AUTOBIOGRAPHY OF TED NELSON
POSSIPLEX
ovies, Intellect, Creative Control,
ly Computer Life
d the Fight for Civilization
ABOUT THIS BOOK
ABOUT THE AUTHOR
Diet Coke

http://www.flickr.com/photos/jaycross/5063853771/

Tim Berners-Lee and the World Wide Web

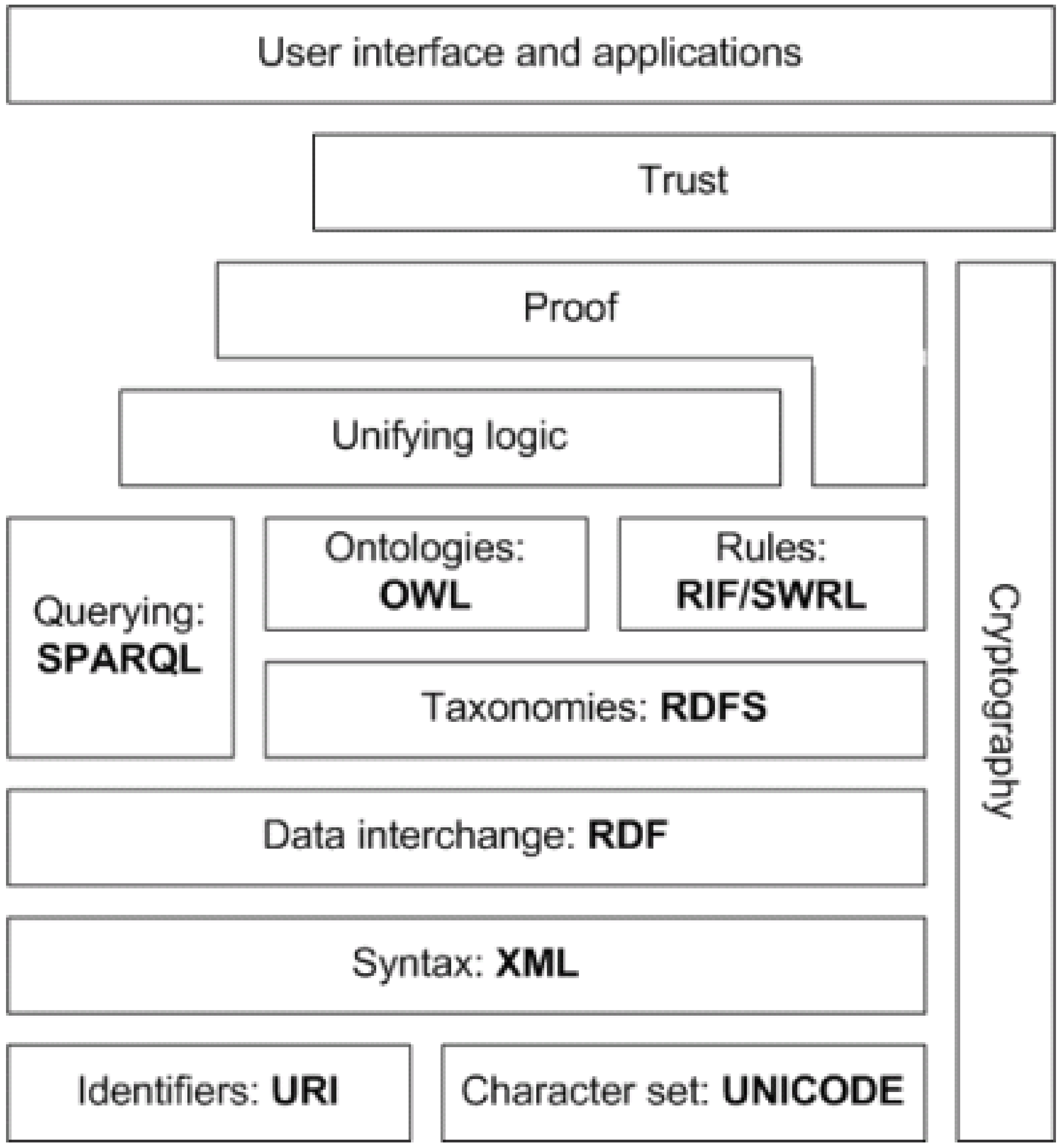


“This machine is a server – DO NOT POWER IT DOWN!!”

So what, then, is the *Semantic* Web?

- The web is the largest and most successful distributed system ever constructed
- But it is a Web of Documents
- The Semantic Web is the effort to create the equivalent Web of Data

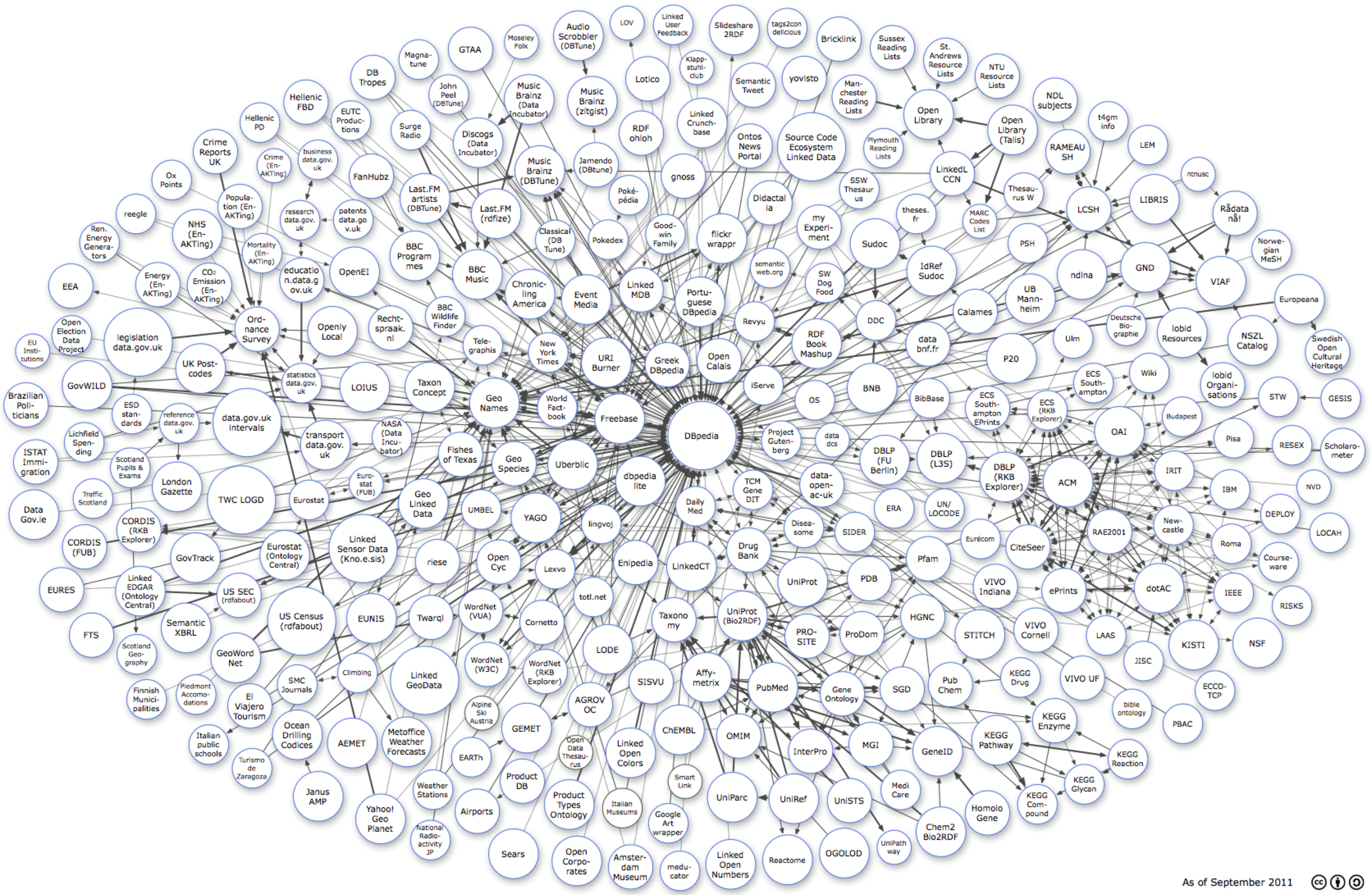
- Semantic Web activities have been baking since the late 1990s
- The infamous layer cake has evolved
- Linked Data is a more recent movement...



Linked Data

- There are two words in Semantic Web
- Both are important!

So how's it going?

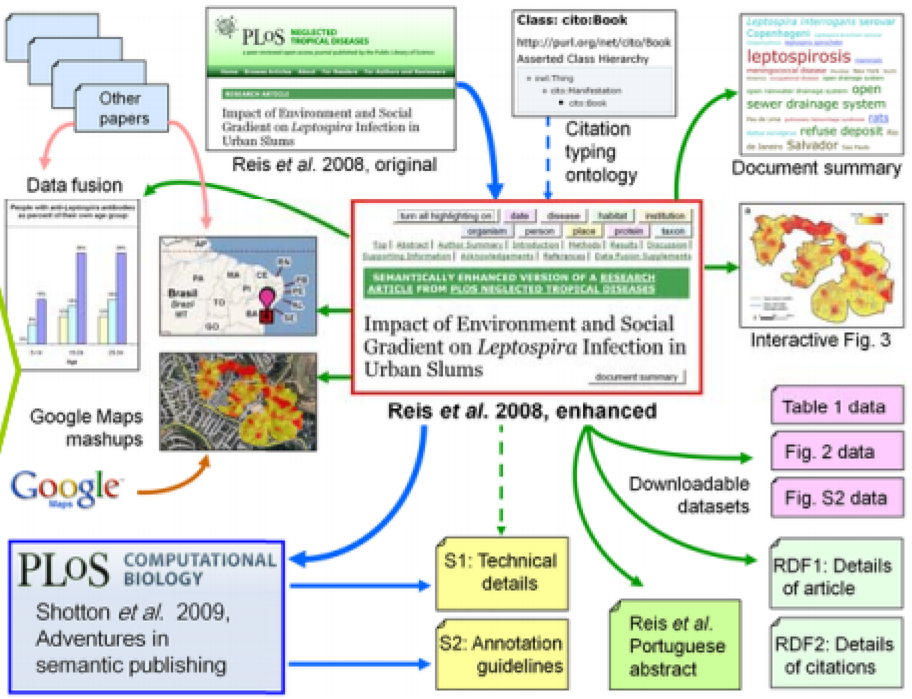


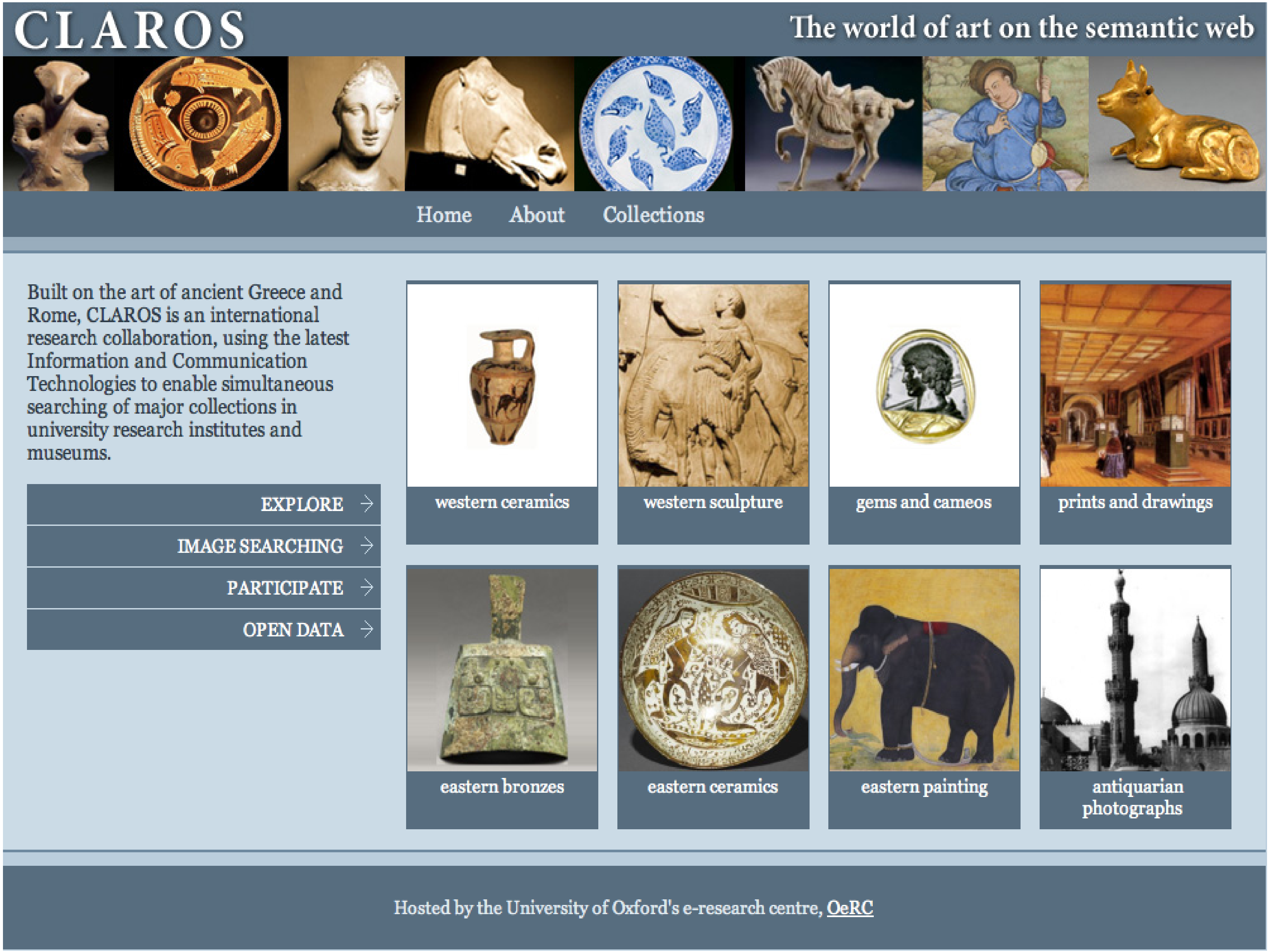
Richard Cyganiak and Anja Jentzsch. http://lod-cloud.net/

Researchers are looking for tools to help them find, integrate, and re-use content

"**Scientific innovation** depends on finding, integrating, and re-using the products of previous research... **Our semantic enhancements led to the creation of a whole "ecosystem"** of articles, documents, spreadsheets, data fusions related to that original work...**frictionless interoperability between papers and datasets is highly desirable.**"

Shotton D. et al, Adventures in Semantic Publishing: Exemplar Semantic Enhancements of a Research Article. *PLoS Comput Biol* 5(4)


Other papers
Data fusion
Reis *et al.* 2008, original
Citation typing ontology
Document summary
Interactive Fig. 3
Google Maps mashups
Google
Reis *et al.* 2008, enhanced
Impact of Environment and Social Gradient on *Leptospira* Infection in Urban Slums
Table 1 data
Fig. 2 data
Fig. S2 data
Downloadable datasets
RDF1: Details of article
RDF2: Details of citations
PLoS COMPUTATIONAL BIOLOGY
Shotton *et al.* 2009, Adventures in semantic publishing
S1: Technical details
S2: Annotation guidelines
Reis *et al.* Portuguese abstract



CLAROS

The world of art on the semantic web

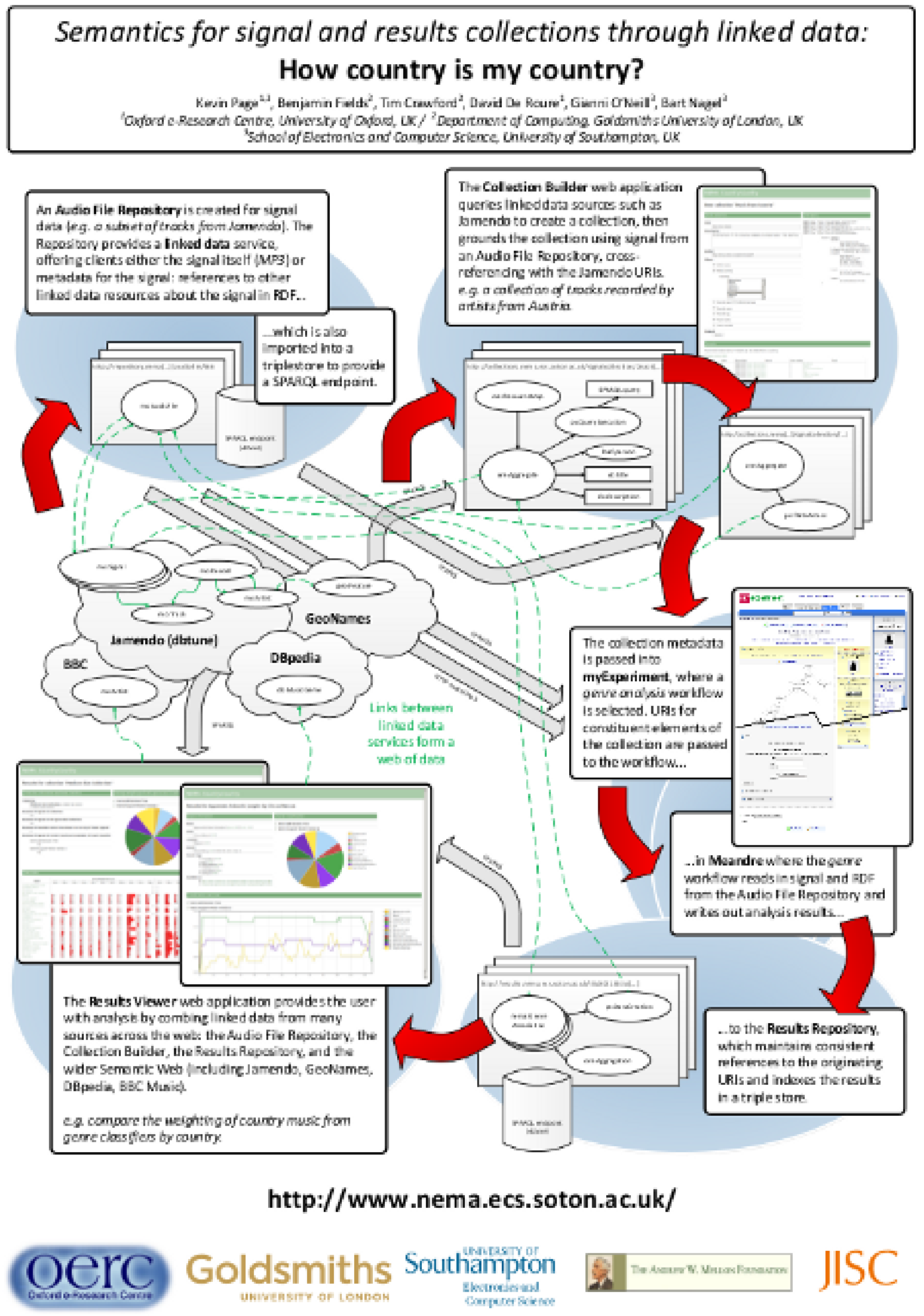
Built on the art of ancient Greece and Rome, CLAROS is an international research collaboration, using the latest Information and Communication Technologies to enable simultaneous searching of major collections in university research institutes and museums.

Hosted by the University of Oxford's e-research centre, OeRC

Music is a great Linked Data opportunity

- there is already data related to the field
- there is general interest and use of music Linked Data
- other people will find *your* data interesting and useful

A simple example



Semantics for signal and results collections through linked data:
How country is my country?

Kevin Page[1,3], Benjamin Fields[2], Tim Crawford[2], David De Roure[1], Gianni O'Neill[3], Bart Nagel[3]
[1]*Oxford e-Research Centre, University of Oxford, UK* / [2]*Department of Computing, Goldsmiths University of London, UK*
[3]*School of Electronics and Computer Science, University of Southampton, UK*

An **Audio File Repository** is created for signal data (*e.g. a subset of tracks from Jamendo*). The Repository provides a **linked data** service, offering clients either the signal itself (MP3) or metadata for the signal: references to other linked data resources about the signal in **RDF**..

...which is also imported into a triplestore to provide a SPARQL endpoint.

The **Collection Builder** web application queries linked data sources such as Jamendo to create a collection, then grounds the collection using signal from an Audio File Repository, cross-referencing with the Jamendo URIs. *e.g. a collection of tracks recorded by artists from Austria.*

The collection metadata is passed into **myExperiment**, where a *genre analysis workflow* is selected. URIs for constituent elements of the collection are passed to the workflow...

...in **Meandre** where the *genre* workflow reads in signal and RDF from the Audio File Repository and writes out analysis results...

...to the **Results Repository**, which maintains consistent references to the originating URIs and indexes the results in a triple store.

The **Results Viewer** web application provides the user with analysis by combing linked data from many sources across the web: the Audio File Repository, the Collection Builder, the Results Repository, and the wider Semantic Web (including Jamendo, GeoNames, DBpedia, BBC Music).

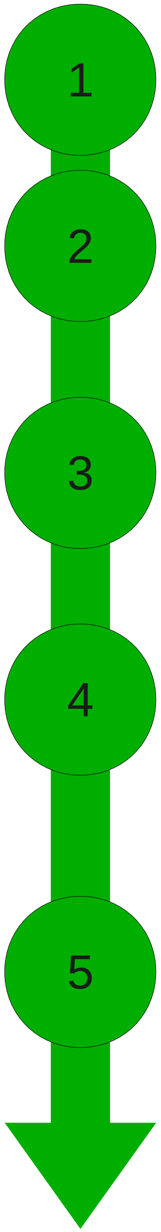
e.g. compare the weighting of country music from genre classifiers by country.

http://www.nema.ecs.soton.ac.uk/

A “simple” example

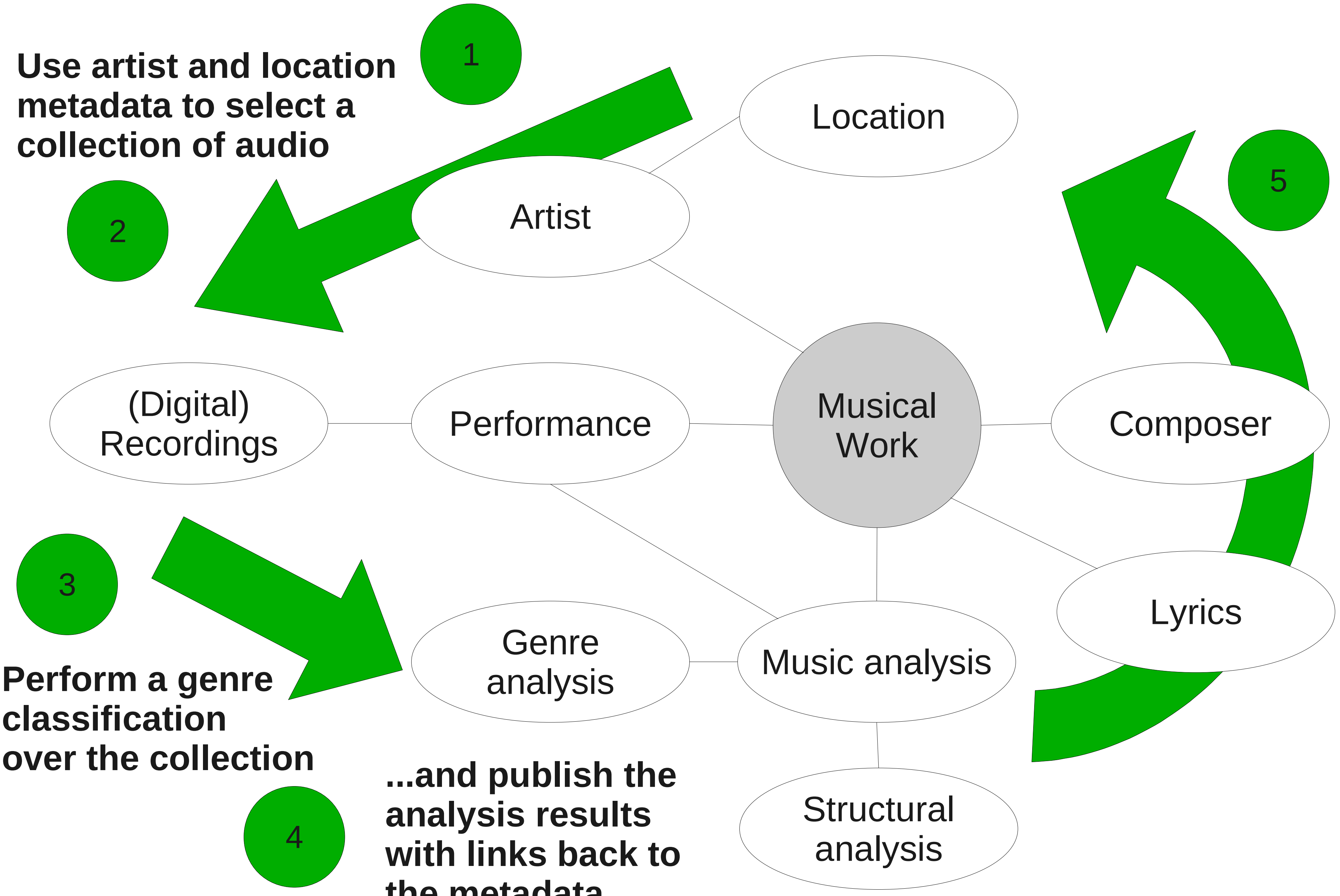
http://www.flickr.com/photos/aptmetaphor/4260444057/

A simple example

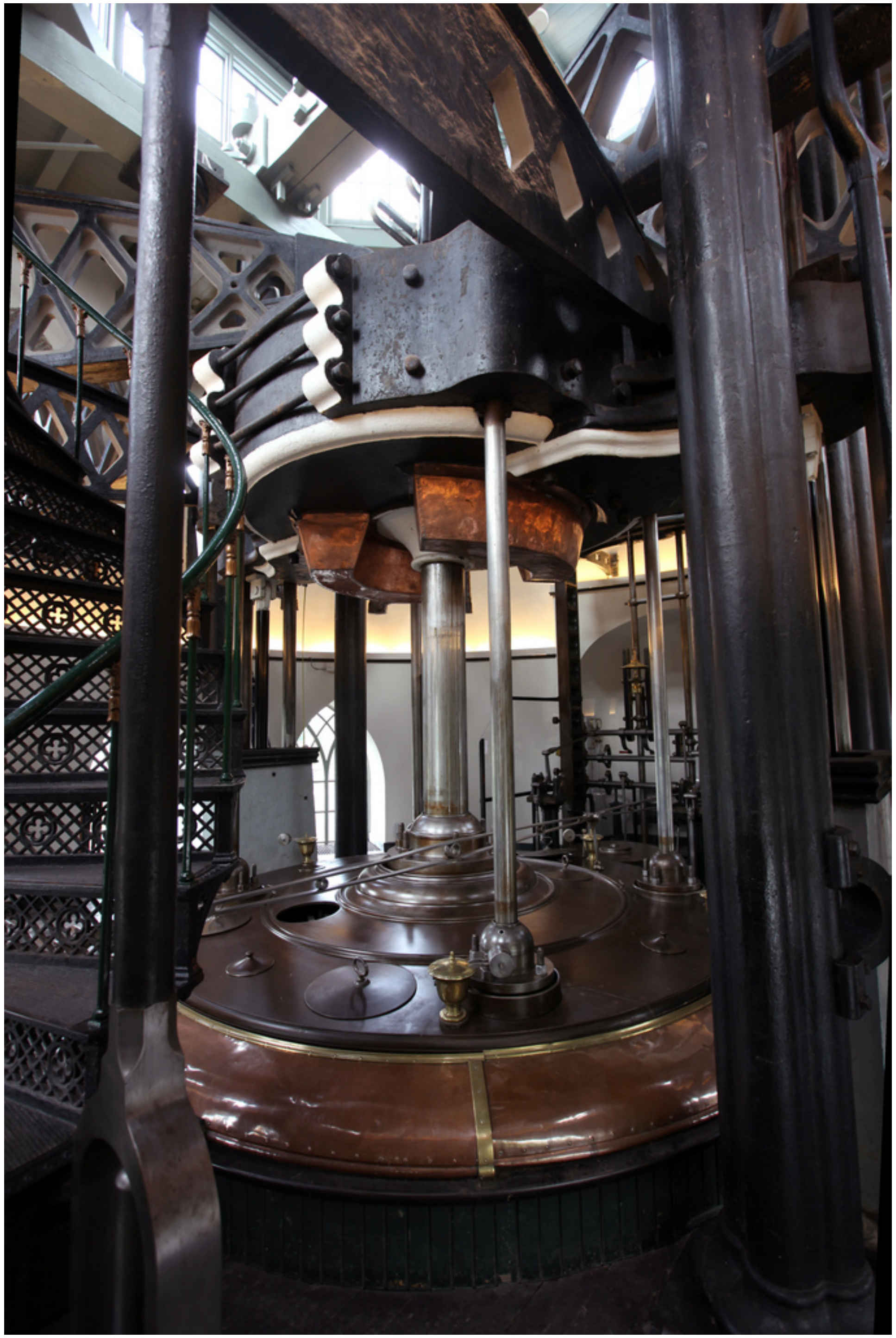


- Use artist and location metadata to select a collection(s) of audio
- Perform a genre classification over the collection
- Publish the genre analysis data with links back to the tracks (and so artist, location, and collections)
- Combine the results with other published metadata (about the artist, location, collections)

A simple example (simplified!)


1
Use artist and location metadata to select a collection of audio
2
3
Perform a genre classification over the collection
4
...and publish the analysis results with links back to the metadata
5
Location
Artist
(Digital) Recordings
Performance
Musical Work
Composer
Lyrics
Genre analysis
Music analysis
Structural analysis

Technologies



http://www.flickr.com/photos/45662099@N07/5948796257/

Key Technologies

- RDF
  - Turtle
- Linked Data
- SPARQL
- Ontologies
- Web Architecture
  - REST and Resource Oriented Architecture

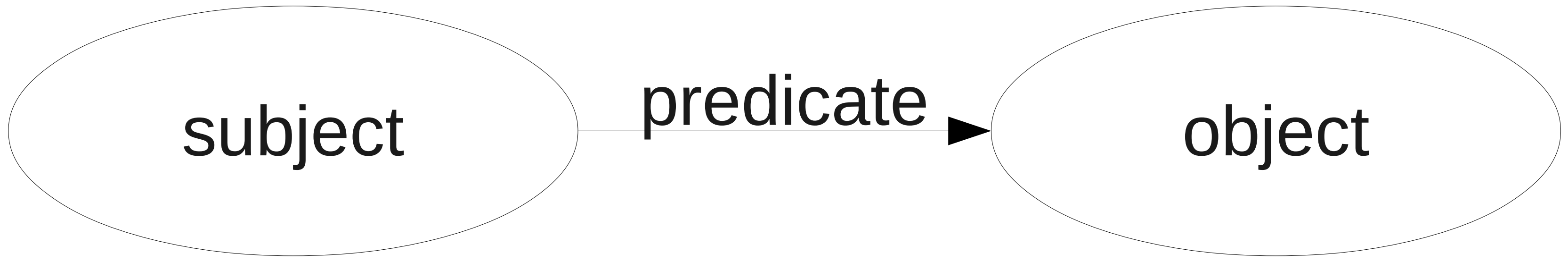
What is RDF?

aka why should I use it instead of XML?

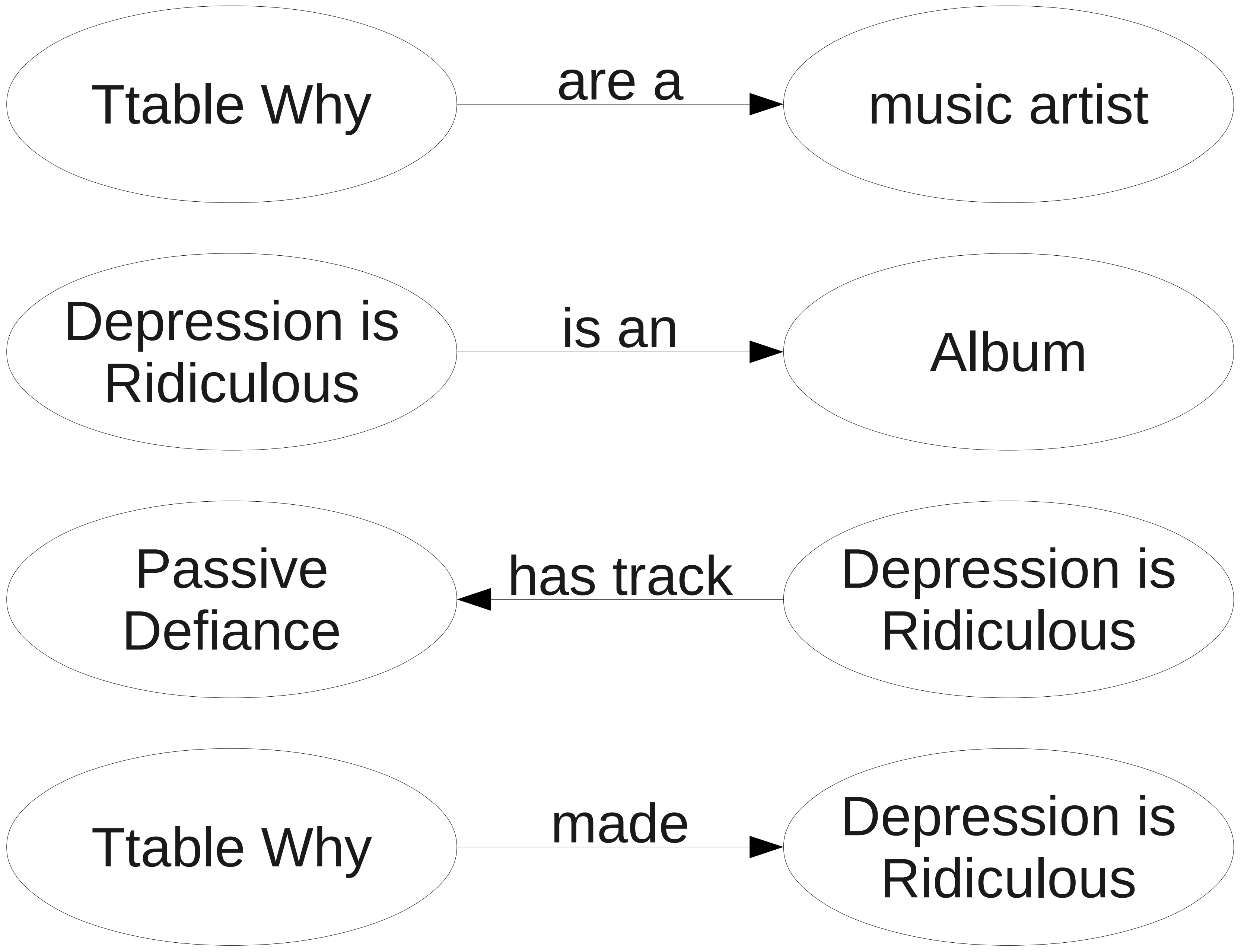
RDF

- Is the Resource Description Framework.
- Is a means of encoding machine-readable, self-describing, meaning.
- RDF is a simple model.

RDF is a simple model

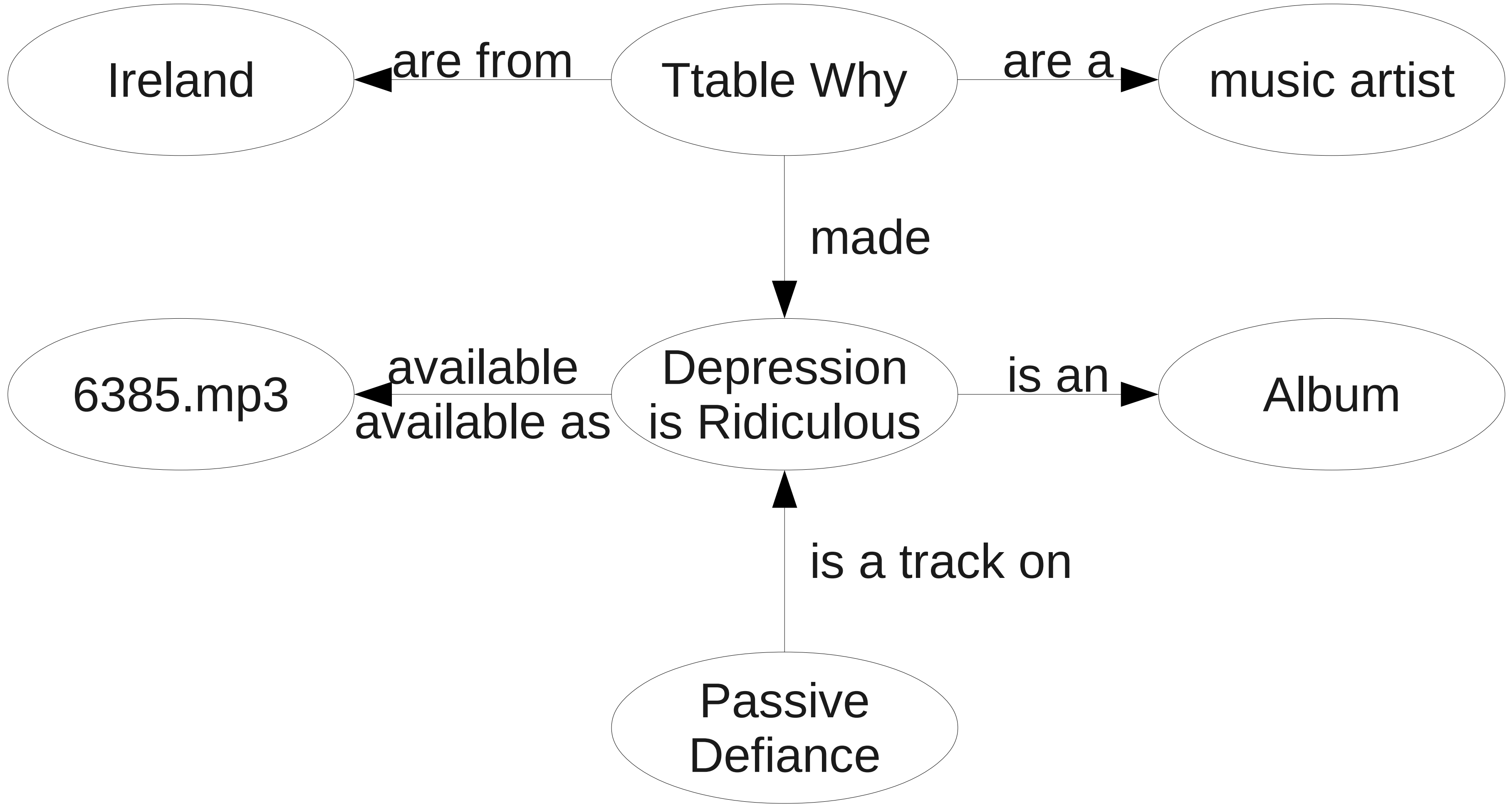


That's it.



Warning: approximated for clarity!

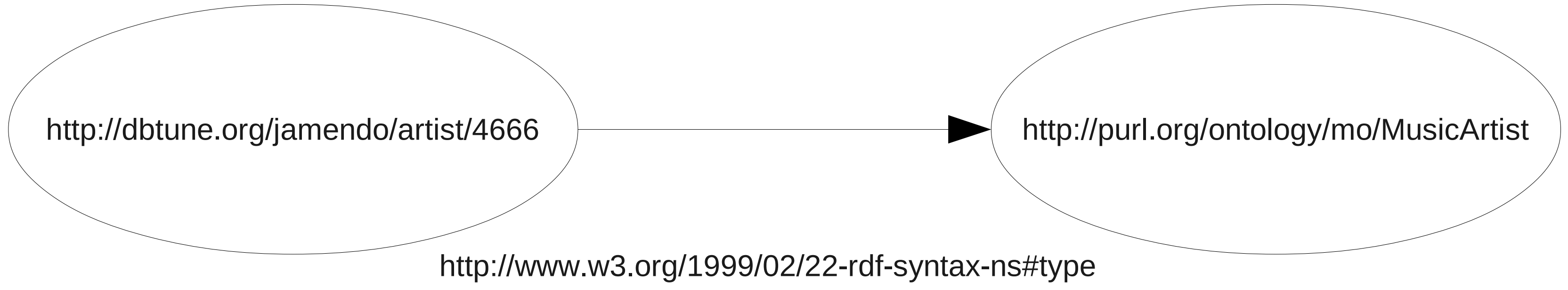
RDF “Triples” joined together form a graph



Warning: approximated for clarity!

URIs everywhere!

- Very significantly, each element of the triple can be a URI – a Uniform Resource Identifier
- e.g. *http://dbtune.org/jamendo/artist/4666*


http://dbtune.org/jamendo/artist/4666
http://purl.org/ontology/mo/MusicArtist
http://www.w3.org/1999/02/22-rdf-syntax-ns#type

Benefits of URIs

- This means we are always clear whether we are referring to the same Thing, or property, or concept, ...
- The graph can be split up and distributed while retaining consistency (if desired)
- We can also use HTTP URIs as a mechanism for retrieving parts of a graph, (but that's a story for a little later...)
- It's a web of data!

How do we express our RDF?

http://www.flickr.com/photos/jin_aili/5365904496/

RDF != RDF/XML

- RDF is an abstract model with several serializations.
- RDF/XML is one of them, but... it's ugly.
- RDF/XML is for our friends the machines, not for us.
- Other serialisations include RDFa, N-Triples, JSON-LD, and...

Turtle

- Much more suitable for humans hacking RDF
- Being standardised by the current W3C RDF Working Group
- “because it is awesome” [1]

[1] http://www.w3.org/2001/sw/wiki/index.php?title=RDF_Core_Work_Items#Turtle

Turtle syntax (1)

- < > for URIs
  - e.g. <http://dbtune.org/jamendo/track/71263>
- “ “ for literals
  - e.g. "1"
- ^^ for typed literals
  - e.g. "1"^^<http://www.w3.org/2001/XMLSchema#int>
- @prefix to set a namespace
  - e.g. @prefix mo: <http://purl.org/ontology/mo/> .
  - Allowing CURIEs such as (no < >)
    - mo:Track

Turtle syntax (2)

- Then we write triples
  - the subject must be a URI
  - the predicate must be a URI
  - The object can be a URI or a literal
- Finish a statement with a full stop .
- Concatenate several statements about one subject with a semi-colon ;
- a is shorthand for rdf:type
- , to separate different objects with the same subject and predicate

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix mo: <http://purl.org/ontology/mo/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix ns1: <http://www.openrdf.org/schema/serql#> .

<http://dbtune.org/jamendo/record/6385> mo:track
     <http://dbtune.org/jamendo/track/71265> .

<http://dbtune.org/jamendo/signal/71265> mo:published_as
     <http://dbtune.org/jamendo/track/71265> .

<http://dbtune.org/jamendo/track/71265> rdf:type mo:Track , rdfs:Resource ;

     mo:available_as <http://www.jamendo.com/get/track/id/track/audio/play/71265> ,
     <http://www.jamendo.com/get/track/id/track/audio/xspf/71265> ;


     dc:title "Disenchantment
                         Anguish"^^<http://www.w3.org/2001/XMLSchema#string> ;


     mo:license <http://creativecommons.org/licenses/by-nc-sa/3.0/> ;


     mo:track_number "3"^^<http://www.w3.org/2001/XMLSchema#int> .
```

Where do we put our RDF?

Where do we get RDF from?

http://www.flickr.com/photos/51686984@N05/6120319026/

On the Web: Linked Data

- The Linked Data principles are specific that the URIs in RDF must be *HTTP* URIs
- HTTP, the Hypertext Transfer Protocol, is familiar to us as the application layer protocol used for the WWW
    - It operates on URIs through GET, POST, PUT and DELETE methods
- With Linked Data as well as *identifying* our triples, we can also use HTTP URIs to retrieve them

GET it

- GET http://dbtune.org/jamendo/signal/71265 and request an RDF content type, and receive...

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix mo: <http://purl.org/ontology/mo/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix ns1: <http://www.openrdf.org/schema/serql#> .

<http://dbtune.org/jamendo/record/6385> mo:track
     <http://dbtune.org/jamendo/track/71265> .

<http://dbtune.org/jamendo/signal/71265> mo:published_as
     <http://dbtune.org/jamendo/track/71265> .

<http://dbtune.org/jamendo/track/71265> rdf:type mo:Track , rdfs:Resource ;

     mo:available_as <http://www.jamendo.com/get/track/id/track/audio/pl.....
```

“Follow your nose”

http://www.flickr.com/photos/publicenergy/3867426297/

There is a slight complication...

- It's a little tiresome, but *is* important.
- The URI http://dbtune.org/jamendo/artist/4666 is not the artist *Ttable Why*. It's a collection of information *about* Ttable Why.
- To distinguish between them, when we request a resource that's a “real world” thing, the web server will return a *303 See Other* code pointing to another resource
  - which is where we'll get our RDF

RDF In a triplestore

- RDF can also be stored in a “triplestore”
- This can also be thought of as a cache where triples are brought together to do useful things to them as a whole
  - e.g. reasoning over the triplestore
- Most usefully from our perspective, we can query such collections of RDF through a powerful interface called SPARQL

SPARQL

- The SPARQL Protocol and RDF Query Language
  - Let's just stick to the query language
- SPARQL has several queries, we'll focus on SELECT
  - Also CONSTRUCT, ASK, and DESCRIBE
- As with Turtle, we start a query by defining prefixes

SPARQL Queries (1)

- Then starting from a Turtle-derived statement, e.g.

```
<http://dbtune.org/jamendo/track/71265> dc:title
                                         "Disenchantment Anguish".
```

- Consider how we can query by making parts of the statement variable, e.g.

```
<http://dbtune.org/jamendo/track/71265> dc:title ?title.

?track dc:title ?title.
```

- and combine

```
?subject dc:title ?title.
?subject rdf:type mo:Track.
```

SPARQL Queries (2)

- Other Turtle-derived syntax works

```
?subject dc:title ?title;
     rdf:type mo:Track.
```

- Using variables we can write “patterns” that match against parts of the graph

```
?artist foaf:name “Ttable Why”.
?album foaf:maker ?artist.
?track mo:track ?album.
```

- Any variables in a SELECT query must be bound in the graph pattern

SPARQL Queries (3)

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX mo: <http://purl.org/ontology/mo/>
PREFIX dc: <http://purl.org/dc/elements/1.1/>

    SELECT ?artistname WHERE {

        ?artist
                a mo:MusicArtist ;
                foaf:name ?artistname ;
                foaf:made ?album .



        ?album
                a mo:Record ;
                dc:title "Depression Is Ridiculous".

    }
```

Linked Data Principles

- Before we take a break, let's relate what we've seen to the Linked Data principles:
  1. Use URIs as names for things
  2. Use HTTP URIs so that people 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 they can discover more things

Hands on session 1

Following your nose, music for the minority, the 303 dance, and the transient nature of (wiki) fame

So far...

- We've learnt what RDF is
- We've found out where to get it
- We've seen how to query it

How do we exchange meaning through our RDF? Ontologies.

What is an ontology?

- An ontology is a description of concepts and their relationships
- It enables us to build semantic models with RDF (more specific models than RDF itself)
- It's about adding meaning to your data so that it can be “understood” and reused

So you just need to...

- Get the model right...
- Get *a* model right...
- Get *your* model right...

Avoiding rabbit holes

http://www.flickr.com/photos/hadevereux/5400490550/

Appropriate Ontologies

- Ontologies don't remove complexity, but they do enable us to scale it (relatively) gracefully.
- More than one “correct” ontology can be applicable to a resource – it depends what you're doing with it.
- Where available, use an applicable existing ontology (or extend it).
- Write the ontologies you *need* – you can extend them later. Better it be limited and right...
- It is probably unwise to expect an ontology for all Things...

One size...

...does not fit all

http://www.flickr.com/photos/saint_will/3335219409/

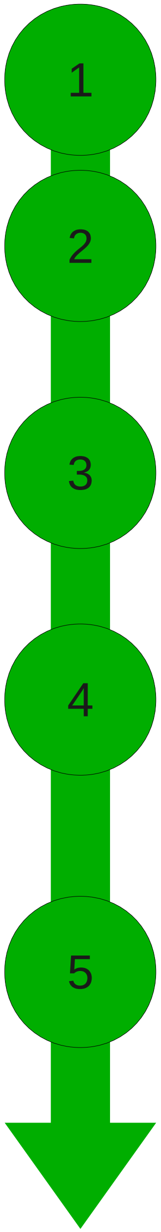
What you might want an ontology for

- Your input.
- Your output.
- Your method.

Ontologies you've already encountered today...

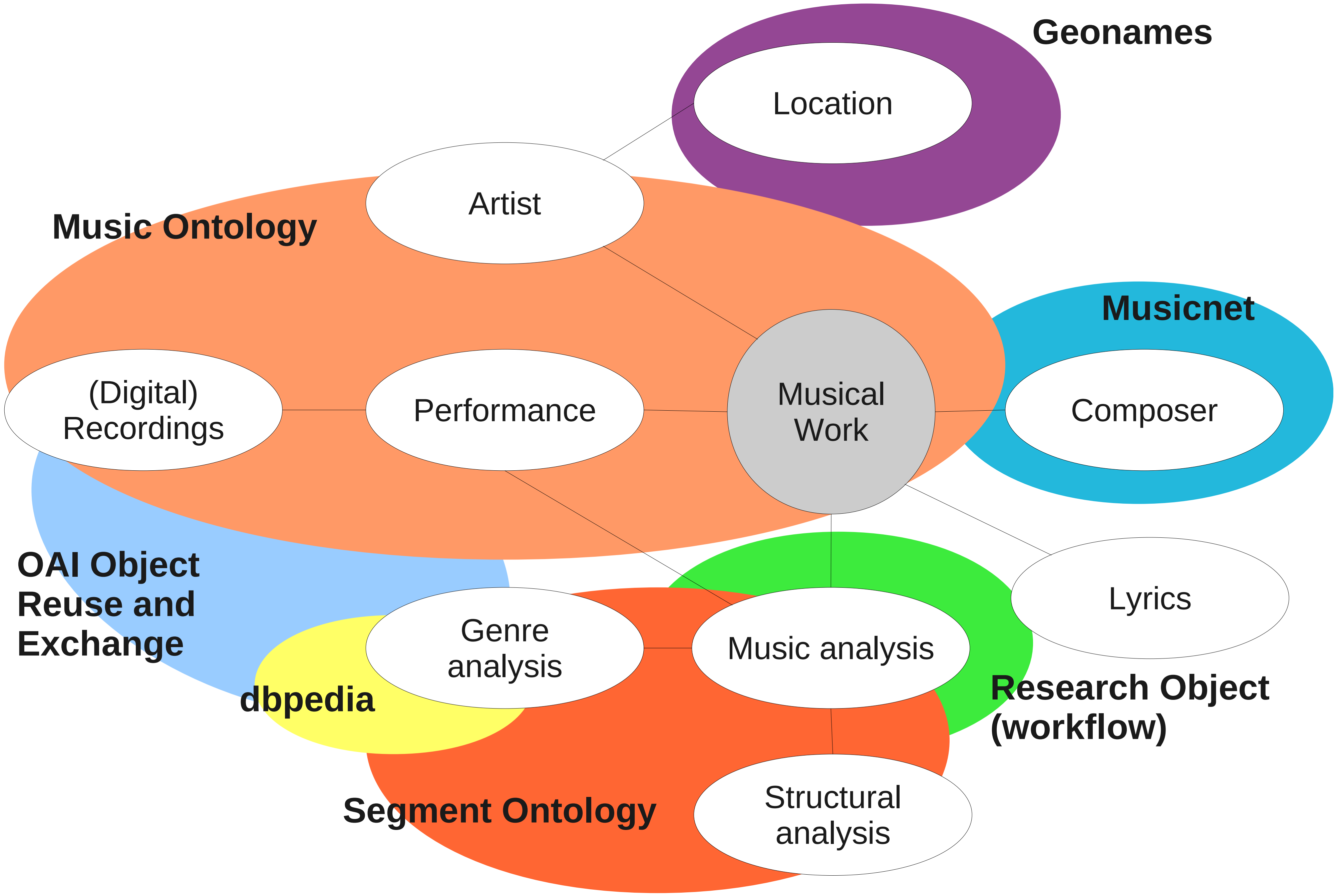
- The Music Ontology
- Friend-of-a-Friend (FOAF)
- Dublin Core
- dbpedia

Our simple example



- Use artist and location metadata to select a collection(s) of audio
- Perform a genre classification over the collection
- Publish the genre analysis data with links back to the tracks (and so artist, location, and collections)
- Combine the results with other published metadata (about the artist, location, collections)

Ontologies in our simple example



Ontologies in our simple example

- Each of these conceptual areas is a specialisation
  - which might be the subject of scholarly study
  - or computational analysis
  - or crowdsourcing, etc.
- There will be overlap
  - one person's metadata is another person's data
  - we can build upon others specialisation and knowledge
- We do not expect complexity to vanish
  - but where it has been studied it should be scaled, shared, and linked

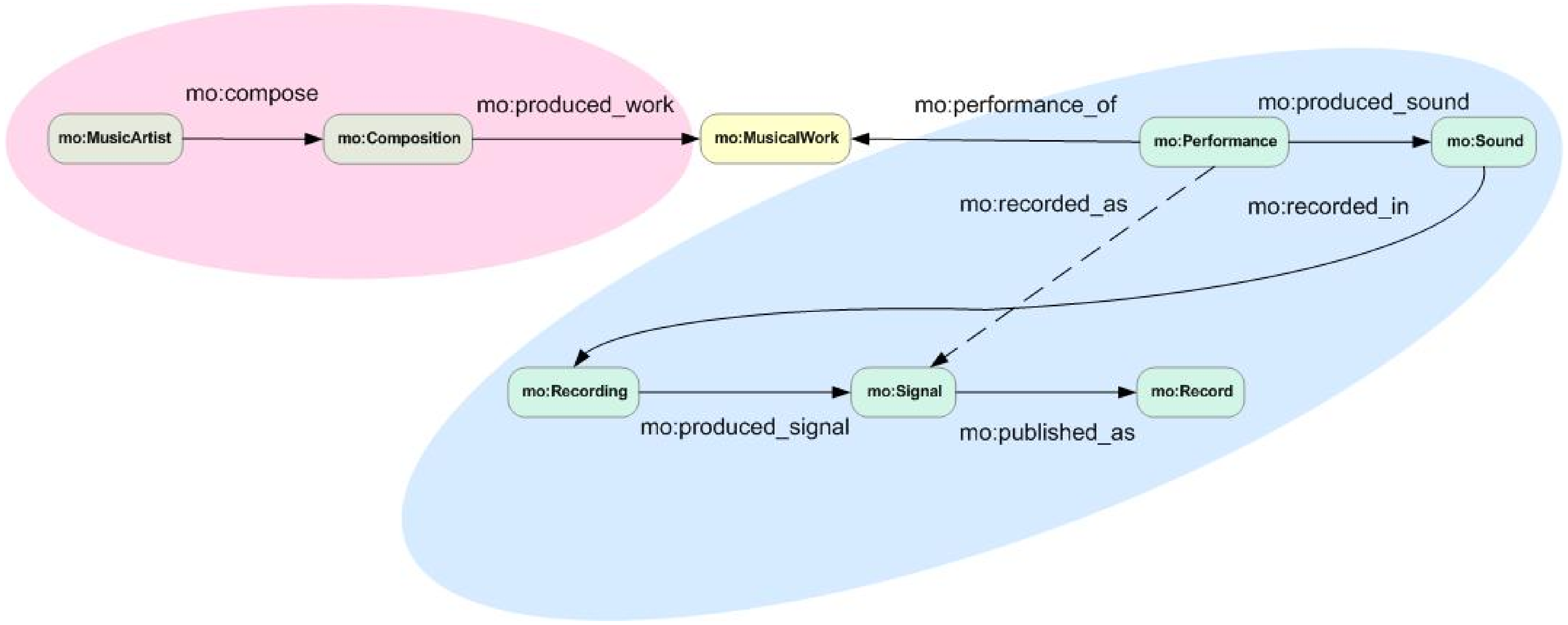
How do we express our ontologies?

(why, in RDF of course!)

RDFS and OWL

- RDFS: RDF Schema
  - The basics required to structure an ontology and exchange vocabularies
  - Classes and properties, super- and sub-classes, range and domain
- OWL: Web Ontology Language
  - More sophisticated structures
  - Constraints for existence and cardinality, transitive, inverse, symmetrical properties, ...

Ontology Examples from the Music Ontology (MO)



- MO extends and builds upon other ontologies: Event (and hence Timeline), FOAF, Dublin Core, ...

The Music Ontology, Yves Raimond *et al*, http://musicontology.com/

Ontology examples from MO (1)

```
mo:MusicArtist
    rdf:type owl:Class ;
    rdfs:isDefinedBy mo: ;
    rdfs:subClassOf foaf:Agent .


mo:AudioFile
    rdf:type owl:Class ;
    rdfs:isDefinedBy mo: ;
    rdfs:subClassOf mo:Medium , foaf:Document .
```

Warning! Curtailed for illustrative purposes.

Ontology examples from MO (2)

```
mo:AnalogSignal
    rdf:type owl:Class ;
    rdfs:isDefinedBy mo: ;
    rdfs:label "analogue signal" ;
    rdfs:subClassOf mo:Signal ;
    owl:disjointWith mo:DigitalSignal .


mo:Recording
    rdf:type owl:Class ;
    rdfs:isDefinedBy mo: ;
    rdfs:subClassOf event:Event .
```

Warning! Curtailed for illustrative purposes.

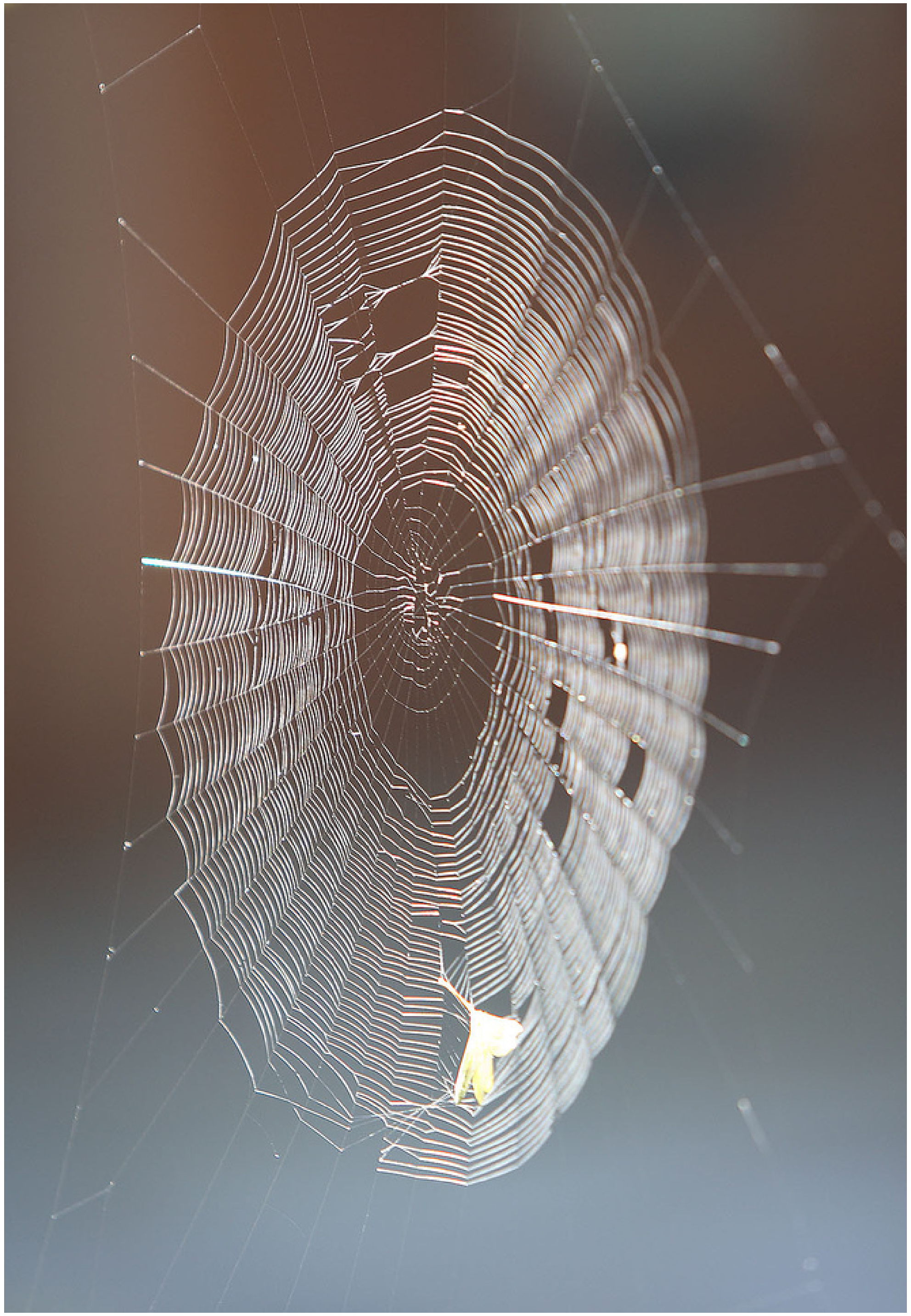
Ontology examples from MO (3)

```
mo:Track
    rdf:type owl:Class ;
    rdfs:subClassOf mo:MusicalManifestation .

mo:Record
    rdf:type owl:Class ;
    rdfs:subClassOf mo:MusicalManifestation .

mo:track
    rdf:type owl:ObjectProperty ;
    rdfs:domain mo:Record ;
    rdfs:range mo:Track .

mo:track_number
    rdf:type owl:DatatypeProperty ;
    rdfs:domain mo:Track ;
    rdfs:range xsd:nonNegativeInteger .
```

Warning! Curtailed for illustrative purposes.

Web Architecture

REST and building Resource Oriented systems



http://www.flickr.com/photos/curtsm/283309601/

- Not just *on* the Web...
- But to use and build upon Web Architecture to scale MIR systems

MIR *Systems*

- Beyond “just” the algorithmic extraction of information from music
- A combination of algorithms, techniques, and data sources
- ... and the use (and re-use) of results and analysis
- (But if you're not linking, this probably isn't for you...?)

MIR System Requirements

- Exchange of music is often restricted
  - licensing and copyright
  - quantity of data
- For comparative evaluation, data sets must be
  - widely shared
  - understood
  - re-usable
- But algorithm development is susceptible to overfitting

Existing MIR Systems

- A wide variety of languages, software engineering approaches, and architectures
- Often built to solve a particular MIR problem and expanded to address others
- Systems interaction through
  - plugins
  - shared libraries
  - syntactic serialisation and file exchange
  - some semantics used, but as an enhancement to traditional systems

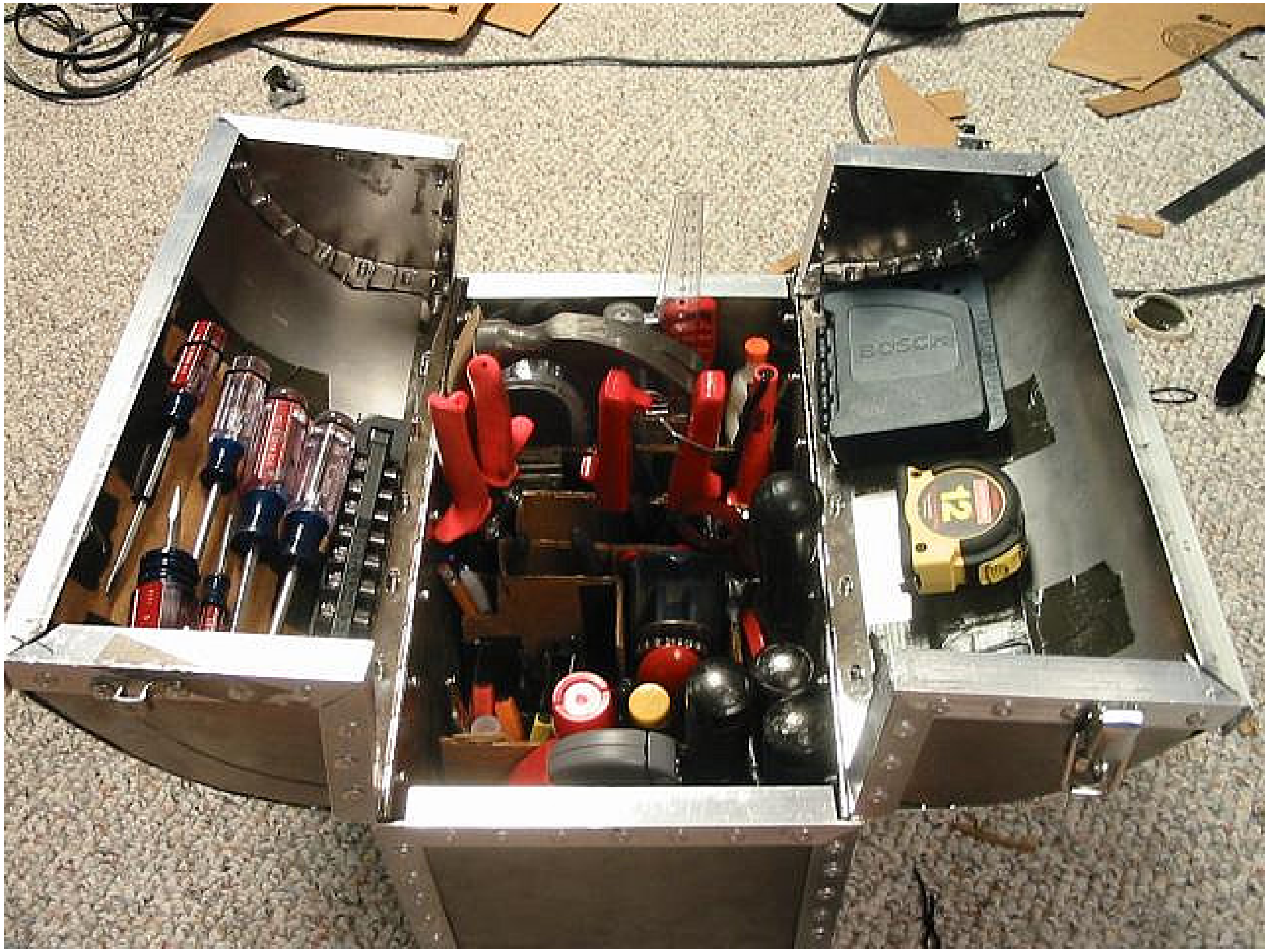
Advantages of adopting a Web Architecture

- Solutions exist for many problems that you can adopt or adapt; much of the stack already exists:
  - Encryption, identification, caching, content negotiation
- A well designed Web API is simple to use and quick to pick up
- A RESTful API is a good fit for Linked Data...

REST

- Roy Fielding, HTTP co-author, thesis “Architectural Styles and the Design of Network-based Software Architectures”
- In summary:
  - everything is a resource which is addressable
  - resources have multiple representations
  - relationships between resources are expressed through hyperlinks
  - all resources share a common interface with a limited set of operations
  - client server communication is stateless
  - (Not just RPC)

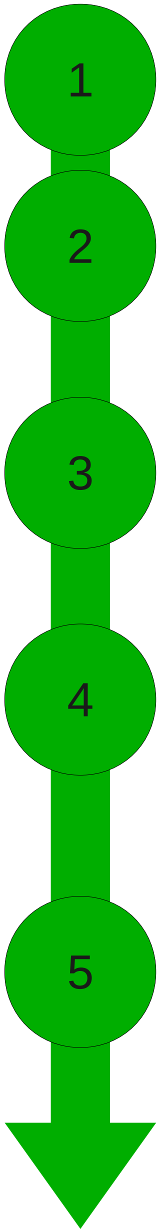
Putting this into practice



Implementing our simple example

ww.flickr.com/photos/zakh/337938459/

Our simple example



- Use artist and location metadata to select a collection(s) of audio
- Perform a genre classification over the collection
- Publish the genre analysis data with links back to the tracks (and so artist, location, and collections)
- Combine the results with other published metadata (about the artist, location, collections)

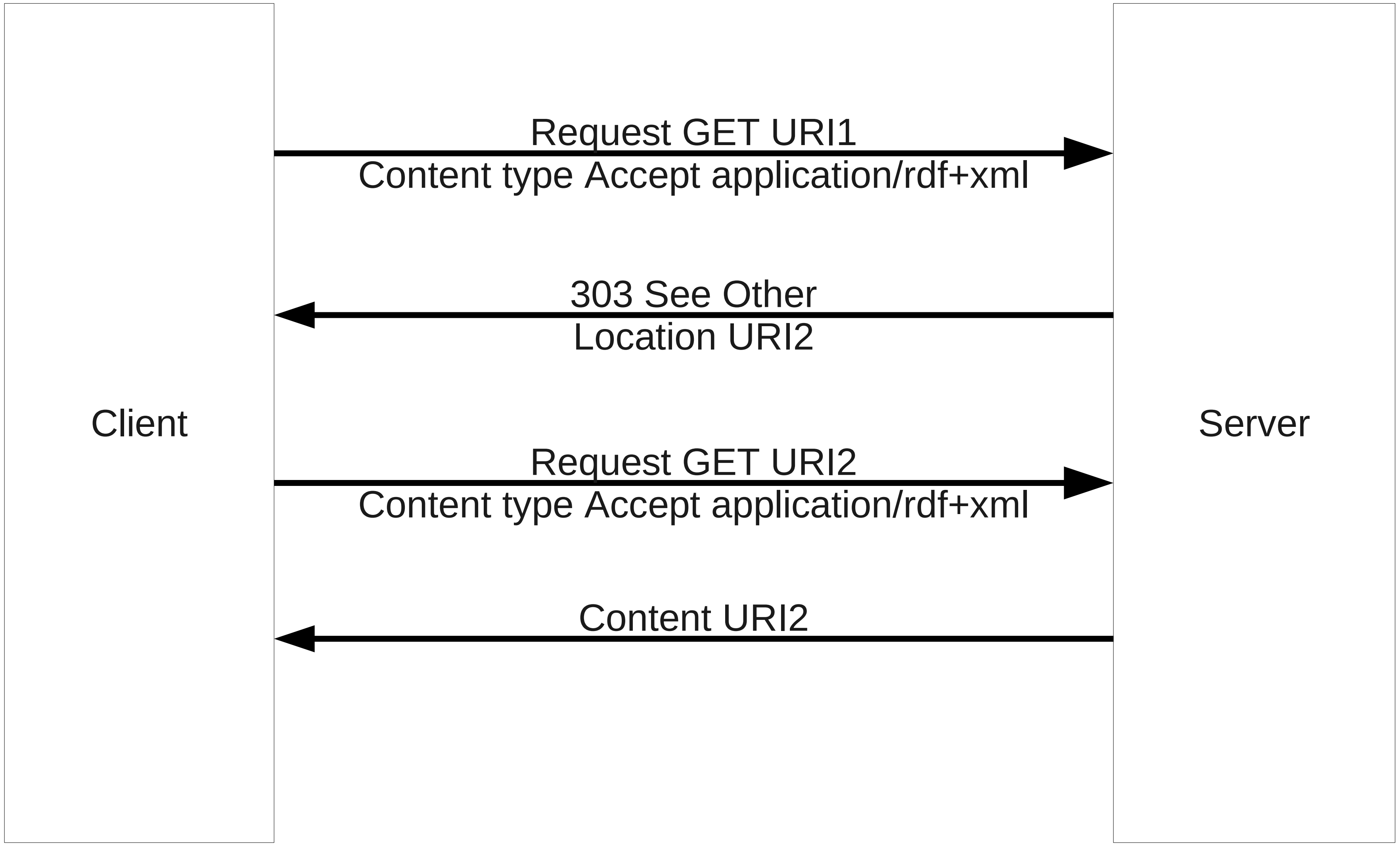
An Audio File Repository

- One service in our distributed architecture
- Provides audio files and metadata to an algorithm hosted by another service
- Each audio file is a resource with an HTTP URI
- A client can use HTTP content-negotiation to retrieve audio (mp3) or metadata (RDF)
- The RDF links each audio file to other data sources
- We also store the RDF in a triplestore and provide a SPARQL endpoint

“Minting” URIs

http://www.flickr.com/photos/a_mason/11071799/

Audio File Content Negotiation



Audio File RDF

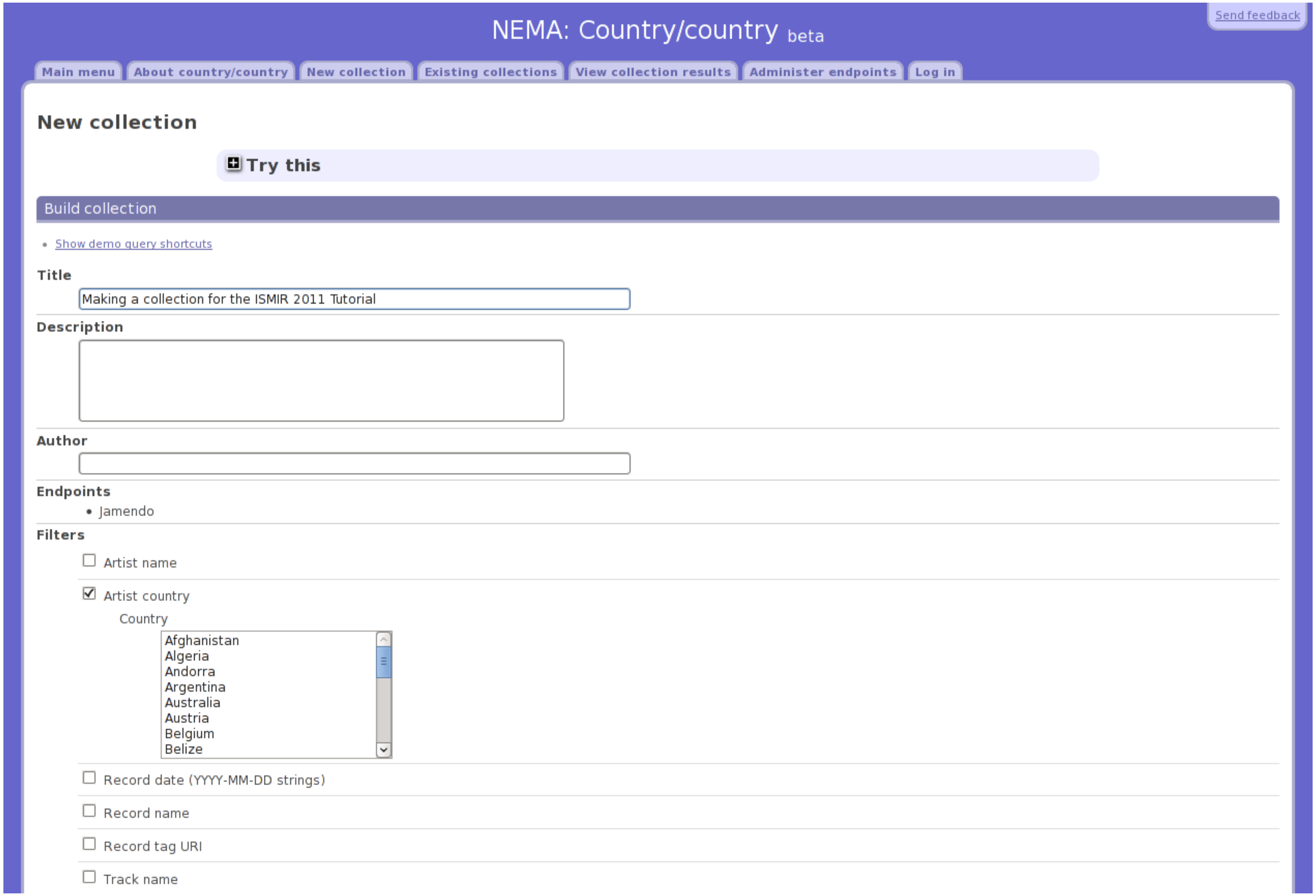
```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix mo: <http://purl.org/ontology/mo/> .


<http://jamendo.legacy.audiofiles.linkedmusic.org/audiofile/98933>
        rdf:type mo:AudioFile ;
        mo:encodes <http://dbtune.org/jamendo/signal/98933> .


<http://dbtune.org/jamendo/track/98933> rdf:type mo:Track ;
        mo:available_as
    <http://jamendo.legacy.audiofiles.linkedmusic.org/audiofile/98933> .
```

Example: Creating signal collections with SPARQL

A Collection Builder

- To begin with, we create our collections using metadata only
- We use SPARQL to query an endpoint for tracks by artists from a particular country
- We publish our collection as RDF
- We can then query our Audio File Repository to find any locally available signal in the collection, and publish this as a derivative collection

Collection Builder: SPARQL



New collection

Try this

Build collection

- Show demo query shortcuts

Title: Making a collection for the ISMIR 2011 Tutorial

Description:

Author:

Endpoints
- Jamendo

Filters

☐ Artist name

☑ Artist country

Country:
Afghanistan
Algeria
Andorra
Argentina
Australia
Austria
Belgium
Belize

☐ Record date (YYYY-MM-DD strings)

☐ Record name

☐ Record tag URI

☐ Track name

Collection Builder: RDF

```
@base <http://collections.nema.linkedmusic.org/signalcollection/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix ore: <http://www.openarchives.org/ore/terms/> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix pv: <http://purl.org/net/provenance/ns#> .

:68a959dcc1f68a950ded985b4a8007ee rdf:type ore:ResourceMap ;
    ore:describes :68a959dcc1f68a950ded985b4a8007ee#aggregate ;
    dc:creator
    <http://www.nema.ecs.soton.ac.uk/countrycountry/publishcollection> .

:68a959dcc1f68a950ded985b4a8007ee#aggregate rdf:type ore:Aggregate ,
                                                      pv:DataItem ;
    dc:title "ISMIR tutorial example collection" ;
    dc:creator <http://id.ecs.soton.ac.uk/person/4394> ;
    pv:createdBy :68a959dcc1f68a950ded985b4a8007ee#execution ;
    ore:aggregates <http://dbtune.org/jamendo/signal/7901> ,
                   <http://dbtune.org/jamendo/signal/7902> ,
                   <http://dbtune.org/jamendo/signal/7903> ,
```

Collection Builder: RDF

```
@base <http://collections.nema.linkedmusic.org/signalcollection/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix ore: <http://www.openarchives.org/ore/terms/> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix pv: <http://purl.org/net/provenance/ns#> .

:ExampleCollection rdf:type ore:ResourceMap ;
    ore:describes :ExampleCollection#aggregate ;
    dc:creator
    <http://www.nema.ecs.soton.ac.uk/countrycountry/publishcollection> .

:ExampleCollection#aggregate rdf:type ore:Aggregate ,
                                      pv:DataItem ;
    dc:title "ISMIR tutorial example collection" ;
    dc:creator <http://id.ecs.soton.ac.uk/person/4394> ;
    pv:createdBy :ExampleCollection#execution ;
    ore:aggregates <http://dbtune.org/jamendo/signal/7901> ,
                   <http://dbtune.org/jamendo/signal/7902> ,
                   <http://dbtune.org/jamendo/signal/7903> ,
```

Collection Builder: checking the Audio File Repository

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX mo: <http://purl.org/ontology/mo/>


SELECT ?audiofile WHERE {
    <http://dbtune.org/jamendo/track/98933>
        mo:available_as ?audiofile .

        ?audiofile a mo:AudioFile .

}
```

Hands on session 2

Content to negotiate and how to SPARQL

So far...

- The Audio File Repository provides locally available signal and RDF to describe it
- The Collection Builder uses SPARQL to create signal collections and find locally available audio files
- So our genre classifier has a collection of signal to analyse...

Example: Algorithm Output in RDF

Genre Classifier Output

- There is so much that could be captured and encoded here
  - Features, workflow, runtime parameters, execution metrics
- We will continue with a simplified example
- But encourage you to consider the possibilities!

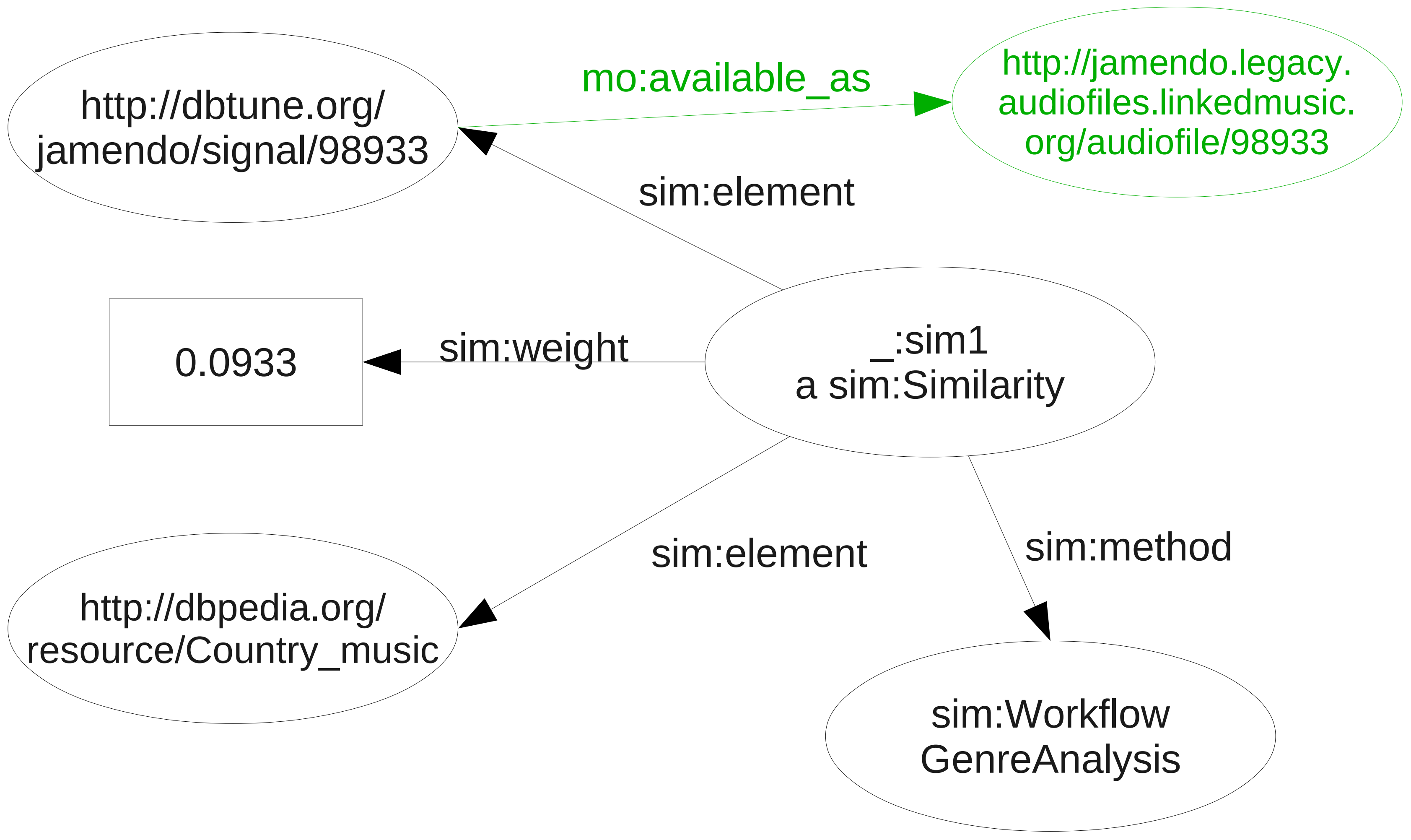
Genre Classifier Output

- Important Things to consider:
  - Maintaining references, through the Audio Files, to external sources of metadata about the signal
  - Making sure any output is modelled correctly
  - Aligning output with any existing relevant concepts (e.g. genre)
  - Providing means for others to re-use and build upon your output

Genre Classifier Output

- e.g.
  - We took the Downie lab's “Son of Blinkie” classifier which runs in the Meandre workflow framework
  - Used myExperiment to run the workflow over all signal in a collection as created by the Collection Builder
  - Modified the output to serialise as RDF with genres aligned (where possible) with dbpedia
  - RDF output is published and held in a triplestore

Genre Classifier Output: RDF


http://dbtune.org/ jamendo/signal/98933
mo:available_as
http://jamendo.legacy. audiofiles.linkedmusic. org/audiofile/98933
sim:element
0.0933
sim:weight
_:sim1 a sim:Similarity
sim:element
sim:method
http://dbpedia.org/ resource/Country_music
sim:Workflow GenreAnalysis

Warning: approximated for clarity!

Example: Viewing output and combining external data

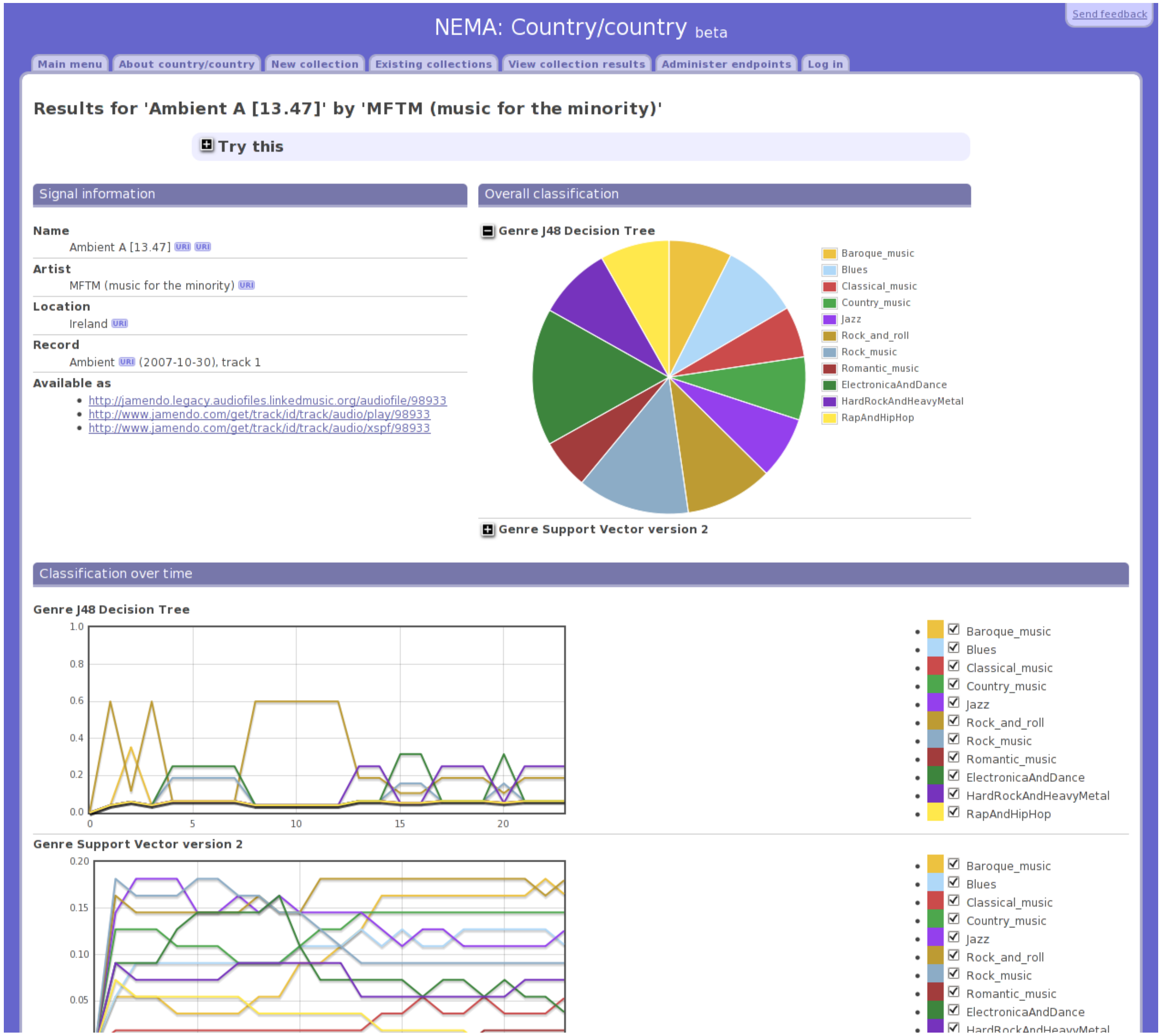
A Results Viewer

- The output from our genre classifier(s), the Collection Builder(s), and Audio File Repository(ies) share a common underlying data model, identifiers, and vocabularies
- The web of links between them allows us to combine their information in a Results Viewer
- Links to external data sources and use of established ontologies allows us to easily incorporate other data in the Results Viewer

Results Viewer

- e.g. find all genre analysis of tracks in a collection, where the collection is of artists from a specific country
- Provide and view examples of other artists in this genre
- For the most heavily weighted track in this genre find audio files for playback

Results Viewer



Results for 'Ambient A [13.47]' by 'MFTM (music for the minority)'

Try this

Signal information

Name
Ambient A [13.47] URI URI

Artist
MFTM (music for the minority) URI

Location
Ireland URI

Record
Ambient URI (2007-10-30), track 1

Available as
- http://jamendo.legacy.audiofiles.linkedmusic.org/audiofile/98933
- http://www.jamendo.com/get/track/id/track/audio/play/98933
- http://www.jamendo.com/get/track/id/track/audio/xspf/98933

Overall classification

Genre J48 Decision Tree

Baroque_music, Blues, Classical_music, Country_music, Jazz, Rock_and_roll, Rock_music, Romantic_music, ElectronicaAndDance, HardRockAndHeavyMetal, RapAndHipHop

Genre Support Vector version 2

Classification over time

Genre J48 Decision Tree

- ☑ Baroque_music
- ☑ Blues
- ☑ Classical_music
- ☑ Country_music
- ☑ Jazz
- ☑ Rock_and_roll
- ☑ Rock_music
- ☑ Romantic_music
- ☑ ElectronicaAndDance
- ☑ HardRockAndHeavyMetal
- ☑ RapAndHipHop

Genre Support Vector version 2

- ☑ Baroque_music
- ☑ Blues
- ☑ Classical_music
- ☑ Country_music
- ☑ Jazz
- ☑ Rock_and_roll
- ☑ Rock_music
- ☑ Romantic_music
- ☑ ElectronicaAndDance
- ☑ HardRockAndHeavyMetal

Results Viewer SPARQL (1)

```
PREFIX dbpedia-owl: http://dbpedia.org/ontology/
PREFIX foaf: http://xmlns.com/foaf/0.1/

SELECT ?artist ?artistname ?place ?placename WHERE {

        ?artist dbpedia-owl:genre <$genreuri> .
        {
            ?artist a dbpedia-owl:Band ;
                dbpedia-owl:hometown ?place .
        } UNION {
            ?artist a dbpedia-owl:MusicalArtist ;
                dbpedia-owl:birthPlace ?place .
        }

        ?artist foaf:name ?artistname .
        ?place foaf:name ?placename .
}
```

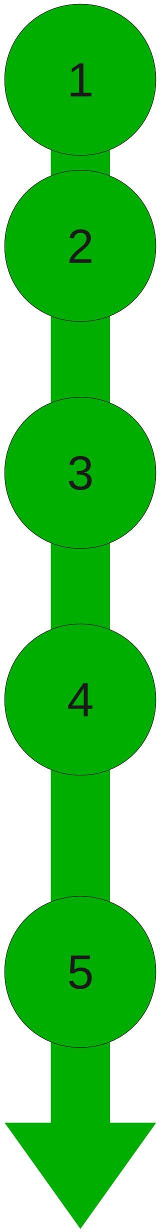
Endpoint: http://dbpedia.org/sparql/

Results Viewer SPARQL (2)

```
PREFIX owl: http://www.w3.org/2002/07/owl#
SELECT * WHERE {

        ?bbcuri owl:sameAs <" . $dbpediauri . "> .

        FILTER regex(str(?bbcuri), \"^http://www.bbc.co.uk/\") .
}
```

Endpoint: http://api.talis.com/stores/bbc-backstage/services/sparql

Summary of our simple example

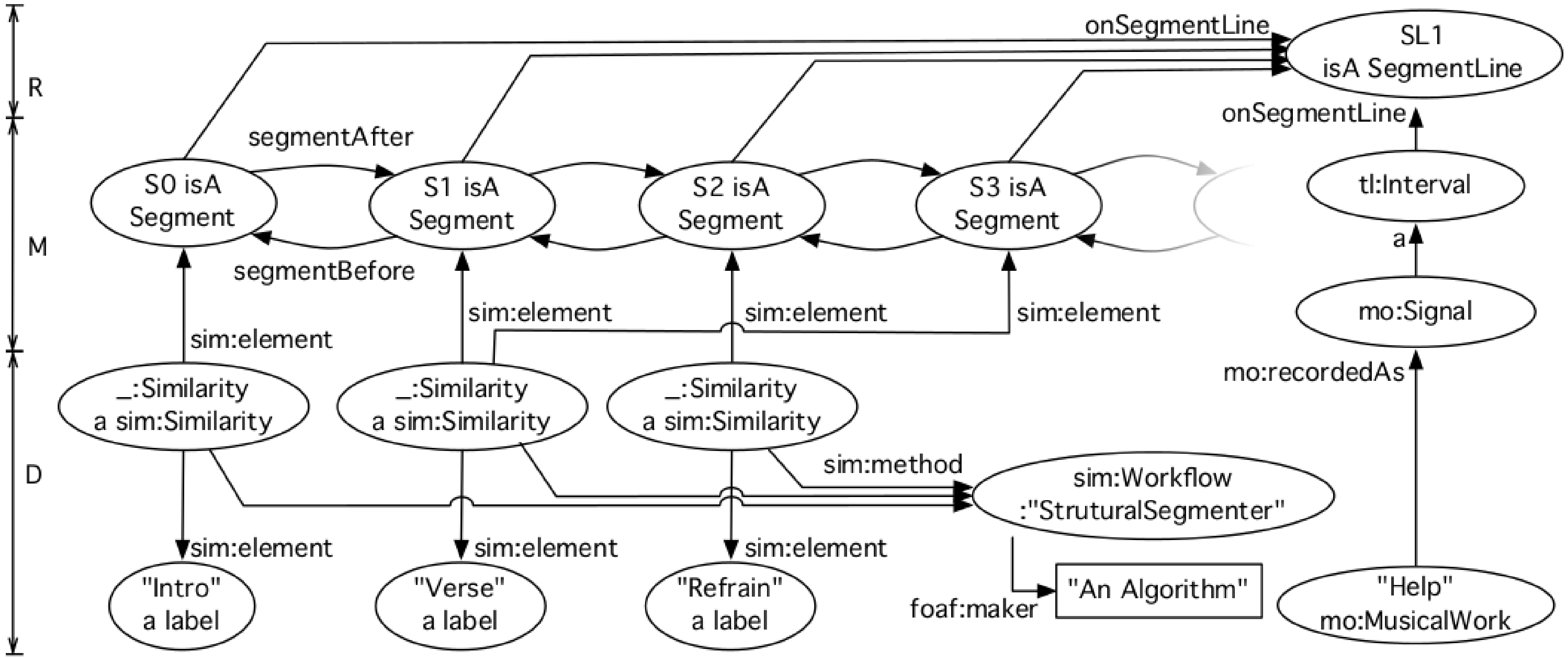


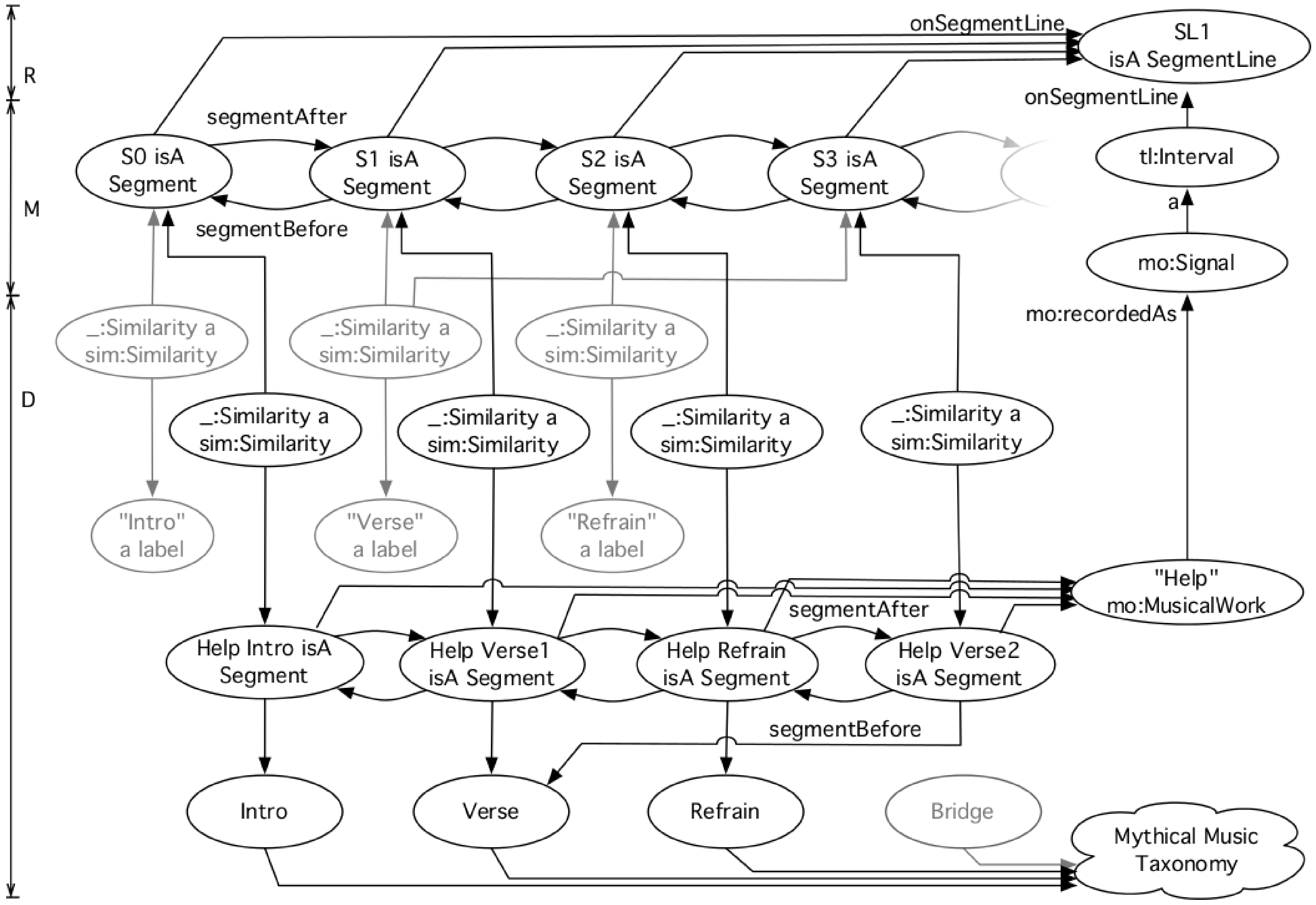
- Use artist and location metadata to select a collection(s) of audio
- Perform a genre classification over the collection
- Publish the genre analysis data with links back to the tracks (and so artist, location, and collections)
- Combine the results with other published metadata (about the artist, location, collections)

Following the rabbit...

...just a little way down the hole

The Segment Ontology



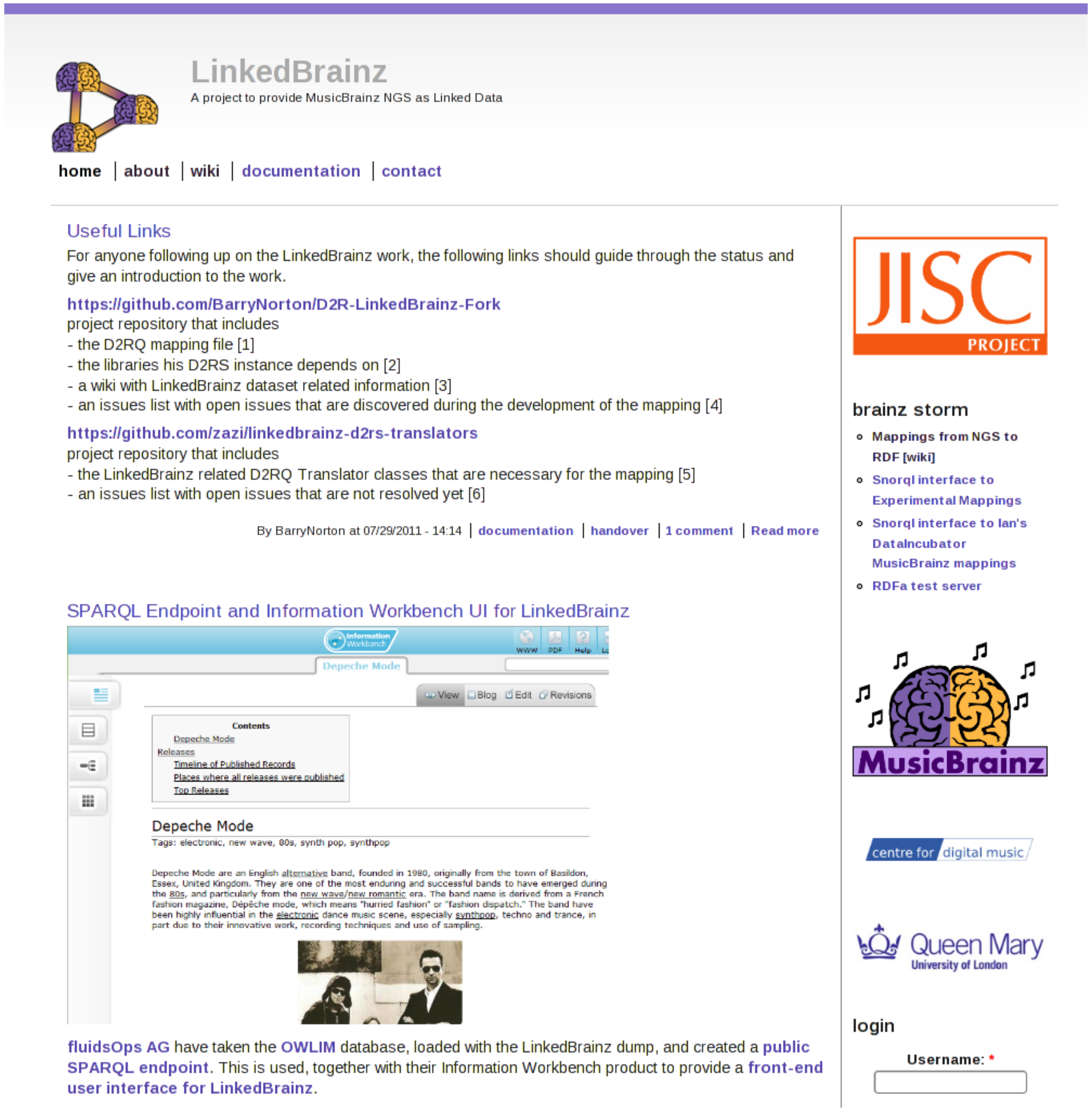

R
M
D
onSegmentLine
SL1 isA SegmentLine
onSegmentLine
segmentAfter
S0 isA Segment
S1 isA Segment
S2 isA Segment
S3 isA Segment
segmentBefore
tl:Interval
a
mo:Signal
mo:recordedAs
_:Similarity a sim:Similarity
"Intro" a label
"Verse" a label
"Refrain" a label
"Help" mo:MusicalWork
segmentAfter
Help Intro isA Segment
Help Verse1 isA Segment
Help Refrain isA Segment
Help Verse2 isA Segment
segmentBefore
Intro
Verse
Refrain
Bridge
Mythical Music Taxonomy

Linking Up

Where to go from here?

Musicbrainz & Linkedbrainz

- Aligning Musicbrainz NGS with Linked Data
- Centre for Digital Music at Queen Mary University of London



LinkedBrainz
A project to provide MusicBrainz NGS as Linked Data

home | about | wiki | documentation | contact

Useful Links

For anyone following up on the LinkedBrainz work, the following links should guide through the status and give an introduction to the work.

https://github.com/BarryNorton/D2R-LinkedBrainz-Fork
project repository that includes
- the D2RQ mapping file [1]
- the libraries his D2RS instance depends on [2]
- a wiki with LinkedBrainz dataset related information [3]
- an issues list with open issues that are discovered during the development of the mapping [4]

https://github.com/zazi/linkedbrainz-d2rs-translators
project repository that includes
- the LinkedBrainz related D2RQ Translator classes that are necessary for the mapping [5]
- an issues list with open issues that are not resolved yet [6]

By BarryNorton at 07/29/2011 - 14:14 | documentation | handover | 1 comment | Read more

SPARQL Endpoint and Information Workbench UI for LinkedBrainz

Depeche Mode

Contents
Depeche Mode
Releases
Timeline of Published Records
Places where all releases were published
Top Releases

Depeche Mode
Tags: electronic, new wave, 80s, synth pop, synthpop

Depeche Mode are an English alternative band, founded in 1980, originally from the town of Basildon, Essex, United Kingdom. They are one of the most enduring and successful bands to have emerged during the 80s, and particularly from the new wave/new romantic era. The band name is derived from a French fashion magazine, Dépêche mode, which means "hurried fashion" or "fashion dispatch." The band have been highly influential in the electronic dance music scene, especially synthpop, techno and trance, in part due to their innovative work, recording techniques and use of sampling.

fluidsOps AG have taken the OWLIM database, loaded with the LinkedBrainz dump, and created a public SPARQL endpoint. This is used, together with their Information Workbench product to provide a front-end user interface for LinkedBrainz.

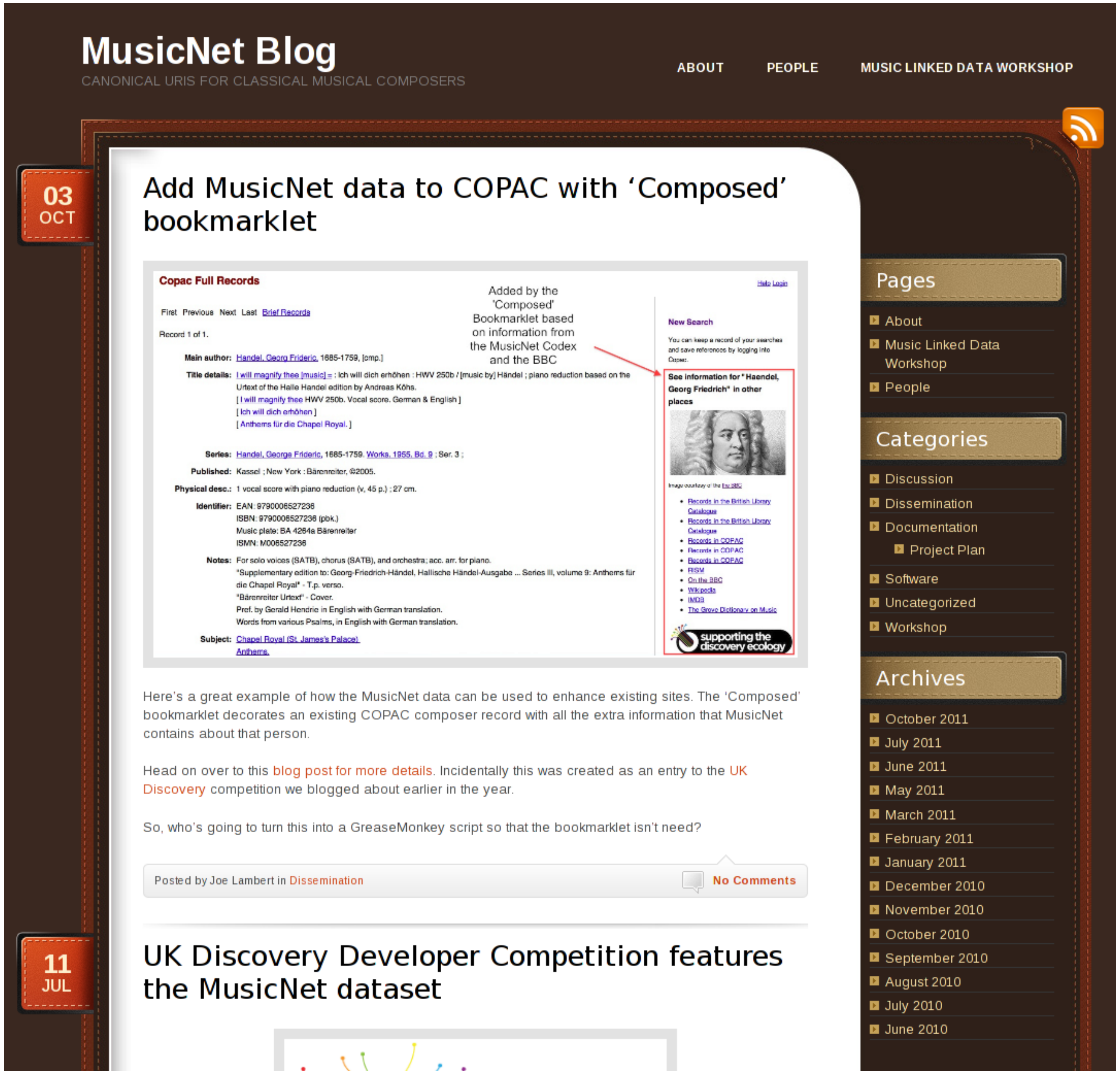
brainz storm
- Mappings from NGS to RDF [wiki]
- Snorql interface to Experimental Mappings
- Snorql interface to Ian's DataIncubator MusicBrainz mappings
- RDFa test server

login
Username: *

http://linkedbrainz.c4dmpresents.org/

Musicnet

- Canonical URIs for composers
- Alignment tool for managing co-references
- University of Southampton



MusicNet Blog

CANONICAL URIS FOR CLASSICAL MUSICAL COMPOSERS

ABOUT PEOPLE MUSIC LINKED DATA WORKSHOP

03 OCT

Add MusicNet data to COPAC with 'Composed' bookmarklet

Copac Full Records

Help Login

First Previous Next Last Brief Records

Record 1 of 1.

Main author:	Handel, Georg Frideric, 1685-1759, [cmp.]
Title details:	I will magnify thee [music] = : Ich will dich erhöhen : HWV 250b / [music by] Händel ; piano reduction based on the Urtext of the Halle Handel edition by Andreas Köhs. [I will magnify thee HWV 250b. Vocal score. German & English] [Ich will dich erhöhen] [Anthems für die Chapel Royal.]
Series:	Handel, George Frideric, 1685-1759. Works. 1955. Bd. 9 ; Ser. 3 ;
Published:	Kassel ; New York : Bärenreiter, ©2005.
Physical desc.:	1 vocal score with piano reduction (v, 45 p.) ; 27 cm.
Identifier:	EAN: 9790006527236 ISBN: 9790006527236 (pbk.) Music plate: BA 4264a Bärenreiter ISMN: M006527236
Notes:	For solo voices (SATB), chorus (SATB), and orchestra; acc. arr. for piano. "Supplementary edition to: Georg-Friedrich-Händel, Hallische Händel-Ausgabe ... Series III, volume 9: Anthems für die Chapel Royal" - T.p. verso. "Bärenreiter Urtext" - Cover. Pref. by Gerald Hendrie in English with German translation. Words from various Psalms, in English with German translation.
Subject:	Chapel Royal (St. James's Palace) Anthems.

Added by the 'Composed' Bookmarklet based on information from the MusicNet Codex and the BBC

New Search

You can keep a record of your searches and save references by logging into Copac.

See information for " Haendel, Georg Friedrich" in other places

Image courtesy of the the BBC

- Records in the British Library Catalogue
- Records in the British Library Catalogue
- Records in COPAC
- Records in COPAC
- Records in COPAC
- RISM
- On the BBC
- Wikipedia
- IMDB
- The Grove Dictionary on Music

supporting the discovery ecology

Here's a great example of how the MusicNet data can be used to enhance existing sites. The 'Composed' bookmarklet decorates an existing COPAC composer record with all the extra information that MusicNet contains about that person.

Head on over to this blog post for more details. Incidentally this was created as an entry to the UK Discovery competition we blogged about earlier in the year.

So, who's going to turn this into a GreaseMonkey script so that the bookmarklet isn't need?

Posted by Joe Lambert in Dissemination No Comments

11 JUL

UK Discovery Developer Competition features the MusicNet dataset

Pages

- About
- Music Linked Data Workshop
- People

Categories

- Discussion
- Dissemination
- Documentation
  - Project Plan
- Software
- Uncategorized
- Workshop

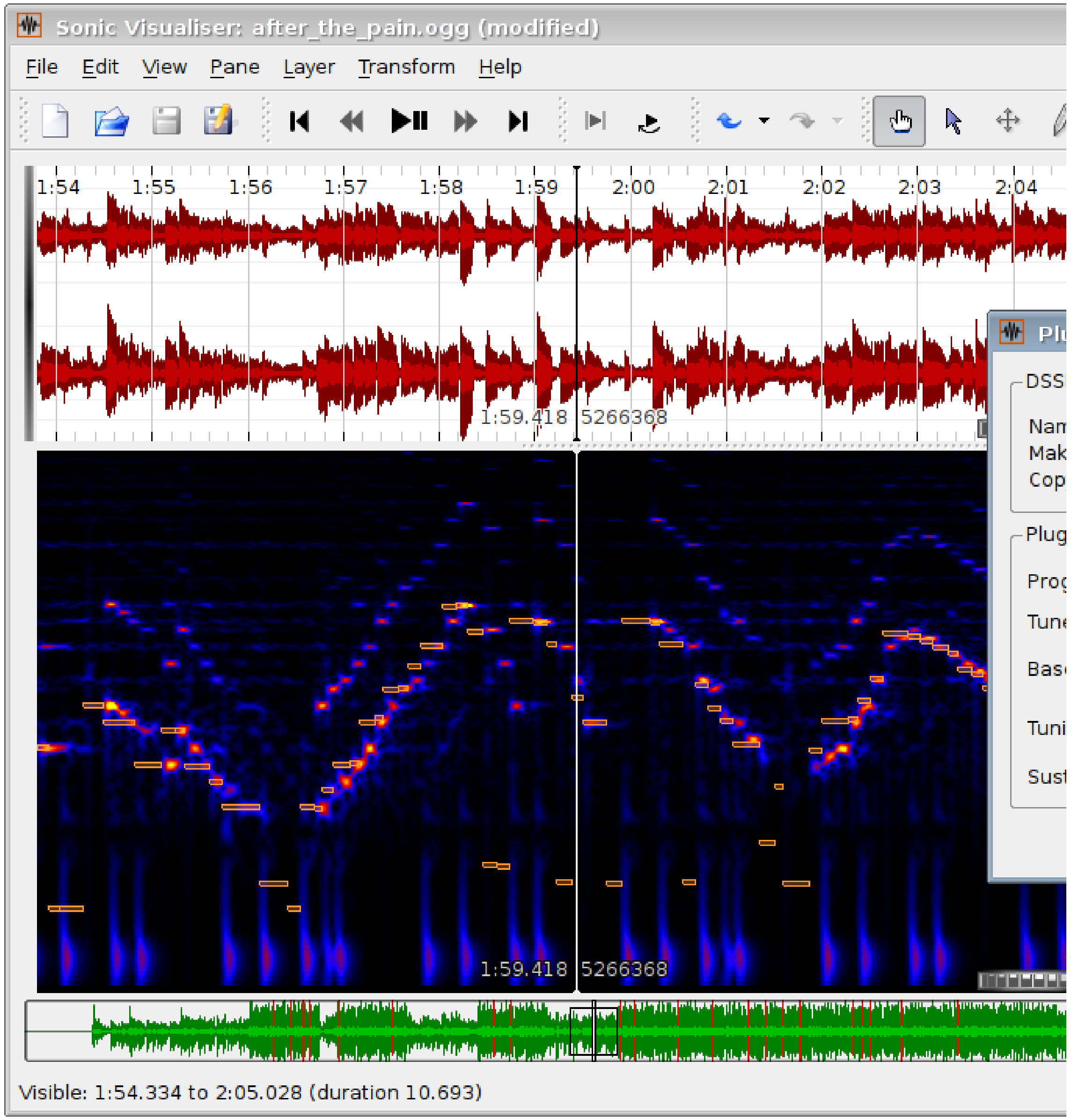
Archives

- October 2011
- July 2011
- June 2011
- May 2011
- March 2011
- February 2011
- January 2011
- December 2010
- November 2010
- October 2010
- September 2010
- August 2010
- July 2010
- June 2010

http://musicnet.mspace.fm/

Sonic Visualiser & Annotator

- Uses RDF for serialisation
- Numerous associated ontologies for tasks
- Centre for Digital Music at Queen Mary University of London

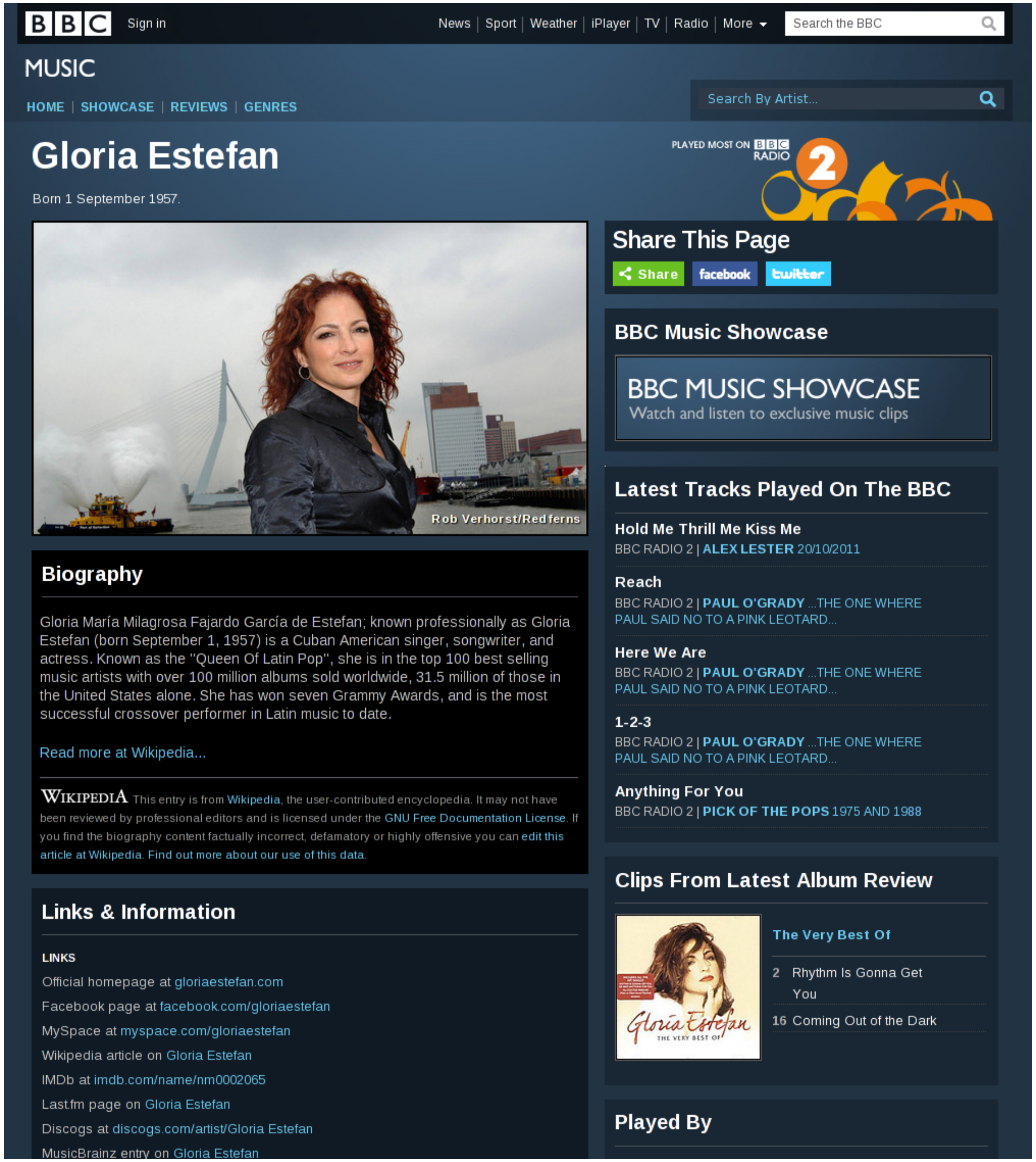


Visible: 1:54.334 to 2:05.028 (duration 10.693)

http://www.sonicvisualiser.org/

BBC Music

- Pages generated using, and available as, RDF
- Linked to other data sources


BBC
Sign in
News | Sport | Weather | iPlayer | TV | Radio | More
Search the BBC
MUSIC
HOME | SHOWCASE | REVIEWS | GENRES
Search By Artist...
PLAYED MOST ON BBC RADIO 2
Rob Verhorst/Redferns
Share
facebook
twitter

Gloria Estefan

Born 1 September 1957.

Share This Page

BBC Music Showcase

BBC MUSIC SHOWCASE
Watch and listen to exclusive music clips

Biography

Gloria María Milagrosa Fajardo García de Estefan; known professionally as Gloria Estefan (born September 1, 1957) is a Cuban American singer, songwriter, and actress. Known as the "Queen Of Latin Pop", she is in the top 100 best selling music artists with over 100 million albums sold worldwide, 31.5 million of those in the United States alone. She has won seven Grammy Awards, and is the most successful crossover performer in Latin music to date.

Read more at Wikipedia...

WIKIPEDIA This entry is from Wikipedia, the user-contributed encyclopedia. It may not have been reviewed by professional editors and is licensed under the GNU Free Documentation License. If you find the biography content factually incorrect, defamatory or highly offensive you can edit this article at Wikipedia. Find out more about our use of this data.

Links & Information

LINKS

Official homepage at gloriaestefan.com
Facebook page at facebook.com/gloriaestefan
MySpace at myspace.com/gloriaestefan
Wikipedia article on Gloria Estefan
IMDb at imdb.com/name/nm0002065
Last.fm page on Gloria Estefan
Discogs at discogs.com/artist/Gloria Estefan
MusicBrainz entry on Gloria Estefan

Latest Tracks Played On The BBC

Hold Me Thrill Me Kiss Me
BBC RADIO 2 | ALEX LESTER 20/10/2011

Reach
BBC RADIO 2 | PAUL O'GRADY ...THE ONE WHERE PAUL SAID NO TO A PINK LEOTARD...

Here We Are
BBC RADIO 2 | PAUL O'GRADY ...THE ONE WHERE PAUL SAID NO TO A PINK LEOTARD...

1-2-3
BBC RADIO 2 | PAUL O'GRADY ...THE ONE WHERE PAUL SAID NO TO A PINK LEOTARD...

Anything For You
BBC RADIO 2 | PICK OF THE POPS 1975 AND 1988

Clips From Latest Album Review

The Very Best Of

2 Rhythm Is Gonna Get You
16 Coming Out of the Dark

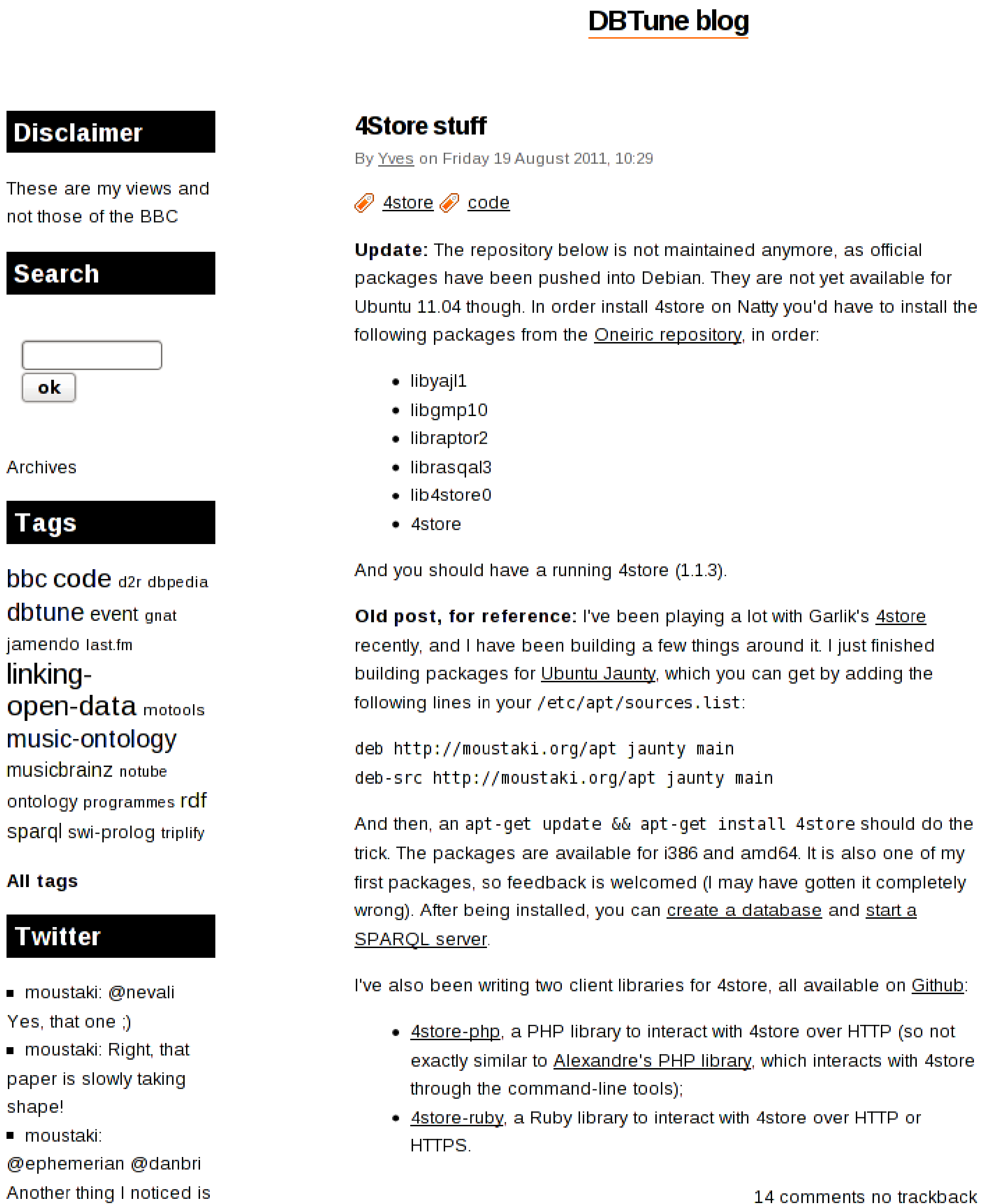
Played By

http://www.bbc.co.uk/music

dbtune.org/*

- Several music related datasets converted and published as Linked Data
- Linked to other data sources
- SPARQL endpoints

http://dbtune.org/



DBTune blog

Disclaimer

These are my views and not those of the BBC

Search

ok

Archives

Tags

bbc code d2r dbpedia dbtune event gnat jamendo last.fm linking-open-data motools music-ontology musicbrainz notube ontology programmes rdf sparql swi-prolog triplify

All tags

Twitter

- moustaki: @nevali Yes, that one ;)
- moustaki: Right, that paper is slowly taking shape!
- moustaki: @ephemerian @danbri Another thing I noticed is

4Store stuff

By Yves on Friday 19 August 2011, 10:29

4store code

Update: The repository below is not maintained anymore, as official packages have been pushed into Debian. They are not yet available for Ubuntu 11.04 though. In order install 4store on Natty you'd have to install the following packages from the Oneiric repository, in order:

- libyajl1
- libgmp10
- libraptor2
- librasqal3
- lib4store0
- 4store

And you should have a running 4store (1.1.3).

Old post, for reference: I've been playing a lot with Garlik's 4store recently, and I have been building a few things around it. I just finished building packages for Ubuntu Jaunty, which you can get by adding the following lines in your `/etc/apt/sources.list`:

```
deb http://moustaki.org/apt jaunty main
deb-src http://moustaki.org/apt jaunty main
```

And then, an `apt-get update && apt-get install 4store` should do the trick. The packages are available for i386 and amd64. It is also one of my first packages, so feedback is welcomed (I may have gotten it completely wrong). After being installed, you can create a database and start a SPARQL server.

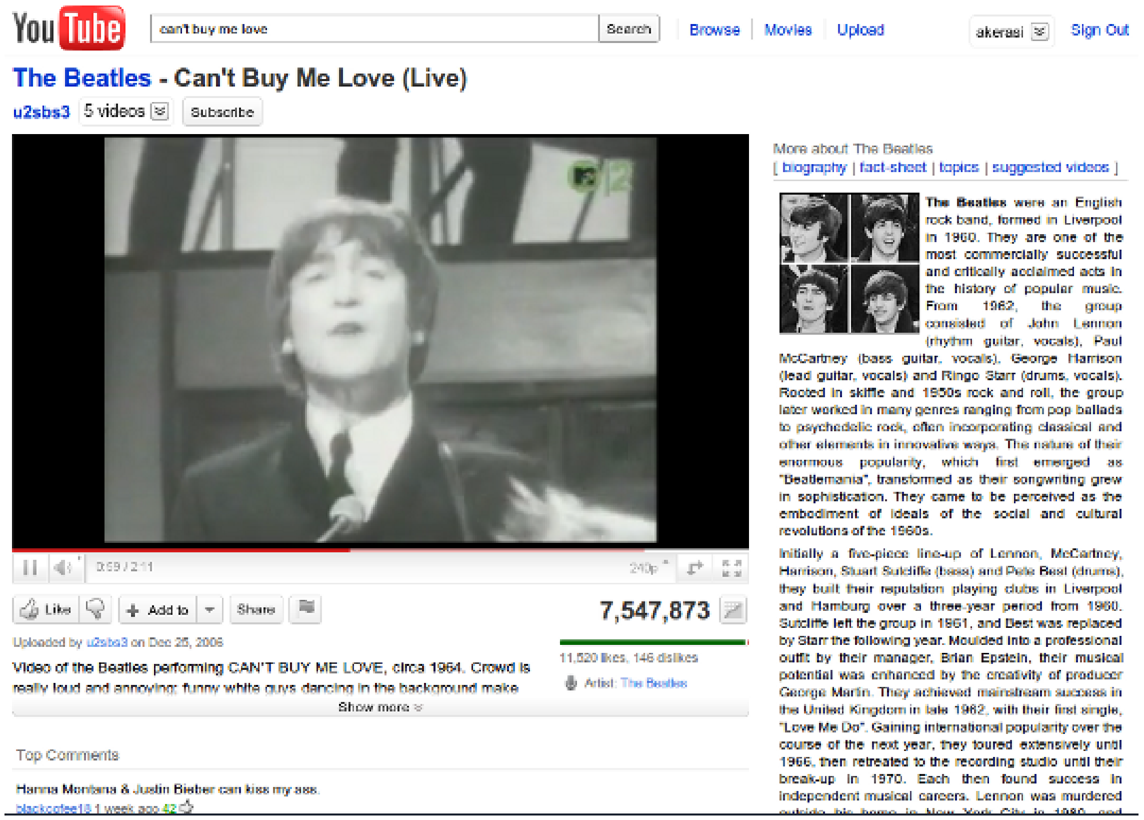
I've also been writing two client libraries for 4store, all available on Github:

- 4store-php, a PHP library to interact with 4store over HTTP (so not exactly similar to Alexandre's PHP library, which interacts with 4store through the command-line tools);
- 4store-ruby, a Ruby library to interact with 4store over HTTP or HTTPS.

14 comments no trackback

seevl

- Contextube plugin for YouTube utilises Linked Data
- They have an API available



The Beatles - Can't Buy Me Love (Live)

Uploaded by u2sbs3 on Dec 25, 2006

Video of the Beatles performing CAN'T BUY ME LOVE, circa 1964. Crowd is really loud and annoying; funny white guys dancing in the background make

11,520 likes, 146 dislikes

Artist: The Beatles

Top Comments

Hanna Montana & Justin Bieber can kiss my ass.
blackcofee18 1 week ago 42

More about The Beatles
[biography | fact-sheet | topics | suggested videos]

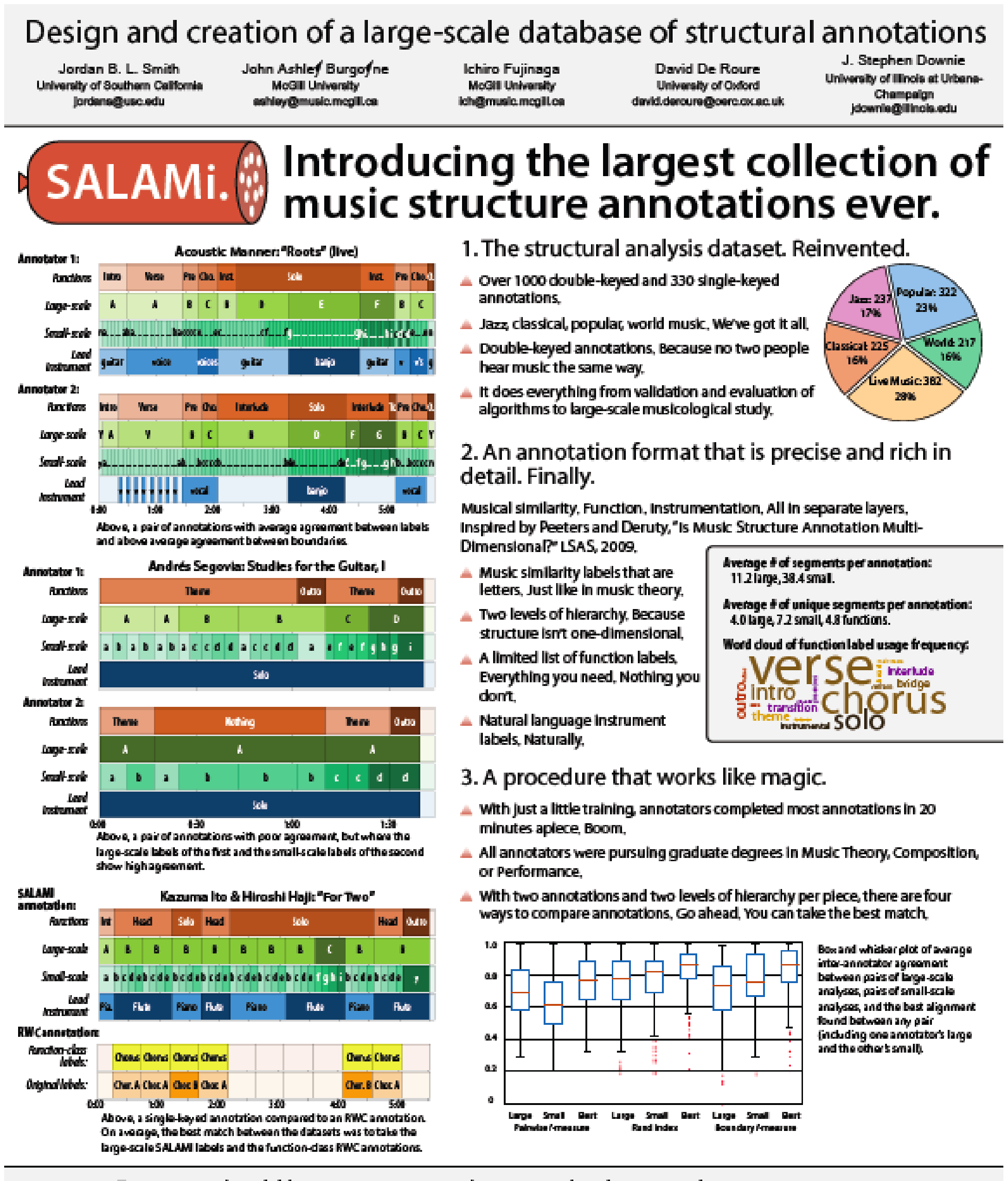
The Beatles were an English rock band, formed in Liverpool in 1960. They are one of the most commercially successful and critically acclaimed acts in the history of popular music. From 1962, the group consisted of John Lennon (rhythm guitar, vocals), Paul McCartney (bass guitar, vocals), George Harrison (lead guitar, vocals) and Ringo Starr (drums, vocals). Rooted in skiffle and 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 innovative ways. The nature of their enormous popularity, which first emerged as "Beatlemania", transformed as their songwriting grew in sophistication. They came to be perceived as the embodiment of ideals of the social and cultural revolutions of the 1960s.

Initially a five-piece line-up of Lennon, McCartney, Harrison, Stuart Sutcliffe (bass) and Pete Best (drums), they built their reputation playing clubs in Liverpool and Hamburg over a three-year period from 1960. Sutcliffe left the group in 1961, and Best was replaced by Starr the following year. Moulded into a professional outfit by their manager, Brian Epstein, their musical potential was enhanced by the creativity of producer George Martin. They achieved mainstream success in the United Kingdom in late 1962, with their first single, "Love Me Do". Gaining international popularity over the course of the next year, they toured extensively until 1966, then retreated to the recording studio until their break-up in 1970. Each then found success in independent musical careers. Lennon was murdered

http://seevl.net/

SALAMI

- *Watch this space!*



Design and creation of a large-scale database of structural annotations

Jordan B. L. Smith, University of Southern California, jordans@usc.edu
John Ashley Burgoyne, McGill University, ashley@music.mcgill.ca
Ichiro Fujinaga, McGill University, ich@music.mcgill.ca
David De Roure, University of Oxford, david.deroure@oerc.ox.ac.uk
J. Stephen Downie, University of Illinois at Urbana-Champaign, jdownie@illinois.edu

SALAMi. Introducing the largest collection of music structure annotations ever.

Above, a pair of annotations with average agreement between labels and above average agreement between boundaries.

Above, a pair of annotations with poor agreement, but where the large-scale labels of the first and the small-scale labels of the second show high agreement.

Above, a single-keyed annotation compared to an RWC annotation. On average, the best match between the datasets was to take the large-scale SALAMI labels and the function-class RWC annotations.

1. The structural analysis dataset. Reinvented.

- Over 1000 double-keyed and 330 single-keyed annotations.
- Jazz, classical, popular, world music. We've got it all.
- Double-keyed annotations. Because no two people hear music the same way.
- It does everything from validation and evaluation of algorithms to large-scale musicological study.

2. An annotation format that is precise and rich in detail. Finally.

Musical similarity. Function. Instrumentation. All in separate layers. Inspired by Peeters and Deruty, "Is Music Structure Annotation Multi-Dimensional?" LSAS, 2009.

- Music similarity labels that are letters. Just like in music theory.
- Two levels of hierarchy. Because structure isn't one-dimensional.
- A limited list of function labels. Everything you need. Nothing you don't.
- Natural language instrument labels. Naturally.

Average # of segments per annotation: 11.2 large, 38.4 small.

Average # of unique segments per annotation: 4.0 large, 7.2 small, 4.8 functions.

Word cloud of function label usage frequency:

3. A procedure that works like magic.

- With just a little training, annotators completed most annotations in 20 minutes apiece. Boom.
- All annotators were pursuing graduate degrees in Music Theory, Composition, or Performance.
- With two annotations and two levels of hierarchy per piece, there are four ways to compare annotations. Go ahead. You can take the best match.

Box and whisker plot of average inter-annotator agreement between pairs of large-scale analyses, pairs of small-scale analyses, and the best alignment found between any pair (including one annotator's large and the other's small).

ISMIR 2011

- Knowledge Representation Issues in Musical Instrument Ontology Design
  - Sefki Kolozali, Mathieu Barthet, György Fazekas, and Mark Sandler

- The Studio Ontology Framework
  - György Fazekas and Mark Sandler

(Poster session 3)

Summary: what you can do

- Linked Data for music is already out there – can it help you in your research?
  - Can you link any of your data to it?
- Can you publish your research output as Linked Data?
  - Perhaps not everything to start with! Every little helps...
  - Are there any existing ontologies available? Can you extend any?
  - Can you work with others to build a common ontology?
- Does existing Linked Data provide an interesting context in which to present and analyse your output?

Summary: why you should do it

- Large datasets are increasingly available, relevant, and should be a fantastic opportunity to apply and refine MIR
- MIR *systems* need to scale to do this – Linked Data and Web Architecture can help you scale
- Domain specific tooling will continue to develop as interest grows; make it work for you
- There's lots of great Linked Data out there already, but every little bit more helps!
    - Linking your data to existing sources makes it more useful and relevant to others – there is great opportunity to reach out to other disciplines

Thanks and acknowledgements

- “Country/country” demonstrator system
  - Bart Nagel, Gianni O'Neill
- Segment Ontology
  - Ben Fields
- SALAMI PIs
  - David De Roure, J. Stephen Downie, Ichiro Fujinaga
- Funders
  - NEMA (Andrew W. Mellon Foundation)
  - SALAMI (JISC, NSERC, NSF)
  - e-Research South (EPSRC grant EP/F05811X/1)