Semantic Web technologies

and Linked Data

Transforming Musicology mini-projects workshop

Lancaster University, 12th February 2014

Kevin Page

Oxford e-Research Centre, University of Oxford.

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

Overview

- Motivation
  - In general, for (computational) musicology, and for Digital Humanities
- Our approach
  - Embracing Web architecture, the Semantic Web, and Linked Data
- A small example

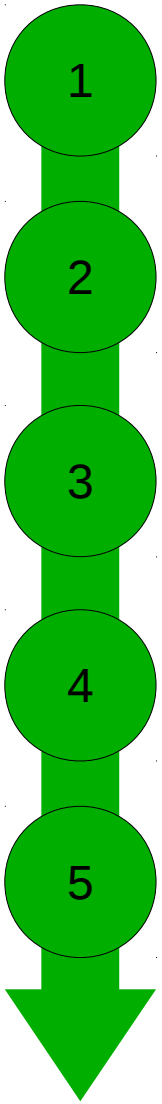
Motivation: in general

- When knowledge has been generated, we should capitalise on its value by
  - capturing it
  - publishing it
  - using it
  - linking it
  - re-using and building upon it (“unintentionally”?)

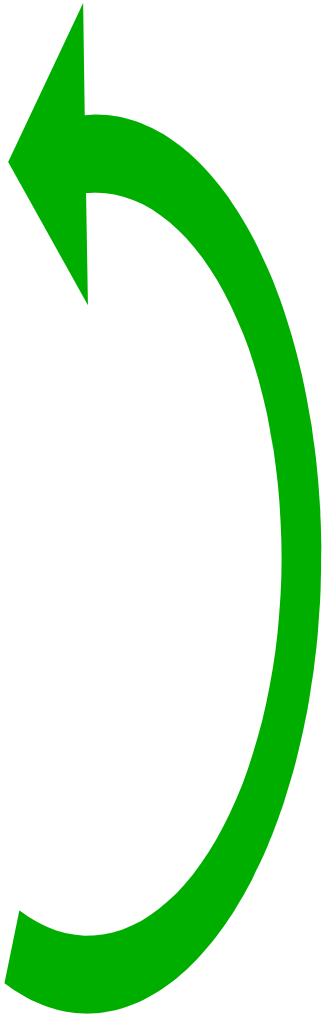
A scholarly process

A digital transformation of this scholarly process

A scholar using digital methods might...



- Building upon previous method 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

On the Web...? (technically speaking!)

Don't just put Digital Humanities content *on* the Web...

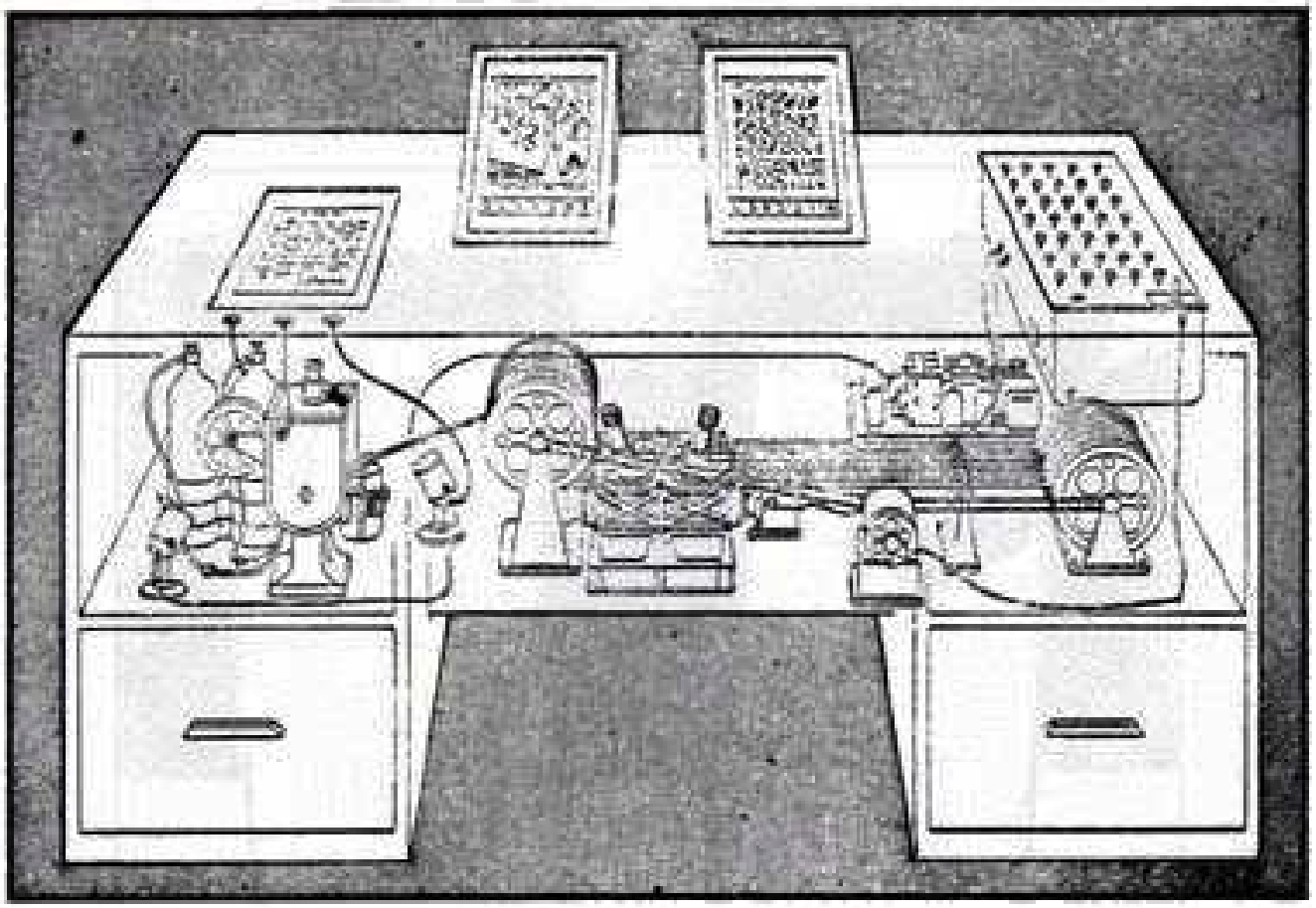
...but use and build upon Web Architecture to scale Digital Humanities activity

The value is in the linking.

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

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

Tim Berners-Lee and the World Wide Web

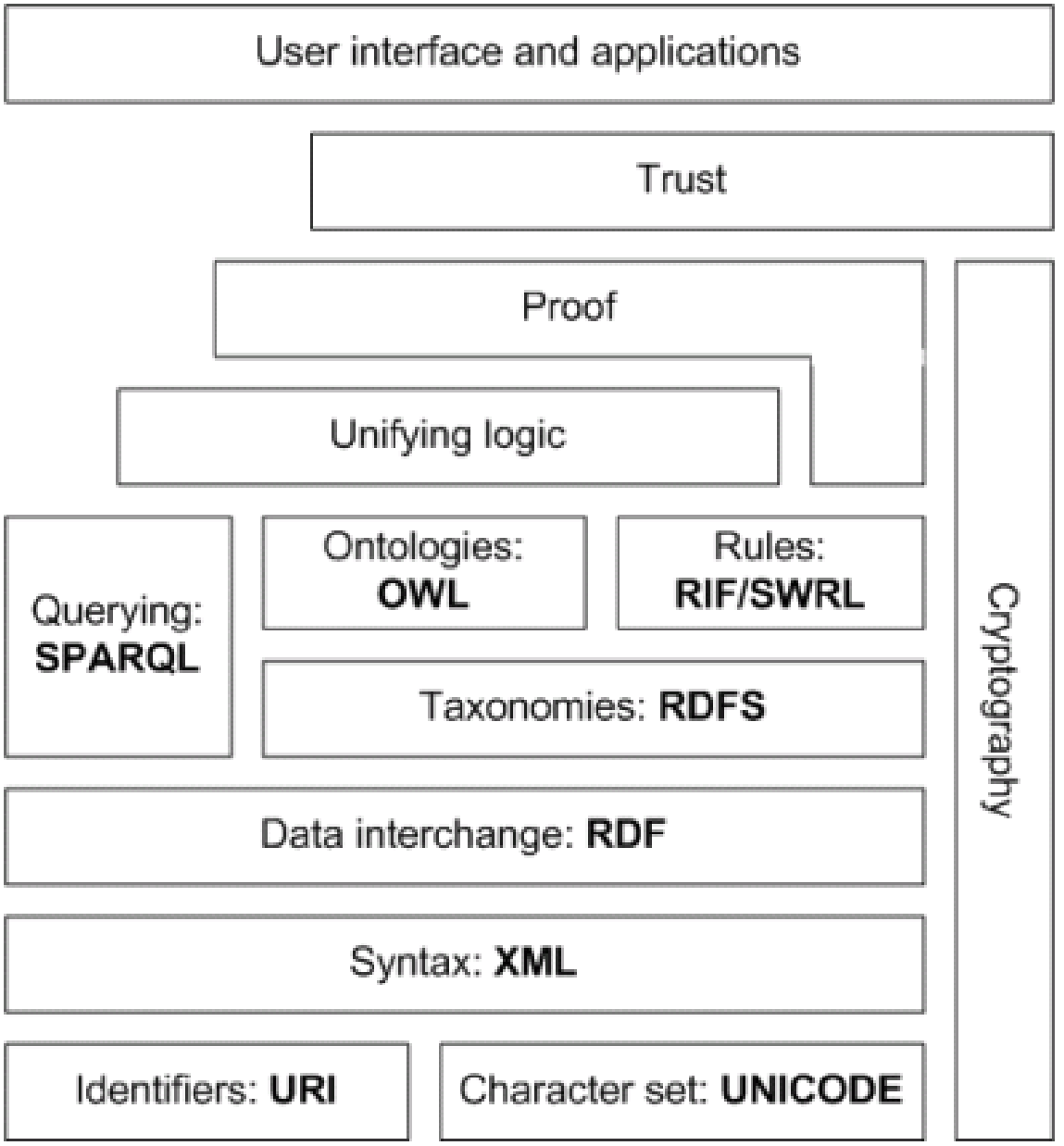
"This machine is a server – DO NOT POWER IT DOWN!!"

http://commons.wikimedia.org/wiki/File:First_Web_Server.jpg

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

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

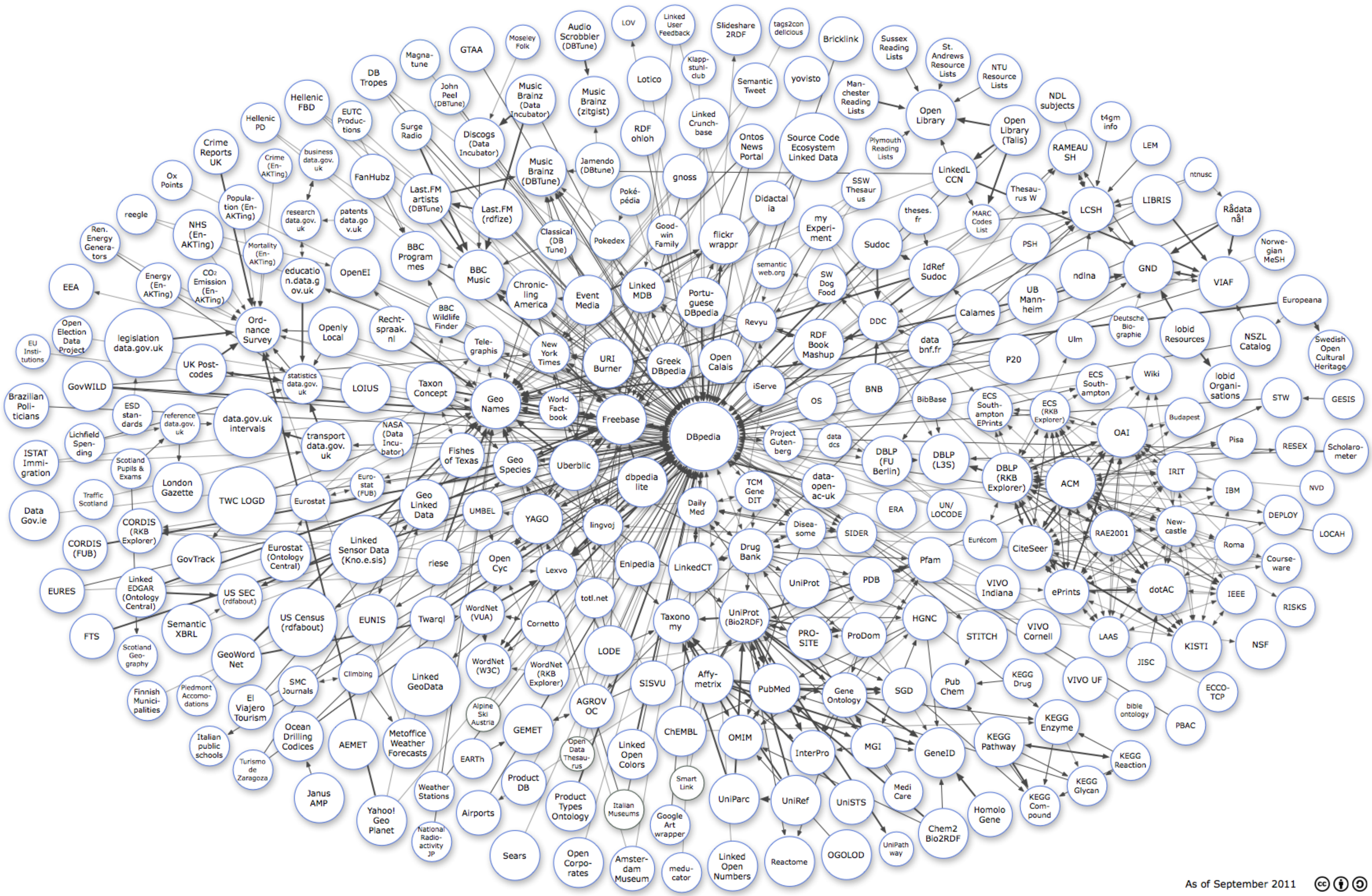
- Semantic Web activities have been baking since the late 1990s
- The infamous layer cake has evolved
- Linked Data is a more recent movement...



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/

CLAROS

The world of art on the semantic web

Home About Collections

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.

EXPLORE →

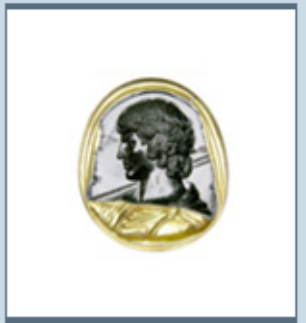
IMAGE SEARCHING →

PARTICIPATE →

OPEN DATA →

western ceramics

western sculpture

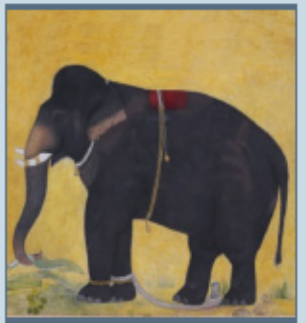


gems and cameos

prints and drawings

eastern bronzes

eastern ceramics



eastern painting

antiquarian photographs

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
  - and scholars!

Technologies: the (more) important bits

Web Architecture (HTTP, URIs, ...)

- a mechanism to unambiguously identify resources
- a standardised means to retrieve content

Resource Description Framework (RDF)

- a simple common information substrate
- incorporating Web linking

Ontologies (OWL, ...)

- encapsulating shared “meaning” using RDF
- an open world
- the ability to simultaneously support multiple ontologies

Semantics in Music and Audio

Many different forms and sources of semantics

- From composition (composer, publishing)
- From the music (score, artistic context)
- From audio production and performance
- From a (digital) artefact
- From an analysis
- From the analysis process and tools

General observations:

- Each a specialism
- Each might be embodied in an ontology or ontologies
- There are links and intersections between them

Each provides context or the others

How can we *usefully* combine these different semantics? How do we enable *end-to-end semantics?*

Semantics in Music and Audio

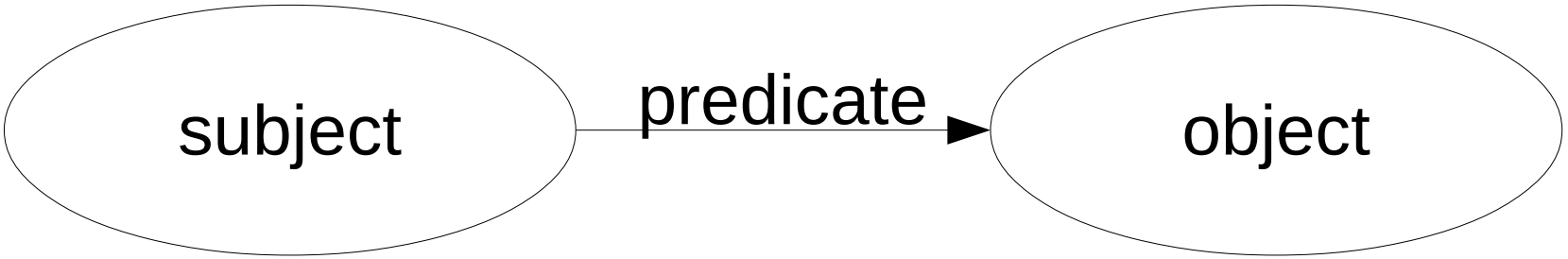
Many different forms and sources of semantics

- From composition (composer, publishing)
- From the music (score, artistic context)
- From audio production and performance
- From a (digital) artefact
- From an analysis
- From the analysis process and tools

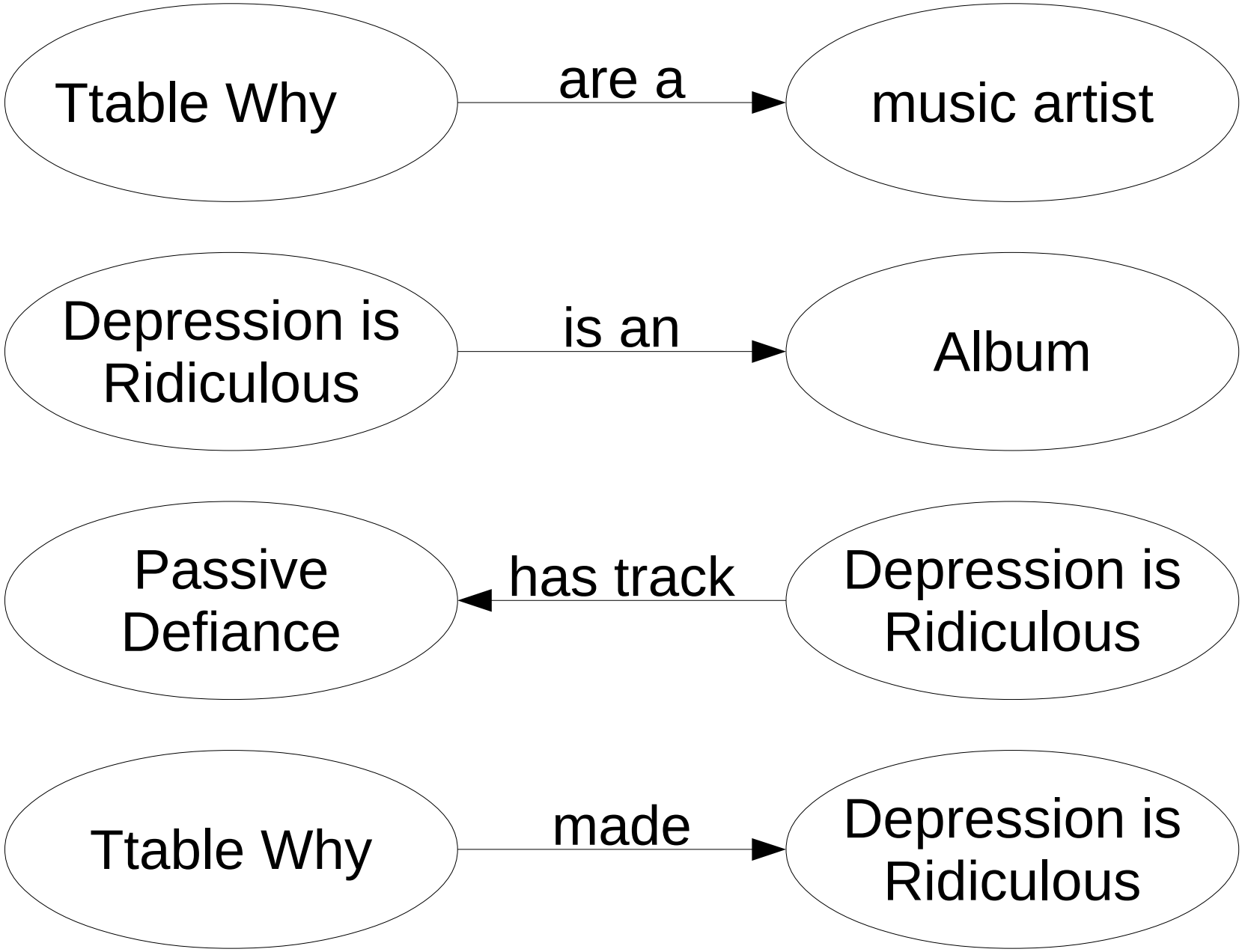
...but it's still important to enable the flow of semantics into the research context

The focus may be on the application of new MIR derived methods...

RDF is a simple model

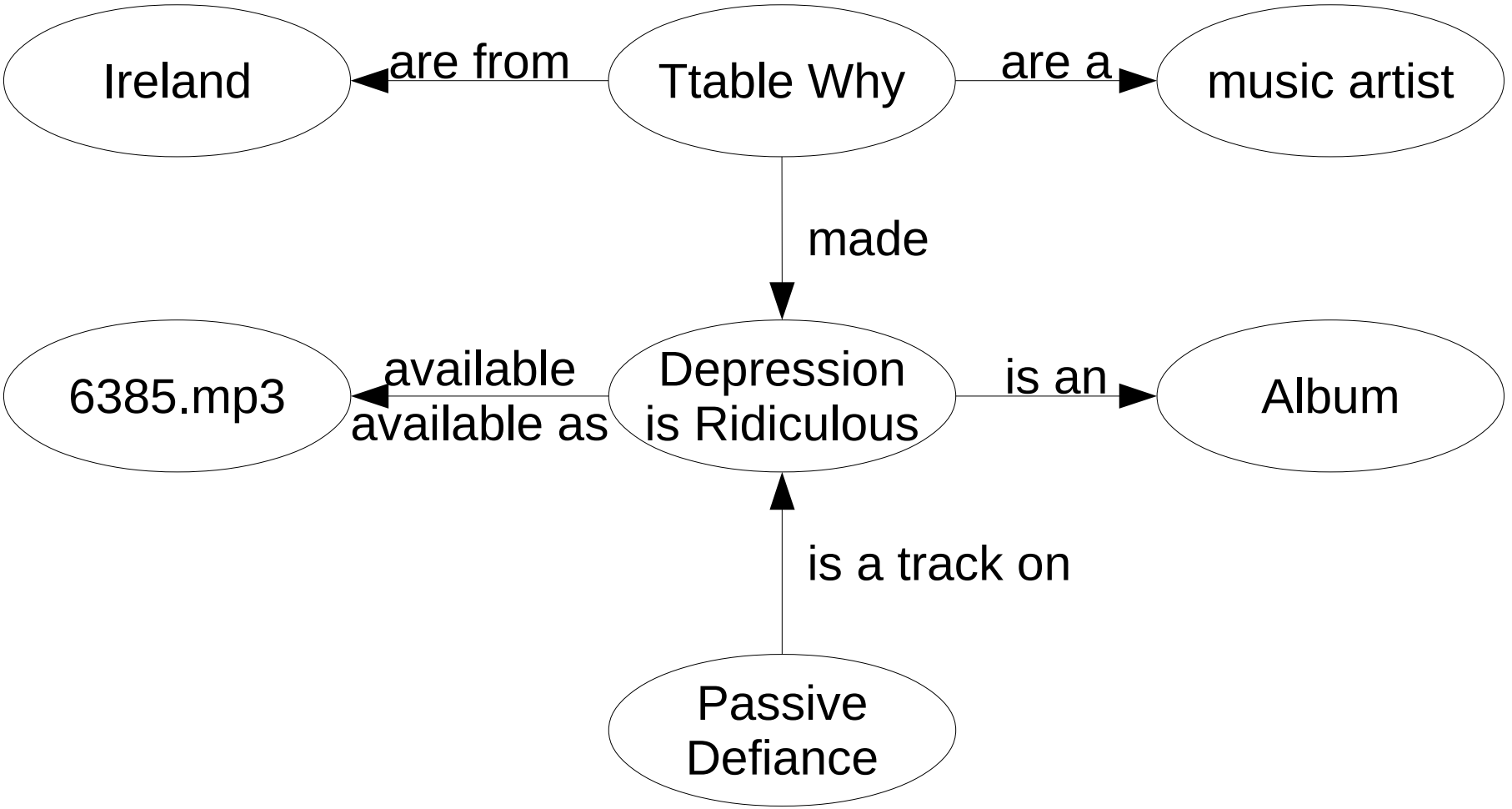


That's it.



Warning: approximated for clarity!

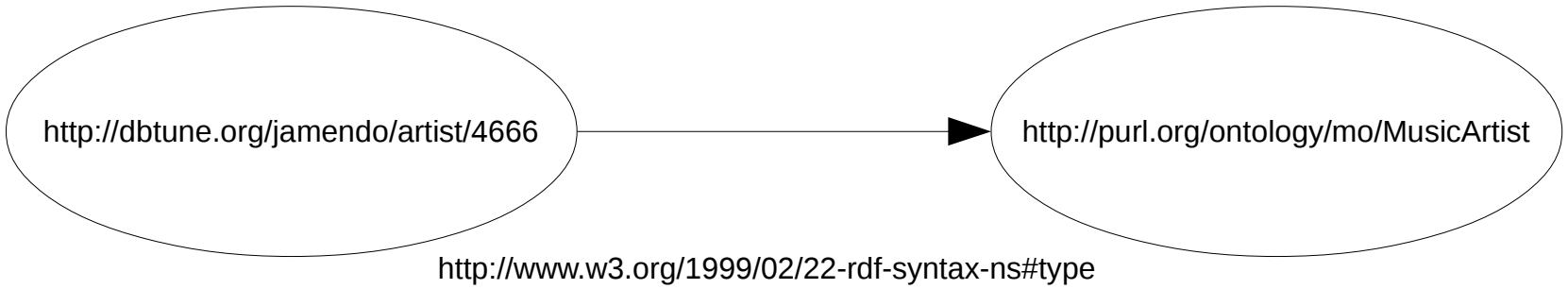
RDF “Triples” joined together form a graph



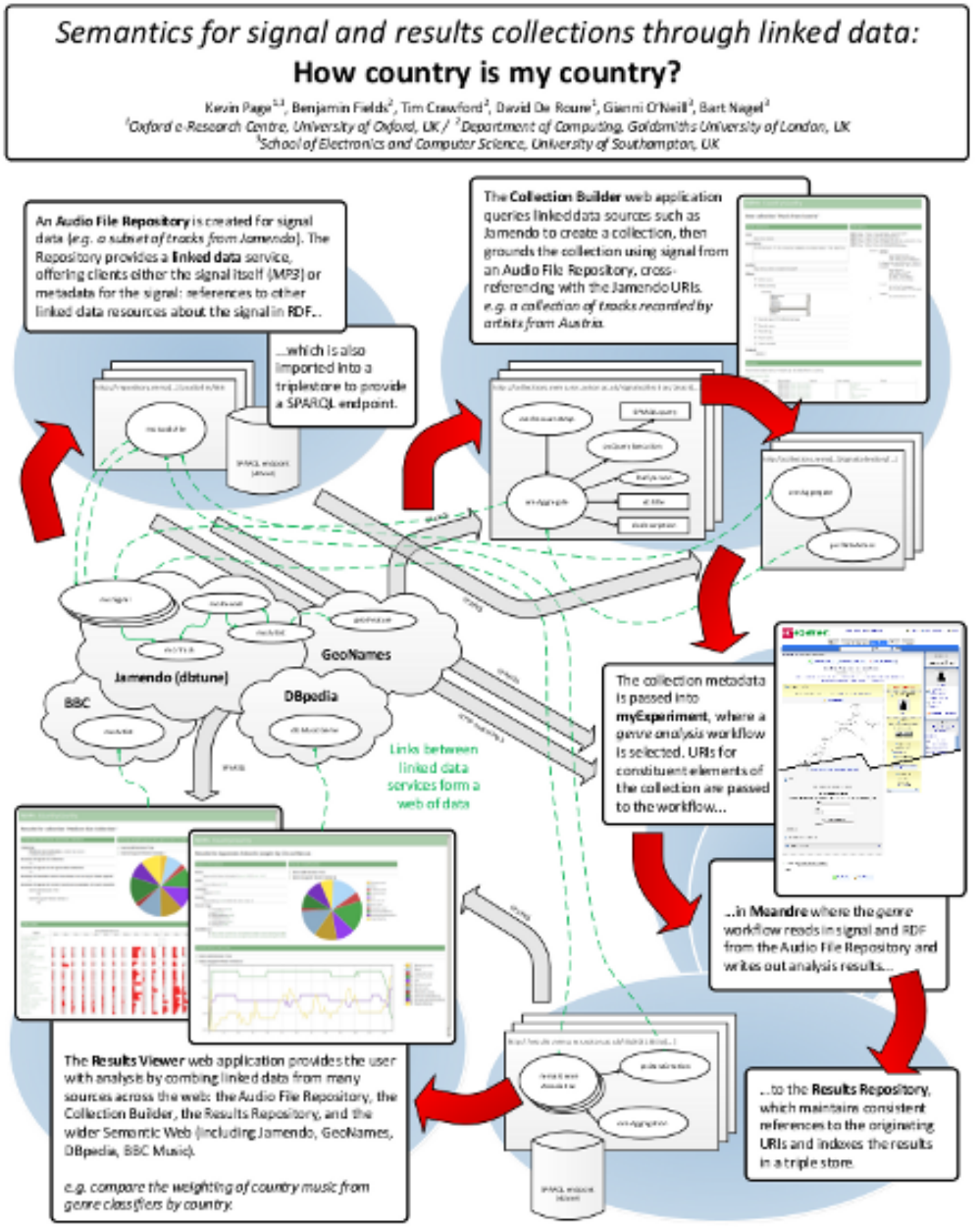
Warning: approximated for clarity!

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*

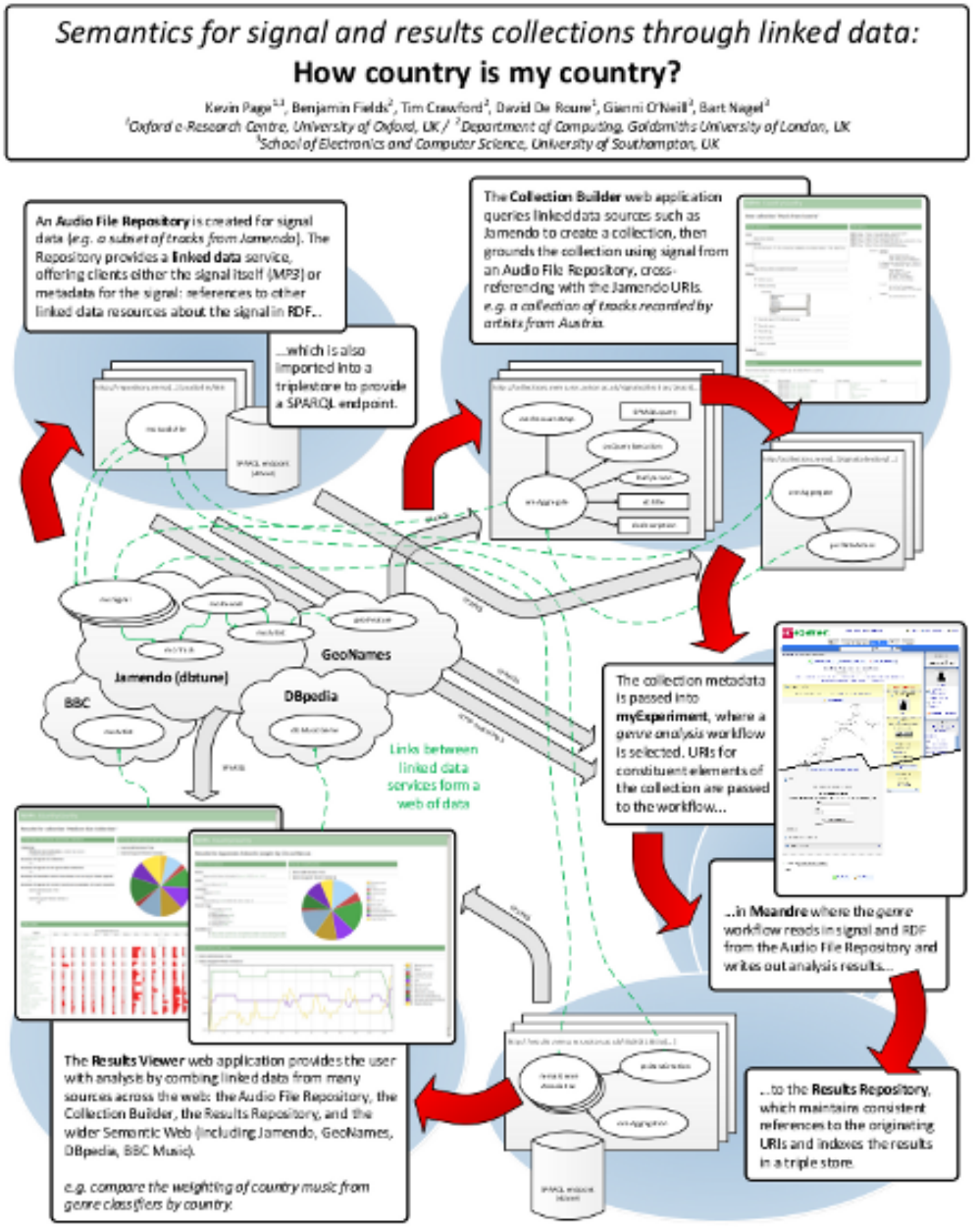


A simple example



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

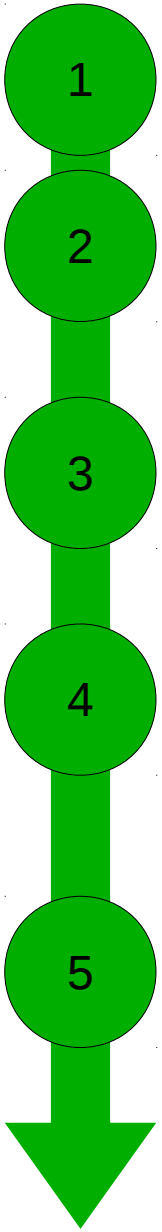
A “simple” example



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

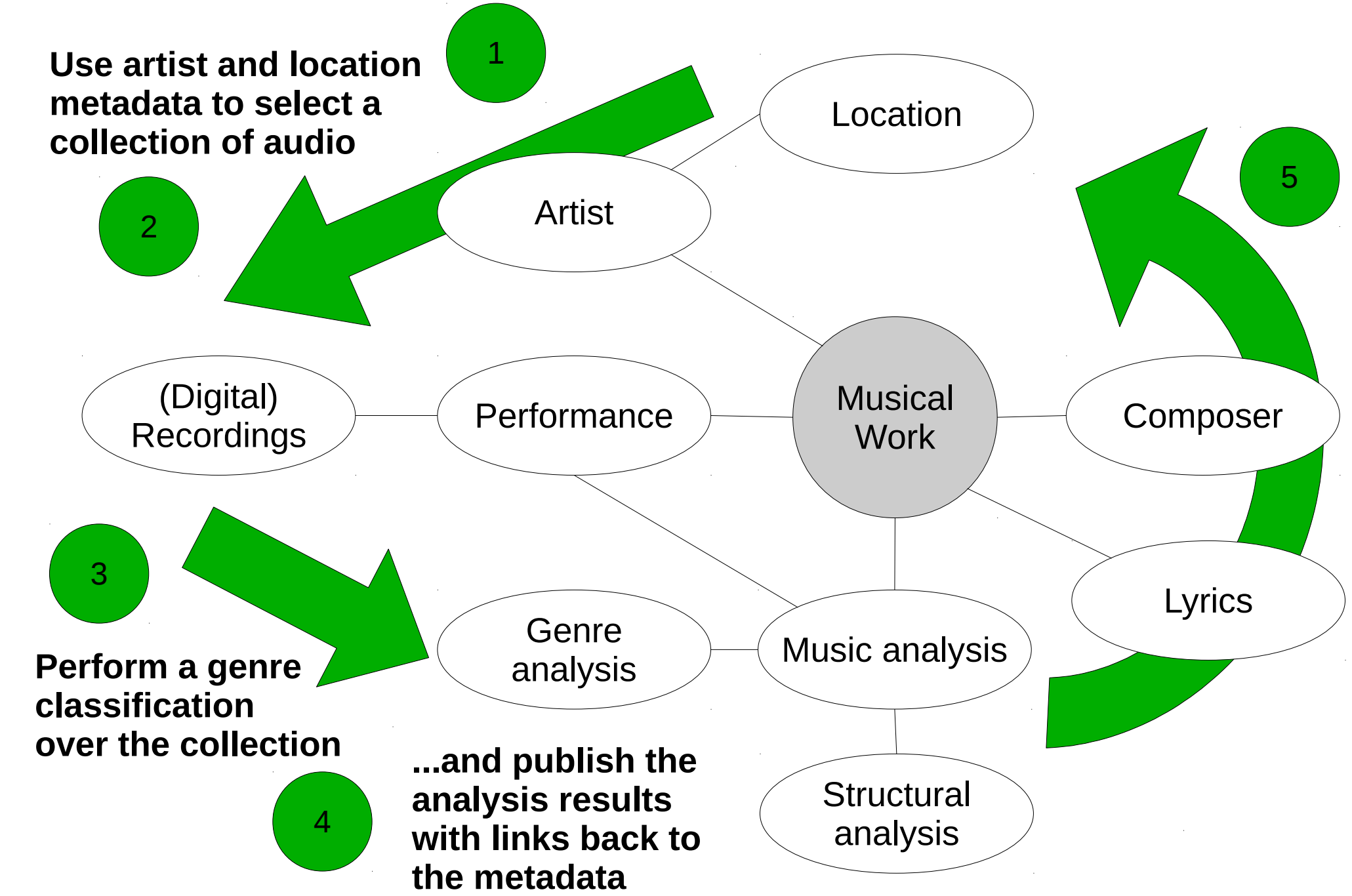
oerc Oxford e-Research Centre · Goldsmiths University of London · University of Southampton Electronics and Computer Science · The Andrew W. Mellon Foundation · JISC

The “How Country is my country?” 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)

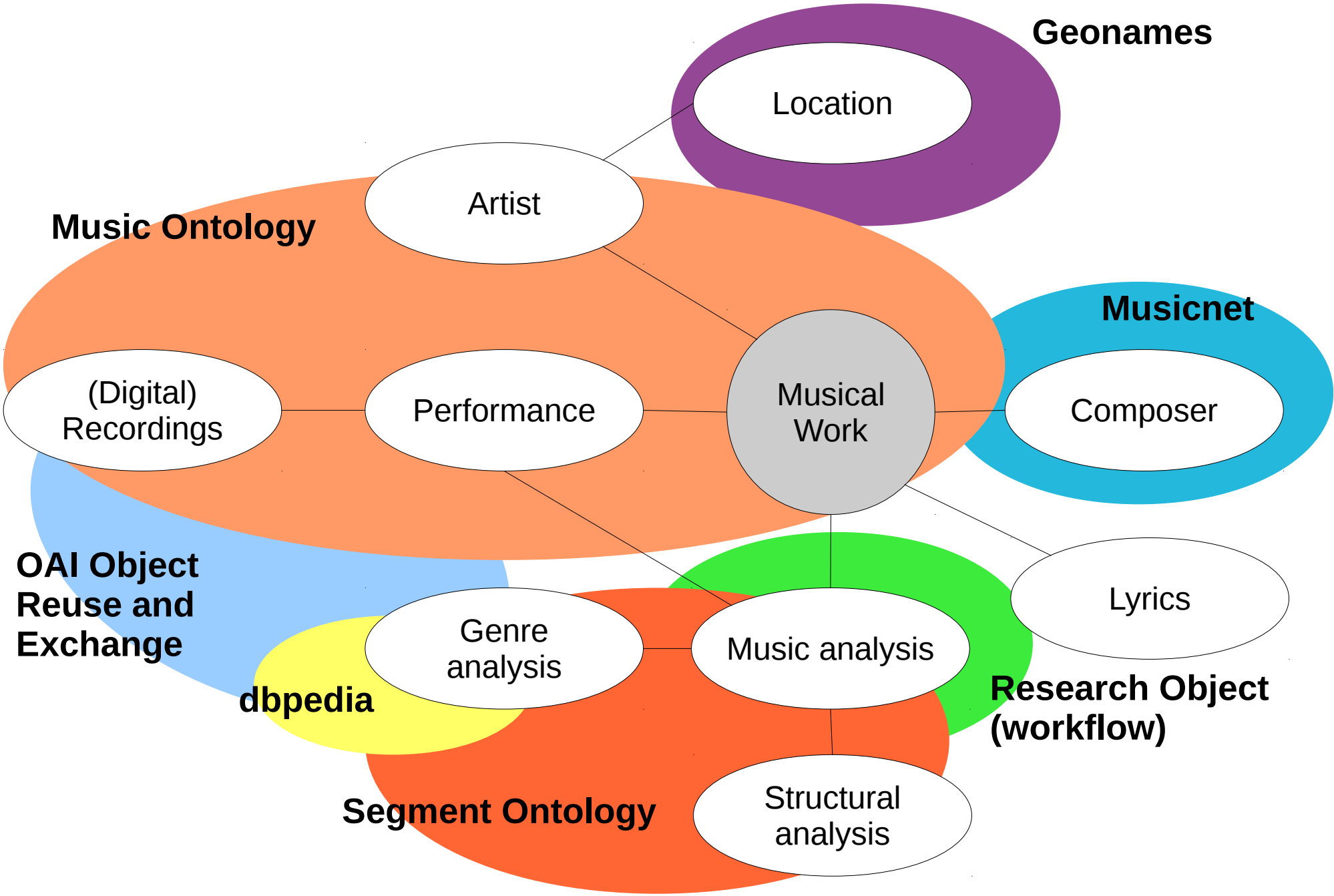
A simple example (simplified!)



Ontologies

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

Ontologies in our simple example



Ontology examples

```
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.

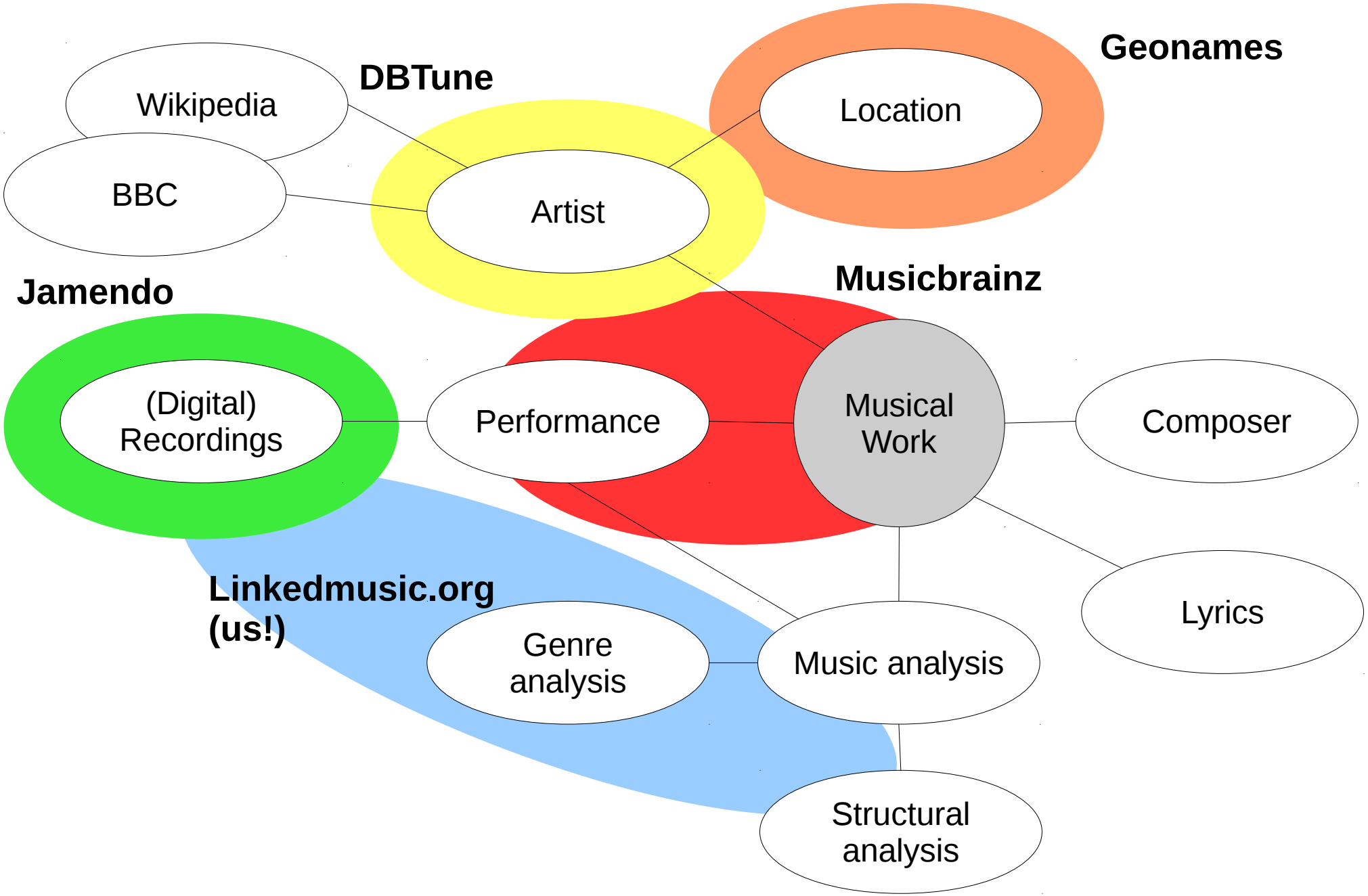
- Not just *on* the Web...
- But to use and build upon Web Architecture to scale humanities systems

System elements

- Audio File Repositories (signal)
- Music Collections
- Algorithms and workflow
- Algorithmic output
- Results and findings

... all joined through a web of linked data

Data & service distribution



Example 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> .
```

Summary

- We can use Web Architecture to publish the output of Transforming Musicology studies
- We can use RDF and ontologies to capture, scale, and link domain knowledge
- This enables us to scale and link different types of work
  - to combine and reuse research input, methods, context, and output

Mini-projects

- Are there Linked Data sources critical to your project?
- Are there Linked Data sources that might bring a richer context to your project?
- How can your output be linked to other Transforming Musicology data and methods for reuse
  - Where is the intersection between common concepts?
  - Where is your specialism? Can you transmit the semantics?

Transforming Musicology Semantic Infrastructure

- Triplestore