# Semantics for music analysis through linked data: How country is my country?

Kevin R. Page [1,2], Benjamin Fields [3], Bart J. Nagel [1], Gianni O'Neill [1], David C. De Roure [2], Tim Crawford [3]

[1] School of Electronics and Computer Science, University of Southampton, UK

[2] Oxford e-Research Centre, University of Oxford, UK

[3] Department of Computing, Goldsmiths University of London, UK

***Abstract*—We present a proof-of-concept system that demonstrates the utility of linked data for enhancing the application of Music Information Retrieval (MIR) workflows, both when curating collections of music signal data for analysis, and publishing results that can be simply and readily correlated to these, and other, collection sets and Linked Data sources.**

**The system includes: linked data implementations of a signal repository, collection builder, and results explorer; an extension to the myExperiment workflow sharing environment to include Meandre workflows; and support within myExperiment and Meandre to retrieve and persist resources from the linked data repositories.**

**By way of example we gather and publish RDF describing signal collections derived from the country of an artist. Genre analysis over these collections and integration of collection and result metadata enables us to ask: "how country is my country?".**

***Index Terms*—music analysis, MIR, workflow, semantic web, linked data.**

## I. Background

Researchers in the field of Music Information Retrieval (MIR) are confronted with problems beyond the design and implementation of systems and algorithms for retrieving information from music. The music recordings over which analysis would be expected to occur are often restricted from exchange amongst researchers, either explicitly through copyrights or implicitly through the high overheads of managing detailed and intricate licensing. As increasingly vast quantities of audio data are digitised, their entanglement with rights management will only make the curation and distribution of ever larger data sets a more complicated and time-consuming task. Even when audio data is freely available, a difficult balance must be found between the need for comparative evaluation of approaches using widely shared, understood, and re-usable data sets, and the avoidance of over-fitting an algorithms during development when a specific data set is repeatedly used for testing.

Evaluation also requires a common structure into which analytic output can be placed for comparison, rather than data structures inherited from the development tool or environment a researcher happened to be using. As faster computational resources become more readily available and can be applied to MIR tasks, the opportunities to undertake analysis on an ever greater scale[1], [2] brings with it the associated problem of managing ever greater quantities of result data.

It is clear, therefore, that systems which could reduce the complexity and workload of managing data and result collections while maintaining the ability to undertake comparative evaluations would aid the MIR researcher. MIREX[3] provides an institutional solution to this problem through the submission of algorithms to be comparatively evaluated using common metrics, when performing common tasks, upon using common datasets; but as an annual event it does not provide a solution for day-to-day research.

Some MIR systems have begun to incorporate data management and interoperability techniques: the Networked Environment for Music Analysis (NEMA) system[4] (used to operate MIREX 2010) adopts a Service Oriented approach of subsuming existing MIR tools as services, but is limited to those which can be aligned with its Java data structures; the jMIR suite uses the ACE XML DTD[5], adoption of which is therefore a prerequisite for interoperability; GNAT and GNARQL[6] use the Music Ontology (a key ontology utilised in this paper) to annotate only personal collections of music; while Henry[6], the Sonic Visualiser and Annotator tools[7] and their VAMP plugins also use the Music, Feature, and associated ontologies for import and export of data using the Resource Description Framework (RDF) model, however these systems could be characterised as traditional MIR solutions that employ semantic web technologies – rather than a Resource Oriented Architecture to support MIR research.

## II. General Approach and System Architecture

The Linked Data movement encourages a Semantic Web built upon HTTP URIs that are published, linked, and retrieved using RDF and SPARQL. Employing existing ontologies including the Music Ontology, GeoNames, and OAI-ORE, we present a proof-of-concept system and use case – informally known as Country/Country – that applies linked data across the lifecycle of MIR development in an effort to enhance the process of research.

The primary motivation of our approach is to utilise metadata associated with digital artefacts used in MIR research to simplify: the process of gathering collections, the application of collections to analysis workflows, publishing results, and undertaking analysis; so the researcher can focus on the development and application of algorithms.

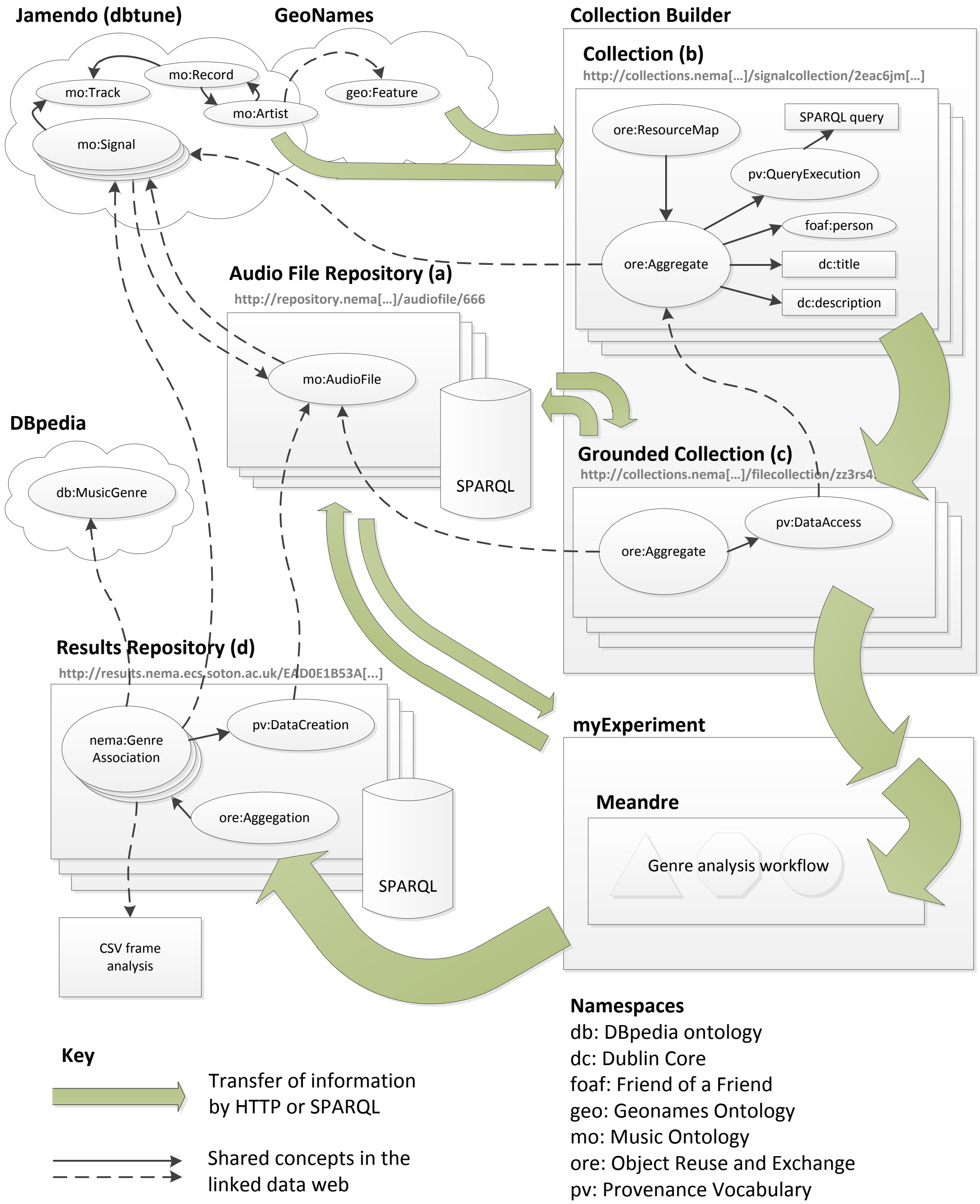


Figure 1. Country/Country system architecture with some connections in the linked data web overlaid

The prototype system consists of several services, described in detail in the following sections and shown in figure 1:

1) An *Audio File Repository* (section III) which serves signal files using the standard HTTP request methods and access mechanisms. The Audio File Repository also publishes a small RDF sub-graph describing each locally stored audio file as linked data.
2) A *Collection Builder* (section IV) that enables a researcher to select a set of signals described by linked data services, then publish the collection as RDF.
3) Music analysis is performed by a Meandre genre classification workflow:
   a) The *myExperiment* (section VI) collaborative environment has been extended to support Meandre workflows.
   b) myExperiment has been modified to accept RDF published by the Collection Builder.
   c) *Meandre workflow components* (section V) dereference resources from the Audio File Repository and persist URIs through the analysis workflow.

d) Output from the analysis is published as linked data in a *Results Repository*.

4) A *Results Viewer* (section VII) retrieves RDF published by the Collection Builder and Result Repository and contextualises one through the other using the common identifiers (URIs) and concepts (ontologies) applied throughout the system.

The services form a highly decentralised and distributed, loosely-federated, and scalable Resource Oriented Architecture[8]; that is, interactions between services occur over HTTP and involve the exchange of representations of resources identified globally by URIs. While the sequence above is repeated through the paper to explain the utility of the services in the context of the use case, there is no requirement for services to interact in this, or any other, specific order.

Since this is a proof of concept, each service is neither the singular nor definitive implementation of its type. For speed of development and clarity of explanation the instances of each service presented here are limited examples – in a true web of data we would expect many providers of all service classes.

## III. Audio File Metadata and Repository

The starting point for most music analysis tasks is the selection of input data for the algorithm under development to process. In our prototype system we focus on the provision of audio signal data in the form of MP3 files, but the technique could equally be applied to symbolic source material such as MIDI.

There are many bases upon which a researcher might assemble and manage a selection of input signal data; often this may be down to the practicalities of local availability of physical media, or freely accessible remote collections. Such limited collections cause not only validity issues such as overfitting algorithms to test data, but also preclude the discovery of novel research techniques and results that might be expected when analysing the massive and increasing digital corpus[9]. In this demonstrator we show how metadata can be used to automate the assembly and management of larger, distributed, and more dynamic collections. While limited metadata is often available through mechanisms such as ID3, this is usually no more than a simple string tag and is limited in scope to the specific audio file in question – here we apply Semantic Web techniques to retrieve and combine metadata both directly and indirectly related to the signal data.

A powerful and flexible feature of the Semantic Web is the ability to distribute metadata across the Web while maintaining a common foundation in the underlying (RDF) model and, as is often desirable, shared ontologies. For example metadata about an artefact such as an audio file can be maintained on a different web server than both the audio file itself and other distinct sources of metadata, but when required the metadata can be dynamically combined to form a coherent statement of information about, and with, the audio file. Building and maintaining this web of distributed information is a key motivator for the Linked Data community.

In our system, the first links to this web are made by an *Audio File Repository*, which serves both MP3 audio files and linked data about the audio files. While this represents a generic collection of audio files to which there may be open or restricted access, to enable a public demonstrator we have amassed a subset of the freely available Jamendo collection[1]. The repository consists of an Apache web server that has been configured to conform to REST[10] and Linked Data principles[11] such that:

- the primary (non-information) resources are *AudioFiles*[2] as described by the Music Ontology[12]. URIs are minted for these resources within the namespace of the repository, e.g. `http://repository.nema.ecs.soton.ac.uk/audiofile/100002`
- if a client fetches the URI representing an AudioFile non-information resource, and uses the HTTP Accept header to request `audio/*` (e.g. `audio/mpeg`), then the server issues a 303 redirect to the audio signal file (an information resource) which the client can then download.
- if, through the HTTP Accept header, a client requests `application/rdf+xml`, then the server issues a 303 redirect to a linked data RDF file (another information resource) containing metadata pertaining to the AudioFile.
- the RDF sub-graphs for each AudioFile are written to a 4store[3] triplestore which provides a SPARQL query endpoint.

A second motivation for populating our audio repository with music from the Jamendo label is to utilise the Linked Data endpoint for Jamendo available at dbtune[4]. This in turn enables us to publish Linked Data ourselves (figure 1(a)), served when a client requests RDF (as above), and using the Music Ontology to assert that:

- an audio file resource found in our repository is an instance of an AudioFile.
- each AudioFile in our repository encodes a specific linked *Signal* instance as defined in the Jamendo linked data set (where Signal is the concept as defined by the Music Ontology).
- a specific Track instance in the Jamendo RDF graph is encoded by each AudioFile in our repository.

While the open licensing of Jamendo enabled us to build a public demonstrator, there is no fundamental requirement for the audio files to be sourced from the same provider as the linked data – as shown in later sections, our aim is to encourage the opposite. For example, the audio files could be transcodings from a private collection with access restricted on an institutional basis, while album metadata would be linked from the Musicbrainz endpoint[5].

[1] http://www.jamendo.com/

[2] Capitalised terms throughout this paper refer to concepts defined in ontologies, e.g. the Music Ontology.

[3] http://4store.org/

[4] http://dbtune.org/jamendo/

[5] http://dbtune.org/musicbrainz/

## IV. Collection Builder Web Application

### A. *Creating collections*

While provision of a linked data Audio File Repository was a necessary building block in construction of the Country/-Country prototype, a key motivation for our approach is to free MIR researchers from data sets that are directly derived from specific signal repository contents. Dynamic collections spanning multiple repositories could instead be selected using criteria relevant to the research being undertaken, whether from within or outside the MIR domain; earlier experimental results could be fed back into this process as further criteria for creation of derivative collections.

For purposes of our demonstrator, we envision a simplified use case where a researcher wishes to investigate the possible correlation between the genre of a performance (e.g. country, jazz) as detected by an MIR algorithm, and the domicile of the performing artist.

The *Collection Builder* web application (figure 2) provides the user with an interface to create collections from the entire Jamendo community, rather than being limited to the sub-set of signal served by our Audio File Repository. As the user selects filters, beneath the UI SPARQL queries are built up, using concepts within and beyond Jamendo and the music ontology to query for *Signal* instances.

For example, the SPARQL query:

```
PREFIX geo: <http://www.geonames.org/ontology#>
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 * WHERE {
   ?artist a mo:MusicArtist ;
   foaf:name ?artistname ;
   foaf:based_near ?basednear .
   { ?basednear geo:inCountry <http://www.geonames.org/
       countries/#BE> }
   OPTIONAL { ?basednear geo:inCountry ?country . }
   ?record
      a mo:Record ;
      foaf:maker ?artist ;
      mo:track ?track ;
      dc:date ?recorddate ;
      dc:title ?recordname .
   ?track
      dc:title ?trackname ;
      mo:track_number ?tracknumber .
   ?signal
      mo:published_as ?track .
}
```

is used return to details of tracks recorded by artists from the country of Belgium, where the location of an artist is asserted in the Jamendo data, but the country of that location is encoded by GeoNames[6].

Once the user has applied sufficient filters to achieve their desired criteria and a SPARQL query constructed to enact it, the user may "publish" their collection. This takes the form of RDF (figure 1(b)), whereby the Collection Builder:

- mints a URI for the RDF collection in the `http://collections.nema.ecs.soton.ac.uk/` namespace.

[6]http://www.geonames.org/ontology/

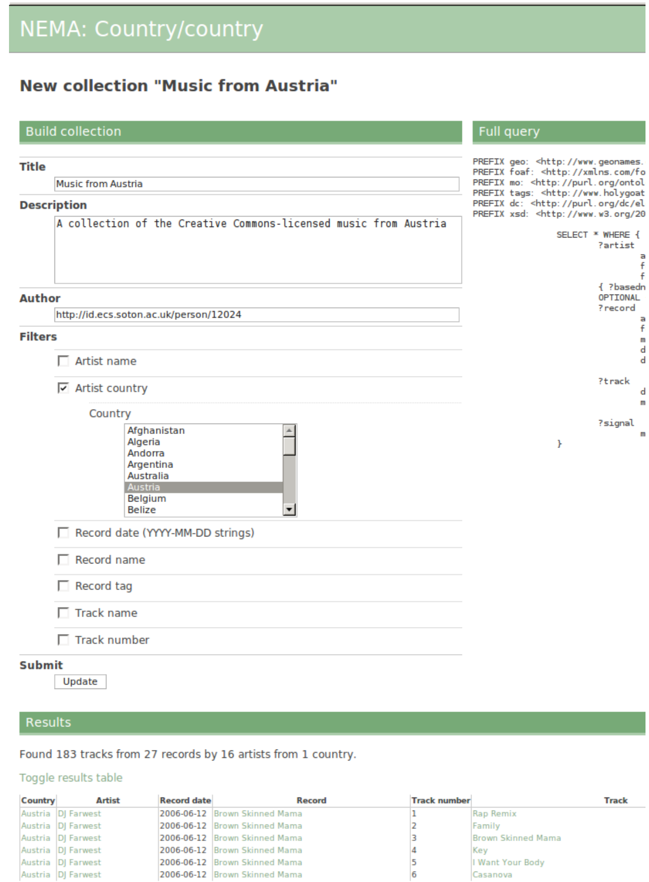


Figure 2. The Country/Country Collection Builder web application

- asserts the collection as an ORE[13] *Aggregate* of Signal, where the Signals are URIs from the Jamendo namespace that match the SPARQL query.
- uses the Provenance Vocabulary[14] to record the SPARQL query used.
- asserts user specified additional metadata including authorship and description.

### B. *Grounding collections with audio files*

The collections described in the previous section have been selected by criteria unbound by an audio file repository, but they only contain the abstract notion of Signal as defined by the Music Ontology; to be used as input by an MIR algorithm they must be "*grounded*" as AudioFiles that encode corresponding Signal in one, or several, signal repositories.

A second stage of the Collection Builder web application enables a user to do just this. By querying the SPARQL endpoint provided by the Audio File Repository (section III) for the Signal URIs aggregated in the abstract collection, a second RDF aggregation (figure 1(c)) is published:

- a URI is minted for this grounded collection
- an ORE aggregate is asserted, this time containing AudioFile URIs from the Audio File Repository (section III) that encode Signal from the existing abstract collection.

- the Provenance Vocabulary expresses the relationship between the grounded collection and its abstract precursor.

It should be noted that an AudioFile collection may not be a complete grounding of a Signal collection: coverage is restricted to that of the Audio File Repository (or repositories) available. On the other hand, multiple corresponding AudioFiles may be available and encoded in the grounded collection, whereby the appropriate repository would be selected when the AudioFile required is determined by network speed, locality, or license restricted access.

## V. Meandre Workflow and Results Repository

Meandre is a data-intensive flow framework[15] which has been adopted as the workflow enactment engine at the core of the NEMA system. The heart of the NEMA system design is an extensible Java data model that incorporates MIR data structures from existing tools such as jMIR[16] and the Sonic tools[7]; in combination with the distributed execution environment of Meandre this allows the NEMA system to host and run MIR workflows authored in a wide variety of existing tools[4]. Most recently the NEMA system has been used as the submission and evaluation framework for MIREX 2010[7].

For the MIR stage of the Country/Country prototype we have adopted an existing Meandre workflow that performs genre and mood analysis. i.e. it takes an audio signal as input, and through a workflow of feature extraction and a number of trained classifiers (e.g. CART decision tree, J48 decision tree, Linear Discriminant) provides a weighted ranking of genre (e.g. country, baroque, jazz, rock) and mood (e.g. aggressive, wistful, cheerful) for each audio signal.

### A. Meandre Components

Each component in a Meandre flow is encapsulated by a Java object, and to integrate with the linked data services provided by Country/Country we have modified the "head" and "tail" components of the flow such that:

- the head component, which retrieves and passes an audio signal to the feature extractor, has been adapted to parse a linked data AudioFile URI – such as one provided by our Audio File Repository (section III) – as its input. The component dereferences this non-information resource twice: once with the `audio/mpeg` HTTP Accept header to get the audio signal file, and again requesting `application/rdf+xml` to retrieve the linked data pertaining to the AudioFile.
- the RDF sub-graph retrieved from the Audio File Repository is stored using an in-memory Jena[8] model so that the URIs can persist through the flow. This maintains the crucial links between the audio signal retrieved from the repository and processed by the flow, and the global identifiers – the URIs – of the Signal of which the AudioFile is an artefact, and – via Signal related concepts such as Artist and Track – linked data sources such as Jamendo.
- the tail component outputs the weighted rankings by genre and mood from the classifiers: the results of the analysis. Because the RDF sub-graph includes concepts from the Music Ontology for both global identifiers (e.g. for Signal) and local artefacts (AudioFile) we can distinguish between these when recording results. *Genre*, for example, is a concept applied to a *Signal*, for which the AudioFile is a digital artefact of a Signal (that in turn encodes a *Performance*).
- analysis is performed on a frame-by-frame basis within the workflow, so output is written both as a CSV file containing detailed classifier values for each frame, and as a linked data RDF model with the average analysis for the whole Performance (i.e. per AudioFile).
- the RDF result graphs are also inserted into a 4store triplestore to provide a SPARQL query endpoint.

### B. Results Repository RDF

The tail component of the workflow uploads output from the analysis to a *Results Repository* (figure 1(d)). The fundamental resources in the repository are ORE Aggregations containing *Associations* (as defined in the Music Similarity Ontology[17]), where the Aggregation of Associations corresponds to the results from a single classifier analysing a single AudioFile. URIs are minted for these associations in the `http://results.nema.ecs.soton.ac.uk/` namespace.

For example, output from the genre classifiers is modelled using a locally declared *GenreAssociation* subclass of Association, which has as its subject a Signal instance (derived from the AudioFile via the Audio File Repository linked data), and as its object a *MusicGenre* instance as defined by the DBpedia[18] ontology.

Further Provenance Vocabulary is used to record the Meandre flow execution instance that performed the analysis (*createdBy*), the classifier within the flow (*usedGuideline*), and the AudioFile input to the analysis (*usedData*, as distinct from the parent Performance of which the AudioFile is a derivative artefact). The CSV file containing frame-by-frame analysis is linked using the Opaque Features File ontology[9].

## VI. myExperiment Workflow Management

The myExperiment[19] web-based virtual research environment provides discovery, sharing, and management of workflows and associated Research Objects throughout their lifecycle, providing specific support for Taverna workflows.

We have added support for Meandre workflows, as used by the NEMA system and the Country/Country prototype, to myExperiment[10]. This includes a preview page for Meandre flows, the same ability to share and manage Meandre flows as for Taverna, and functionality to enact the flow on a specified

[7]http://www.music-ir.org/mirex/wiki/2010:Main_Page

[8]http://jena.sourceforge.net/

[9]http://purl.org/ontology/off/

[10]Meandre support will be merged into the main myExperiment version over the coming months.

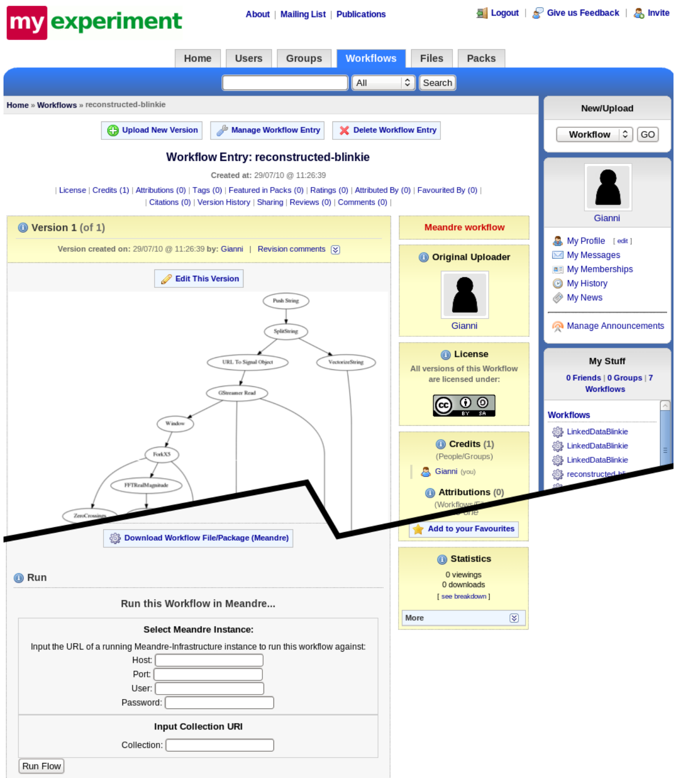


Figure 3. myExperiment previewing a Meandre flow (truncated)

Meandre flow server (figure 3). The underlying implementation stores MAU files (a complete self-contained Meandre workflow including executable components and workflow metadata) within the myExperiment system.

We also extend the myExperiment API to support importing collections from the Country/Country Collection Builder (section IV). This new API method takes the URI of a grounded collection as its argument; when accessed myExperiment loads the Collection metadata and makes it available to a user as potential input to a workflow. Should the user then apply the collection to the Country/Country genre analysis workflow, myExperiment will iterate through the collection and enact the workflow for each AudioFile URI within (each AudioFile URI is then dereferenced within the workflow; see section V).

A link utilising this API call is appended to the end of the Collection Builder grounding process so that a user can quickly and simply move from collection maintenance to application of the collection to workflows in myExperiment.

## VII. Results Viewer Web Application

The final service provided as part of the Country/Country prototype system is a web application which allows a researcher to view the analysis results, cross-reference against collections, and combine the analysis with other linked data sources. More than any other component the Results Viewer is a proof-of-concept that highlights only a select number of the many possible data sources and combinations.

The Results Viewer demonstration implementation (figure 4) begins by combining two linked data sets: it takes a collection (as created in the Collection Builder, section IV) and queries the Result Repository SPARQL endpoint (section V) matching Association results for Signal contained in the collection(s). The demonstrator is focussed on our country-centric genre analysis scenario: using country derived collections cross-referenced with result data a number of statistics and visualisations pertinent to this scenario are calculated and rendered (figure VII) including:

- for a collection (and comparison of multiple collections)
  - the number of signals
  - the number of signals which have been grounded in an Audio File Repository
  - number of classifiers which were run on any of the signals according to a Results Repository
  - the numbers of results (Genre Associations) available for each workflow enactment of the classifier
- for each classifier over a collection
  - the songs (by artist and title) which are most and least weighted for each genre
  - a pie chart taking the highest genre weighting for each signal
  - a pie chart showing average weightings for each genre over the collection
- for each signal in the collection (figure 4)
  - a full listing of genre weightings from each classifier
  - a playback page which retrieves the AudioFile (using linked data from the Audio File repository) and the frame analysis data (using the data link from the Results Repository RDF) and, while playing the audio, scrolls a graph showing the frame-by-frame genre weightings for each classifier
  - references to other relevant information from linked data sources

It is through this final element that we demonstrate the further potential of linked data in bringing together a wide variety of information sources.

In our first example we take the highest weighted genre for a given artist or collection and link to other artists in DBpedia who perform in the same genre and are also from the same country. This illustrates how it is possible to link between imperfectly aligned data sets: not only do GeoNames (and the linked Jamendo data set and the Country/Country collections) and DBpedia use different ontologies for countries, but artists in DBpedia can be associated with a wide variety of geographic coverages (town, region, country, etc.) and through various relationships (residence, place of birth, etc.).

We can overcome this because geographic entities below the level of country in DBpedia have been asserted “sameAs” specific features in GeoNames – in other words, even though it is conceptually incorrect to align the ontologies at the level of country, it is possible at other levels (e.g. cities and towns).

A SPARQL query to DBpedia for a list of geographic locations associated with all artists of a specific genre can then be cross-referenced against their sameAs features in

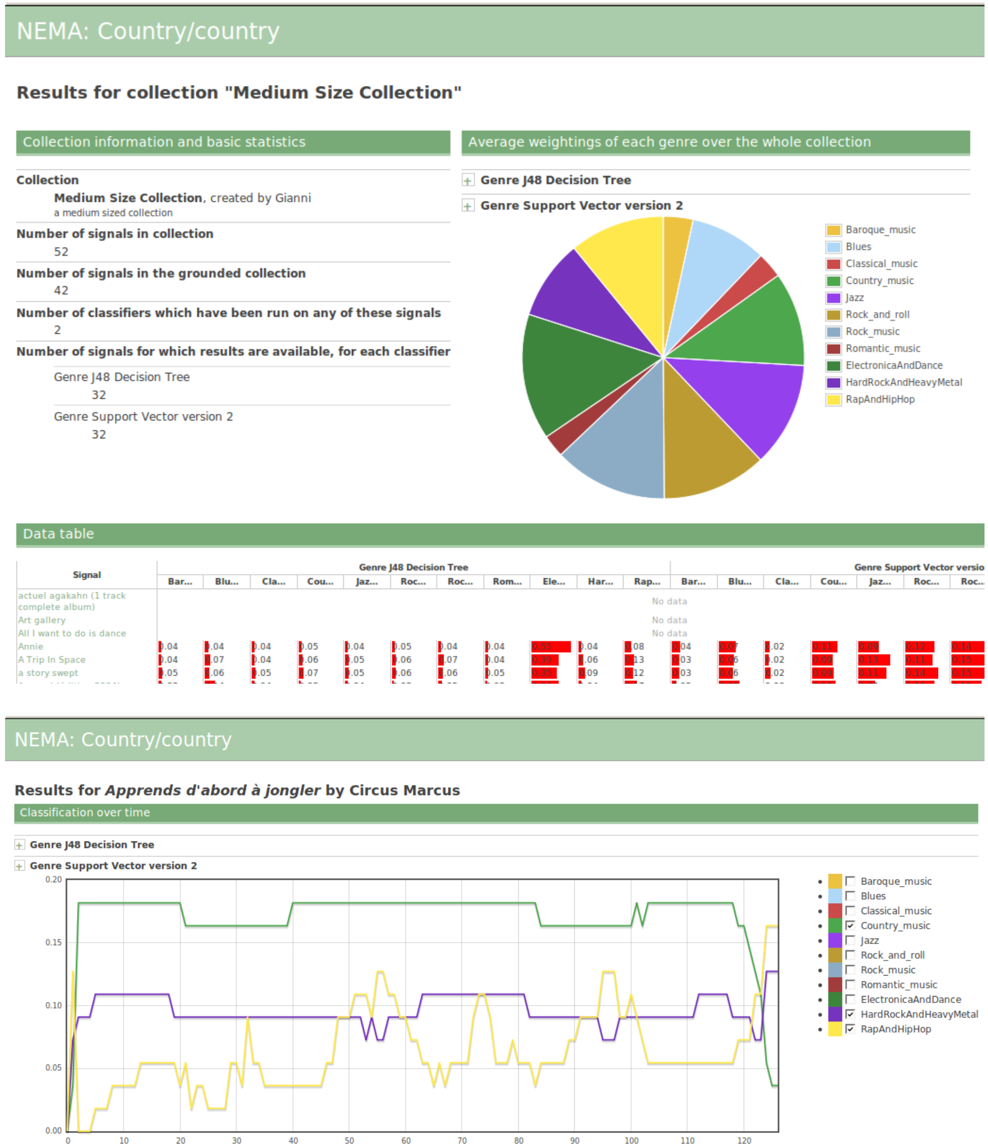


Figure 4. The Result Viewer web application showing analysis for a collection (top) and genre weightings over time for a specific Signal (bottom).

GeoNames, which can finally be culled by the GeoNames country they are located in (as used by the Country/Country collections). Although the relationship between artist and country in DBpedia can be one of several types, the common RDF model allows us to process them all.

In our second example we take this list of artists and, using the provided sameAs assertions, link to the same artists on the BBC Music website[11].

## VIII. Conclusions and Future Work

The Country/Country prototype demonstrates the utility of semantic web technologies: the consistent use of globally unique identifiers (in the form of URIs) that can persist within and between systems; a resource oriented architecture which enables highly distributed, lightweight, and dynamic services when publishing data; a common underlying model in RDF and shared ontologies for information exchange; and the power of merging distributed information through a web of linked data.

We have shown how even a relatively limited linking of semantic web data sources can provide an MIR researcher with a far greater flexibility when selecting input sources than previously available. When links to the RDF graph are maintained through an analysis workflow, we have demonstrated that the results when published as linked data can be quickly, easily, and usefully cross-referenced with other results, signal collections, and further sources of data beyond the obvious day to day purview of the researcher.

While the demonstrator embodies a specific analysis (genre) and collection selection (by nation) in a basic use case, the approach and technologies are more generic and widely

[11] http://www.bbc.co.uk/music/

applicable. The common data model of RDF extends a myriad of possibilities for linking with data models and categorisations within and without the MIR community, which future iterations of our architecture will address.

That is not to say one should overstate the availability of linked data that is currently available, for while there are plentiful opportunities for improving the lot of researchers using the current sparse link density, information exposed as linked data is but a tiny fraction of that available on the World Wide Web – the document web. Herein lies something of a bootstrapping problem, and one which we hope can be overcome by the easy, inconspicuous, and simple data publishing techniques we have illustrated (where we have exposed details such as SPARQL queries to the user, this is only in its role of proof of the concept!).

The tools we have created offer increased automation and simplification of day to day tasks, greater impact of results through easy access and in turn more frequent re-use and validation by peers. While a researcher may not immediately or directly recognise the benefits of idealised Linked Data Principles, such practical benefits must surely be an attractive motivation which could kick-start a virtuous circle of reuse and automation: one researcher's results can form the basis for another's input collection, so data and techniques can be combined, the web of linked data grows, and the scale of re-use and automation grows further.

Future work will more completely and accurately model the data and processes within analysis workflows – "black boxes" within the current system, but which are the focus of any MIR researcher's work and interest. While Meandre is nominally underpinned by an RDF data model, the structure of this model is as yet insufficient for direct publication as linked data (the extensive use of string literal key/value pairs limits the opportunities for linking). We also recognise that procedures such as collection building are themselves workflows, and as the quantity of linked data available for collection building increases the value of applying workflow techniques and sharing environments can only grow.

Finally, while we have demonstrated how RESTful and Linked Data techniques enable the distributed serving and separation of content and metadata, a future implementation should demonstrate how standard HTTP access and authentication mechanisms can take advantage of this separation for the purpose of adhering to digital rights restrictions on audio content. If applied to the system presented in this paper, the resulting Resource Oriented Architecture would make a compelling implementation of the "OMEN"[20] approach.

## ACKNOWLEDGEMENTS

Many thanks to Stephen Downie and his team at IMIRSEL University of Illinois for access and source code to their "Son of Blinkie" genre classification workflow, the myExperiment team at the University of Southampton for help and guidance while developing the Meandre extension, and to Michael Jewell for the thought that led to an inspiring title!

This work was carried out through the Networked Environment for Musical Analysis project funded by the Andrew W. Mellon Foundation, and the Structural Analysis of Large Amounts of Musical Information (SALAMI) project funded by the JISC Digitisation and e-Content programme as a part of the Digging into Data challenge.

## REFERENCES

[1] A. Berenzweig, B. Logan, D. Ellis, and B. Whitman, "A large-scale evaluation of acoustic and subjective music-similarity measures," *Computer Music Journal*, vol. 28, no. 2, pp. 63–76, 2004.

[2] E. Al-Shakarchi, P. Cozza, A. Harrison, C. Mastroianni, M. Shields, D. Talia, and I. Taylor, "Distributing workflows over a ubiquitous p2p network," *Scientific Programming*, vol. 15, no. 4, pp. 269–281, 2007.

[3] J. S. Downie, "The music information retrieval evaluation exchange (2005-2007): A window into music information retrieval research," *Acoustical Science and Technology*, vol. 29, no. 4, pp. 247–255, 2008.

[4] K. West, A. Kumar, A. Shirk, G. Zhu, J. Downie, A. Ehmann, and M. Bay, "The Networked Environment for Music Analysis (NEMA)," in *6th IEEE World Congress on Services*, Jul. 2010, pp. 314–317.

[5] C. McKay, J. A. Burgoyne, J. Thompson, and I. Fujinaga, "Using ACE XML 2.0 to store and share feature, instance and class data for musical classification," in *Proceedings of the International Society for Music Information Retrieval Conference 2009*, Oct. 2009, pp. 303–8.

[6] Y. Raimond and M. Sandler, "A Web of Musical Information," in *Proceedings of the International Conference on Music Information Retrieval 2008*, Sep. 2008, pp. 263–268.

[7] C. Cannam, C. Landone, M. Sandler, and J. Bello, "The sonic visualiser: A visualisation platform for semantic descriptors from musical signals," in *Proceedings of the 7th International Conference on Music Information Retrieval*, 2006, pp. 324–327.

[8] L. Richardson and S. Ruby, *RESTful Web Services*. O'Reilly & Associates, May 2007.

[9] D. De Roure, D., "SALAMI: Structural Analysis of Large Amounts of Music Information," in *UK e-Science All Hands Meeting 2010, Software Sustainbility Workshop*, Sep. 2010.

[10] R. T. Fielding, "Architectural Styles and the Design of Network-based Software Architectures," Ph.D. dissertation, Information and Computer Science, University of California, Irvine, California, USA, 2000.

[11] L. Sauermann and R. Cyganiak, "Cool URIs for the Semantic Web," W3C Semantic Web Education and Outreach Interest Group Note, 31 Mar. 2008.

[12] Y. Raimond, S. Abdallah, M. Sandler, and F. Giasson, "The music ontology," in *Proceedings of the International Conference on Music Information Retrieval*, 2007, pp. 417–422.

[13] C. Lagoze and H. Van de Sompel, "Object Reuse and Exchange (OAI-ORE)," Open Achives Initiative, Tech. Rep., 2008.

[14] O. Hartig and J. Zhao, "Publishing and Consuming Provenance Metadata on the Web of Linked Data," in *Proceedings of the 3rd International Provenance and Annotation Workshop (IPAW)*, Jun. 2010.

[15] X. Llorà, B. Ács, L. Auvil, B. Capitanu, M. Welge, and D. Goldberg, "Meandre: Semantic-Driven Data-Intensive Flows in the Clouds," in *4th IEEE International Conference on eScience*, Dec. 2008, pp. 238–245.

[16] C. McKay and I. Fujinaga, "jMIR: Tools for automatic music classification," in *Proceedings of the International Computer Music Conference 2009*, Oct. 2009, pp. 65–8.

[17] K. Kurt Jacobson, Y. Yves Raimond, and M. Sandler, "An Ecosystem for Transparent Music Similarity in an Open World," in *Proceedings of the International Conference on Music Information Retrieval 2009*, Oct. 2009, pp. 33–38.

[18] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, and Z. Ives, "DBpedia: A Nucleus for a Web of Open Data," in *Proceedings of the 6th International Semantic Web Conference*, 2007.

[19] C. Goble, J. Bhagat, S. Aleksejevs, D. Cruickshank, D. Michaelides, D. Newman, M. Borkum, S. Bechhofer, M. Roos, P. Li *et al.*, "myExperiment: a repository and social network for the sharing of bioinformatics workflows," *Nucleic Acids Research*, 2010.

[20] D. McEnnis, C. McKay, and I. Fujinaga, "Overview of OMEN," in *Proceedings of the International Conference on Music Information 2006*, 2006, pp. 7–12.