Combining Social Music and Semantic Web for music-related recommender systems

Alexandre Passant[1], Yves Raimond[2]

[1] DERI, National University of Ireland, Galway,
IDA Business Park, Lower Dangan,
Galway, Ireland,
alexandre.passant@deri.org
[2] Center For Digital Music,
Queen Mary, University of London,
England,
yves.raimond@elec.qmul.ac.uk

Abstract. This paper introduces various ways to suggest music-related content on the Web thanks to Semantic Web technologies. Rather than focusing on features of musical signals or running statistical analysis over listening habits, we detail how social networking, user contributions, and other interlinked data published within the scope of the Linking Open Data initiative can be combined to provide *data-rich* recommendations.

Key words: Web 2.0, Music, FOAF, SIOC, MOAT, Linked Data, Recommendation systems

1 Introduction

Recent Web 2.0 trends introduced new paradigms regarding the way information is produced on the Web. Users were mainly consumers of content, but now tend to become producers by spontaneously publishing and exchanging data [9]. In particular, sharing musical tastes is a frequently used practice. For instance, Last.fm[3] allows its users to publish their listening habits on the Web and offers musical recommendation based on these behaviors. MySpace[4] allows its users to declare themselves as friends of artists that are member of the platform – since many of them are using it to interact with their audience or to promote their work – while Facebook allows one to declare himself as a "fan" of someone. Moreover, people can blog about the latest gig they have seen, tag their favourite band pages on del.icio.us[5] or maintain discographies of artists on Wikipedia[6]. Thus, while the term *social music* was coined by last.fm to mainly identify the act of sharing musical tastes, we will here refer to the publishing and sharing of

[3] http://last.fm/
[4] http://www.myspace.com/
[5] http://del.icio.us/
[6] http://wikipedia.org/

music-related data on the Web, whatever the format is: blog posts, wiki pages, community databases, mp3s or playlists.

In this paper, we will see how *social music* data can be leveraged to the Semantic Web and how it can be used to let people find related musical-content regarding what they are producing or consuming. Rather than focusing on well-known practices for musical recommendations system (that will be described at the end of this paper), our approach consists in using relationships between various types of data (social networks, published content, tags, artist information, etc.) that have been modeled in RDF from those websites. The rest of the paper is organized as follows. In the first section, we describe current issues with music-related social networks and we see how FOAF and linked data can be used to provide a completely open and distributed *social graph*. We also see how SIOC can be used to model user contributions on the Web. In the second part, we overview the music-related part of the Linking Open Data project, providing machine-understandable and interlinked description of artists, bands, and so on. We also see how MOAT can help to move from simple tagging systems to semantic indexing using reference web identifiers. We then describe how the results of these different efforts can be used to suggest related content, whether it is music itself, blog posts, wiki pages... Finally, we go through an overview of existing musical recommendation systems and compare it with the ideas exposed in this paper. Then, we conclude the paper with overview of some future works.

2 Weaving social networks and music-related social data to the Semantic Web

2.1 From closed-world data silos to open social networks

One well-known feature of Web 2.0 websites is the ability to define and maintain social networks. These applications let their users add some other users in their network so that they can receive updates, send direct messages or access some semi-private data. Yet, most of the available social networking sites are isolated amongst each other. When subscribing to a new site, a user must invite his friends again, even if he already defined his network on another website. The process must be repeated each time a new website is joined, leading to what has been called *Social Network Fatigue*[7]. Another common limitation is that it is impossible to add as a friend on a particular service someone who is registered on another service.

Weaving these social networks into the Semantic Web can solve this problem. FOAF – Friend Of A Friend [5] –, a well-known vocabulary to describe agents and their relationships is especially adapted for this use-case. Various exporters for social websites have been written, as for Flickr [14], Last.fm[8] or MySpace[9], providing representations of social networks using FOAF and the

[7] `http://slashdot.org/articles/07/01/02/237223.shtml`
[8] `http://dbtune.org/last-fm/`
[9] `http://dbtune.org/myspace/`

`foaf:knows` relationship. Such networks can therefore be queried uniformally using the SPARQL query language, instead of relying on a specific API for each service. Moreover, various networks can be merged together thanks to resource consolidation, by explicitly defining `owl:sameAs` properties, or relying on the implicit use of `owl:InverseFunctionalProperties`. Such interlinking between various social networks provide a complete distributed and open *social graph*[10], that can be then queried and processed in an uniform way.

2.2 Representing Web 2.0 content on the Semantic Web

As for social networks, the content produced by users of Web 2.0 is held within closed systems, and not represented uniformally. There is a need for a shared semantics in order to represent in a common way user-generated content coming from multiple places. SIOC – Semantically-Interlinked Online Communities [4] – achieve this goal, by offering a model to represent activities of online communities and their contributions. More than 40 applications are now available for SIOC, both for creating or browsing and querying data, from PHP APIs to dedicated SIOC browsers[11].

Furthermore, as Web 2.0 content is not limited to textual data, SIOC defines a Type module[12], so that new types of user-generated content can be easily specified in data exporters, sometimes by reusing terms from other domain-specific ontologies (for example, `dcmi:MovingImage` or `mo:Playlist`). The following example represents a playlist using SIOC and Music Ontology terms.

```
:myRadio a mo:Playlist ;
  mo:track :song1 ;
  sioc:has_creator :me ;
  sioc:site <http://lastfm.com> ;
  dc:title ‘‘Alex’s last.fm playlist’’ .
:song1 a mo:Track ;
  dc:title ‘‘Monkey Man’’ ;
  foaf:maker dbpedia:The_Specials .
```

3 Music and the Linking Open Data project

The Linking Open Data community project [3] aims at publishing and interlinking a wide range of open data sources, by following the four principles outlined by Berners-Lee in [2]: (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 and (4) Include links to other URIs so that they can discover more things. For example, the Flickr exporter mentioned above provides

[10] `http://www.bradfitz.com/social-graph-problem/`
[11] `http://sioc-project.org/applications`
[12] `http://rdfs.org/sioc/types`

the geolocation of the user using a link to a Geonames resource, allowing an user agent to jump from one dataset to another.

A wide range of music-related data sources have been interlinked within the Linking Open Data initiative [17]. For example, the DBTune project[13] exports the datasets depicted in Fig. 1 in RDF. These datasets encompass detailed editorial information, geolocation of artists, social networking information amongst artists and listeners, listening habits, Creative Commons content, public broadcasting information, and content-based data (eg. features extracted from the audio signal characterising structure, harmony, melody, rhythm or timbre, and content-based similarity measures derived from these). These datasets are linked to other ones. For example, Jamendo is linked to Geonames, therefore providing an easy to build geolocation-based mash-up for musical data. Artists within Musicbrainz are linked to DBPedia artists, MySpace artists, and artists within the BBC Playcount data.

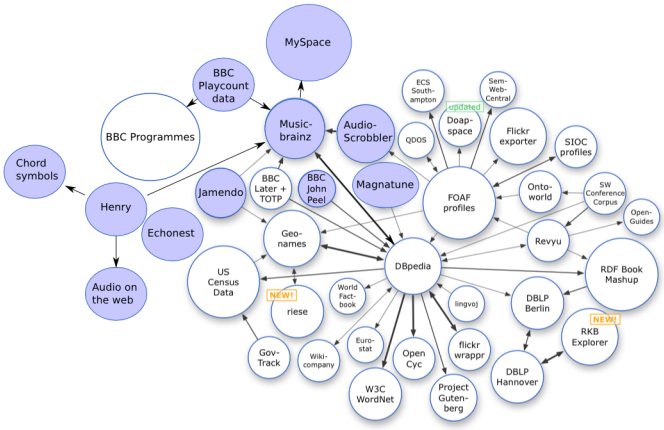

Fig. 1. Depiction of different music-related datasets and their interlinks. The coloured circles correspond to dataset made available within the DBTune project.

4 Using distributed RDF data for musical recommendation

We now have, to the extent of our knowledge, the largest publicly available distributed database of music-related information. We detail in the following the usefulness of such a database for musical recommendations.

[13] `http://dbtune.org/`

4.1 Social-networking and music recommendations

Using social-networks is a first method to suggest musical recommendations, based on listening habits in the network. The Music Ontology (MO [16]) and the related Event Ontology can be used to represent listening habits, as done on the last.fm exporter. Another way is to rely on the `foaf:topic_interest` to provide a direct link between someone and his interests, as artists or bands URIs. An inference rule can be defined to go from the first representation to the second one, represented as follows (using the SPARQL `CONSTRUCT` pattern):

```
CONSTRUCT { ?user foaf:topic_interest ?artist }
WHERE {
  ?ev a event:Event ;
     event:agent ?user ;
     event:factor ?track .
  ?track foaf:maker ?artist .
}
```

Once such `foaf:topic_interest` links have been defined, a single SPARQL query can be used to find listening habits of one's friends, so that they can be suggested to him:

```
SELECT ?artist
WHERE {
  <$uri> foaf:knows [
    foaf:foaf_interest ?artist .
  ] .
}
```

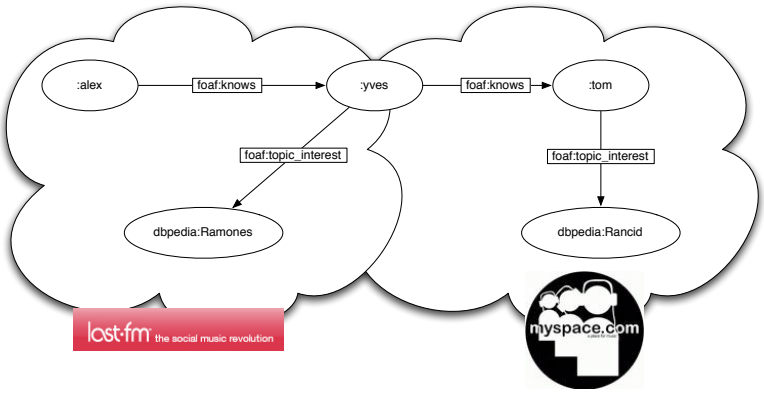

Fig. 2. Using distributed social-networks in recommendation systems

This query therefore implements a really simple collaborative filtering algorithm, that already exist in most of *social music* services. Going further, we can extend this relationships path, for instance to suggest tastes of friends of a friend, and so on. Yet, the most interesting aspect here is that, thanks to the distributed social graph model detailed in Sec. 2.1, the relationships paths can be distributed in various social networks. A recommender system is therefore not limited to a single network, but can combine the aggregation of various social networks and listening habits (or interests), unified using Semantic Web technologies, as outlined in Fig. 2.

4.2 Tag-based recommendations

Tagged content, by using the MOAT framework [15], can also be used to drive music-related recommendations. MOAT allows people to tag their content with URIs, rather than simple keywords, and can be used on existing tagged content, thanks to the LODr application[14].

Once people have tagged their data using URIs, we can rely on relationships between those URIs to suggest related data. For instance, a recommender system for music-related content might suggest to browse a picture tagged with the URI of Joe Strummer on Flickr when browsing a blog post about The Clash, since there is a relationship between both defined in DBpedia, as seen in Fig. 3. By leveraging existing Web 2.0 content on the Semantic Web, as for social networks, we break the bareers between applications and can use data combined, mashed and eventually interlinked through various paths and datasets, that are completely disconnected at a usual "Web of documents" level.

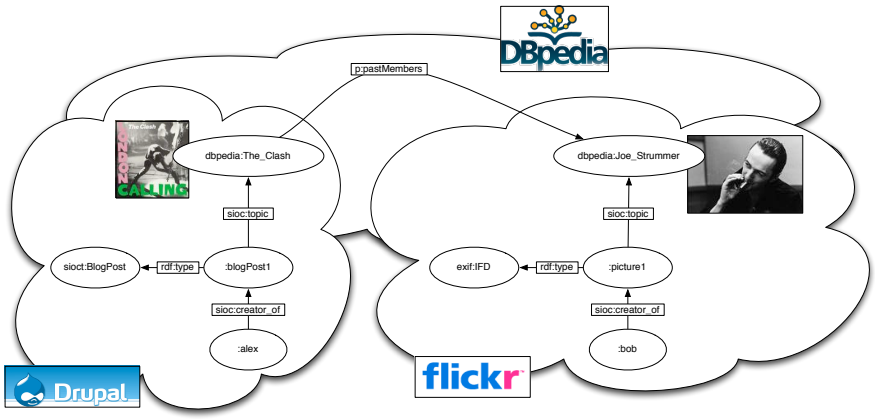

Fig. 3. Interlinking distributed tagged content

[14] http://lodr.info

4.3 Using LOD to find relevant musical content

Using RDF data available from the LOD cloud, various strategies can be adapted to find related content, especially bands or artists. While we can consider direct relationships, as suggesting someone to look for solo members when browsing a band's page, one of our interest is to exploit the structure of linked data and provide recommendations based on paths between artists and bands, for instance, because two bands played at the Cavern in Liverpool or at the CBGB in New York City. Here, one challenge is to find relevant paths between concepts, especially since a lot of various paths exist. As a first step, we analyzed 400 random bands and artist pages from DBpedia to find which are the most used properties, and see how they could be used for music-based recommendations. The 20 most popular are described in the table below (we voluntary excluded plain-text properties as using plain-text relationships may lead to noise and lack of precision):

Rank	Property	Number of relationships
1	`skos:subject`	1930
2	`rdf:type`	882
3	`dbpedia:reference`	847
4	`dbpedia:genre`	450
5	`dbpedia:page`	400
6	`dbpedia:hasPhotoCollection`	400
7	`dbpedia:origin`	355
8	`dbpedia:wikiPageUsesTemplate`	333
9	`dbpedia:label`	265
10	`dbpedia:wordnet_type`	194
11	`dbpedia:associatedActs`	189
12	`foaf:homepage`	178
13	`dbpedia:currentMembers`	151
14	`dbpedia:url`	114
15	`dbpedia:pastMembers`	108
16	`dbpedia:occupation`	97
17	`owl:sameAs`	95
18	`foaf:depiction`	89
19	`foaf:img`	89
20	`dpbedia:wikipage-de`	85

From this list of property, we voluntary excluded those leading to nonrelevant results (as `dbpedia:wikiPageUsesTemplate`), and designed a lightweight facet browser that suggest related artists based on the other properties (and their values) (Fig. 4). We may in the future rely on automatic extraction [12] so that the browsing interface could be automatically adapted to the changes in the Wikipedia structure (and so on the DBpedia one), and it would help to exclude property / values tuples leading to too many results. Another challenge regarding facets definitions is also to automatically find which paths might be

Fig. 4. Finding related artists using Linked Data

relevant to musical recommendations. Especially, future work includes the automatic generation of relevant recommendation paths in linked data. So far, our prototype is closer to a faceted browsing interface than a recommender system, although it demonstrates the richness of the available music-related linked data

As Fig 5 shows, there is a similar path between artists related because of their music genre of vegetarian habits. Yet, on some case this is not so trivial and may also depend on the interest of the users and musical styles. For instance, some might be interested when listening to The Ramones by being suggested to Listen to the Sex Pistols not because they are somehow musically-related by 3-chords songs, but because graph browsing would have tell us that members of both of them died from a drug overdose in a NYC hotel.

5 Comparison with existing approaches

Current music recommendation systems can be clustered in two main categories: collaborative filtering and content-based.

Collaborative filtering consists in recommending items to an user based on the stated tastes of other related users. For example, a user u_2 might like a song s_2 if he likes a song s_1, and that a user u_1 likes both s_1 and s_2. Usually, music recommendations service based on such a methodology use a closed set of information, gathered through a particular service, eg. listening habbits for last.fm

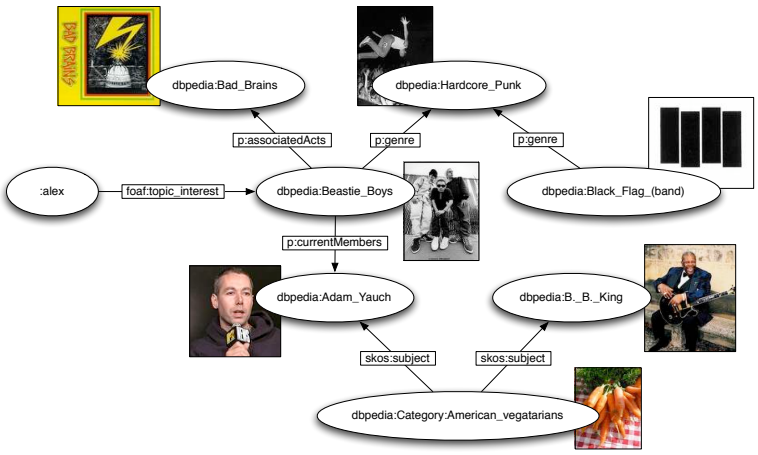

Fig. 5. Exploring different paths between artists

or consumer behaviors for Amazon. Some systems adapt similar ideas in open information environments, such as in [8], where structured taste information is extracted from web documents. Collaborative filtering can also be enhanced by integrating contextual data, reflecting the user's interests at a particular moment [10]. However, collaborative filtering does not tackle the problem of finding items within the *long tail* of music production — those for which the amount of taste data is limited. This approach is only suitable when a large number of users have heard the music and expressed their interest in it.

A recent trend for music recommendation is to use automated analysis of the musical items to drive recommendations, by modeling musical audio similarity [13]. For example, a particular song might be recommended to an user because its rhythm is similar to a song mentioned in his listening habbits. A system based on such a methodology will therefore be able to recommend items within the long tail [7] — it will be able to recommend unknown artists, as the system has no notion of 'popularity'. Current content-based recommendation systems focus on low-level representations of the musical audio content, relative to one of its facet (timbre, structure, melody, harmony or rhythm). However, higher-level representations may lead towards a more widespread adoption of such music recommenders.

Several works combine these two methodologies (collaborative filtering and content-based) to achieve better recommendation [1,18]. Indeed, content-based recommendation can help for the *cold start* problem of collaborative filtering — if the user made no statements about his tastes at first, it will be impossible to provide him with an acceptable recommendation and therefore obtaining taste data from him. The Foafing-the-music project [6] is particularly related to our

work, as it uses distributed social networks using FOAF as well as content-based data available in RDF.

We believe our approach, using not only these two methodologies (collaborative filtering and content-based similarities), but also a wide range of interlinked data in multiple domains, allows the user to get much more justified recommendations [11], such as the ones detailed in sec 4.3. We therefore move towards *data-rich* musical recommendations.

6 Conclusion and future works

In this paper, we outlined various ways to extract RDF data from social music websites and see how it can be used in recommendation systems. While we defined various strategies (based respectively on social networks, tagged data and data published and interlinked within the Linking Open Data community project), they can be combined together for advanced querying and suggesting data, especially as the underlying models are also interlinked (for instance, SIOC and MO reuses the FOAF ontology and MOAT allows to link SIOC data to LOD URIs). Using this global data integration, a query could be used to suggest all bands that will play in your town next month for less than 15euros, that one of your friend blogged about and that play cover songs from a band that you listened to more than 10 times on the last week. Going further, wen can also combine content-based similarity information and interest data as in the following query. The Henry[15] dataset within DBTune indeed publishes `mo:similar_to` statement between musical audio items, based on timbre similarity, so that we could also imagine SPARQL queries using extracted keys and rhythms published by Henry to generate a playlist of recommendations with smooth transitions.

```
SELECT ?track1 ?track2
WHERE {
  <$me> foaf:topic_interest ?artist .
  ?artist foaf:made ?track1 .
  ?track1 mo:available_as ?audio .
  ?track2 mo:available_as ?audio2 .
  ?audio mo:similar_to ?audio2 .
}
```

Regarding those kind of query, relying on highly-distributed data, while not directly related to our work, some efforts should be made regarding dedicated crawlers or distributed querying systems as well as using voiD[16] to find the relevant datasets.

Also, a lot of future work still needs to be done to find the best way to browse the graph of linked data to provide good recommendations. Using attention and taste data available in a user's profile is a first step towards it, but we could also

[15] `http://dbtune.org/henry/`
[16] `http://community.linkeddata.org/MediaWiki/index.php?MetaLOD`

imagine using lots of other data, such as political interests, interests in other arts, etc. While we worked on a first implementation based on data available from DBpedia, we plan to extend it to re-use other datasets and make it publicly available. Especially, future work includes the evaluation of a recommendation system based on such approach, and compare it to existing approaches, to see how it can augment the user experience of discovering musical content.

Acknowledgments

This material is based (in part) upon works supported by the Science Foundation Ireland under Grant No. SFI/02/CE1/I131.

References

1. Marko Balabanovi and Yoav Shoham. Fab: content-based, collaborative recommendation. *Communications of the ACM*, 40(3):66–72, March 1997.
2. Tim Berners-Lee. Linked data. World wide web design issues, July 2006.
3. Chris Bizer, Tom Heath, Danny Ayers, and Yves Raimond. Interlinking open data on the web. In *Poster, 4th Annual European Semantic Web Conference (ESWC2007), Innsbruck,Austria*, 2007.
4. J.G. Breslin, A. Harth, U. Bojars, and S. Decker. Towards Semantically-Interlinked Online Communities. *2nd European Semantic Web Conference*, May 2005.
5. Dan Brickley and Libby Miller. FOAF Vocabulary Specification. Namespace Document 2 Sept 2004, FOAF Project, 2004. http://xmlns.com/foaf/0.1/.
6. O. Celma, M. Ramirez, and P. Herrera. Foafing the music: A music recommendation system based on rss feeds and user preferences. In *Proceedings of the International Conference on Music Information Retrieval*, 2005.
7. Oscar Celma and Pedro Cano. From hits to niches? or how popular artists can bias music recommendation and discovery. In *2nd Workshop on Large-Scale Recommender Systems and the Netflix Prize Competition (ACM KDD)*, Las Vegas, USA, August 2008.
8. William W. Cohen and Wei Fan. Web-collaborative filtering: recommending music by crawling the web. *Computer Networks*, 33(1-6):685–698, June 2000.
9. Dan Gillmor. *We the Media: Grassroots Journalism by the People, for the People*. O'Reilly Media , Inc.,, Sebastopol, CA, August 2004.
10. C. Hayes. *Smart Radio: Building community-Based Internet Music Radio*. PhD thesis, Trinity College Dublin, October 2003.
11. Jonathan L. Herlocker, Joseph A. Konstan, and John Riedl. Explaining collaborative filtering recommendations. In *CSCW '00: Proceedings of the 2000 ACM conference on Computer supported cooperative work*, pages 241–250, New York, NY, USA, 2000. ACM.
12. Eyal Oren, Renaud Delbru, Knud Möller, Max Völkel, and Siegfried Handschuh. Annotation and navigation in semantic wikis. In Max Völkel, editor, *Proceedings of the First Workshop on Semantic Wikis – From Wiki*, JUN 2006.
13. Elias Pampalk. *Computational Models of Music Similarity and their Application in Music Information Retrieval*. PhD thesis, Vienna University of Technology, 2006.

14. Alexandre Passant. :me owl:sameAs flickr:33669349@N00. In *Proceedings of the WWW 2008 Workshop Linked Data on the Web (LDOW2008), Beijing, China*, Apr 2008.
15. Alexandre Passant and Philippe Laublet. Meaning Of A Tag: A collaborative approach to bridge the gap between tagging and Linked Data. In *Proceedings of the WWW 2008 Workshop Linked Data on the Web (LDOW2008), Beijing, China*, Apr 2008.
16. Yves Raimond, Samer Abdallah, Mark Sandler, and Frederick Giasson. The music ontology. In *Proceedings of the International Conference on Music Information Retrieval*, pages 417–422, September 2007.
17. Yves Raimond and Mark Sandler. A web of musical information. In *Proceedings of the International Conference on Music Information Retrieval*, Philadelphia, USA, 2008.
18. Kazuyoshi Yoshii, Masataka Goto, Kazunori Komatani, Tetsuya Ogata, and Hiroshi G. Okuno. Hybrid collaborative and content-based music recommendation using probabilistic model with latent user preferences. In *Proceedings of the International Conference on Music Information Retrieval*, 2006.