CONNECTING THE DOTS: MUSIC METADATA GENERATION, SCHEMAS AND APPLICATIONS

Nik Corthaut, Sten Govaerts, Katrien Verbert, Erik Duval
Katholieke Universiteit Leuven, Dept. Computer Science
{nik.corthaut, sten.govaerts, katrien.verbert, erik.duval}@cs.kuleuven.be

ABSTRACT

With the ever-increasing amount of digitized music becoming available, metadata is a key driver for different music related application domains. A service that combines different metadata sources should be aware of the existence of different schemas to store and exchange music metadata. The user of a metadata provider could benefit from knowledge about the metadata needs for different music application domains. In this paper, we present how we can compare the expressiveness and richness of a metadata schema for an application. To cope with different levels of granularity in metadata fields we defined clusters of semantically related metadata fields. Similarly, application domains were defined to tackle the fine-grained functionality space in music applications. Next is shown to what extent music application domains and metadata schemas make use of the metadata field clusters. Finally, we link the metadata schemas with the application domains. A decision table is presented that assists the user of a metadata provider in choosing the right metadata schema for his application.

1. INTRODUCTION

Metadata plays an important role in MIR research. In 2007, we proposed a semi-automatic approach for the generation of music metadata [2] in the rockanango project [1]. In this approach, we bundled the power of computational techniques, the wisdom of the crowds and the experience of the music experts of Aristo Music (www.aristomusic.com), the company with whom we collaborate. This was needed because of the increasing volume of music that needs annotation within a reasonable amount of time and cost.

A diversity of available metadata web services can be used to facilitate the annotation process. Amazon.com has a substantial web API to retrieve metadata about their products. Available music metadata for CD's includes reviews from editors and customers, track listings, release dates, genres, labels, popularity and similar items. Last.fm (http://last.fm), the well-known social recommendation music service, provides a web service to retrieve user, artist, group and tag data. The All Music Guide is also a very rich source of music metadata. MusicBrainz (http://www.musicbrainz.org) is a community based music metadata database used for CD recognition. It contains artists, tracks, labels and releases. Discogs (http://discogs.com) offers a neat web service with good pointers to different production houses and publishers.

Another kind of web service is available that, given an audio file, returns the results of different signal processing algorithms. Echo Nest (http://analyze.echonest.com/) recently released a web service that returns a set of musical features, like timbre, pitch, and rhythm. It is aiming at different applications, such as visualizations, games and DJ-software. Furthermore, Echo Nest has been developing software that 'listens' to music and tries to predict which songs will become hits. The MIR Group of the Vienna University of Technology (http://www.ifs.tuwien.ac.at/mir/webservice/) also made a web service available that returns a set of musical features for a given song (e.g. rhythm patterns, statistical spectrum descriptors and rhythm histograms) and allows the training of self-organizing music maps. In conclusion, these systems provide diverse and useful metadata, however not without a substantial amount of overlap.

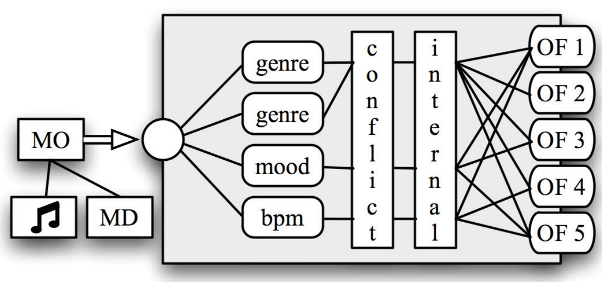

Figure 1. Internals of the metadata framework.

We want to build a framework that makes use of the strength of the different systems (see Figure 1). The input is a musical object (MO) that consists of an audio file (the signal) and possibly some initial metadata (MD). When it enters the system, a federated request for metadata among available generators is done. Generators can use available a priori metadata to enhance predictions. To cope with potentially conflicting results, [3] suggests different conflict resolution approaches. The resulting metadata is stored in an internal format. The generated metadata can cover the full range from low-level features, over factual data e.g. the arrangement of the band, to affective metadata e.g. the mood, similar songs or popularity.

By offering the choice between different existing music metadata formats (OF), we enable the reuse of existing tools and parsers to handle the generated metadata. To be able to store all the generated metadata we are looking for a suitable metadata schema for internal use in the metadata framework. In this paper we will select a set of metadata standards relevant for the task, investigate in which application domains they are useful and evaluate their descriptive richness.

The description of large music works, e.g. the complete opus of a composer, is beyond the scope of this paper. Likewise metadata schemas for metadata exchange (e.g. METS [18]) or rights (e.g. ODRL [19]) are also out of the scope. We focus on descriptive metadata schemas.

2. COMPARISON OF METADATA SCHEMAS

The goal of collecting metadata is to enable functionalities in an application. To understand the usefulness of a metadata schema in different cases, we will investigate which metadata is involved in different use cases. This will be the foundation for defining music application domains and selecting a number of relevant metadata schemas. For the sake of easier comparison, we introduce a level of abstraction to the metadata fields by means of field clustering. The three elements: the selected metadata standards, the different application domains and categories of metadata fields will be introduced in this section. The question which metadata schemas to use when building an application that offers a certain functionality can then be answered through the comparison of the three elements.

2.1. Application domains

Music metadata can be anywhere in the production cycle of music, whether it be copyright information about a drum loop while composing or the number of searches for a musical piece at an online store. For the use of metadata schemas, we limit our application domains to software.

Clustering software applications into application domains allows easier comparison between different applications. The actual clustering is based on the actions people perform when handling music [4] [14]. The eight clusters cover the production cycle from conception (composition) over consumption to transactions.

- *Music library/encyclopedia:* software systems for physical libraries, encyclopediae or companies that license music, describing factual knowledge for large collections of music, e.g. AllMusic Guide.
- *Personal collection management:* software to organize your music collection, e.g. iTunes, Winamp media library, Collectorz.com Music Collector.
- *Commerce and transactions:* applications involved in the act of shopping for music, this includes presenting songs, searching and trading, e.g. iTunes Music Store, Amazon, Magnatune.
- *Music editing/production:* the tools deployed in the creation and adaptation of music, e.g. Logic Pro.
- *Music playback:* applications that render music files to its audible form, e.g. your favorite music player.
- *Music recommendation:* services for discovery of new and similar music, e.g. Pandora, Last.fm.
- *Music retrieval:* search and identification tools with different query interfaces in all their forms, e.g. query by humming.
- *Musical notation:* creation and manipulation tools for musical scores, e.g. Sibelius, WAV2Midi.

A real-life application will most likely use a number of application domains, e.g. playlist generation can be classified as music recommendation and library functionality, some music players also offer management of the personal music collection.

2.2. Metadata standards

Metadata standards originate from different sources: it can evolve out of the design of an application and through wide adoption become a de facto standard, it can be focused on interoperability or it can be designed as a standard from the start. No single metadata standard is at the moment available for music covering all the possible requirements. Based on industry standards and the use in ongoing research we selected eight music metadata standards. The methodology presented in this paper is applicable to the many other available standards (e.g. MARC [20], MODS [21]). Future work includes extending the comparison with other relevant schemas.

- *ID3:* An ID3-tag is a data container of a prescribed format embedded within an audio file. The stored data can contain the artist name, song title and genre of the audio file. ID3 has a wide spread use in music players and devices like iTunes, iPod and Winamp.
- *FreeDB:* is an online database to look up CD information by calculating a unique ID for a CD to query the database. FreeDB is a community version of the commercial Gracenote service; both are used in a variety of playback and MP3-ripping software.
- *MusicBrainz:* the scope of MusicBrainz [6] is the same as FreeDB, but they have a moderated database and use identification on track level. MusicBrainz is an RDF-based [7] web service.
- *Dublin Core:* is a standard for cross-domain information resource description, which provides a simple and standardized set of conventions for

describing things online [11]. It is widely used to describe digital multimedia materials.

- *Music Vocabulary:* describes classic musical works and performances in an RDF-ontology [8], defining classes for musical works, events, instruments, performers and relationships between them.
- *Music Ontology:* the goal is to link information about artists, albums and tracks together and express relationships between musical information for the Semantic Web [10]. The Music Ontology is based on Functional Requirements for Bibliographic Records (FRBR) [13], but removes the Expression entity to be able to model the creational process behind music and is available as an RDF-ontology.
- *MPEG-7:* is a multimedia content description standard. MPEG-7 [9] provides standardized tools for describing different aspects of multimedia at different levels of abstraction. The XML-based syntax enables interchange across applications, but without precise semantics metadata interoperability can be a problem.

2.3. Metadata field clusters

Different metadata standards describe different things. To facilitate the comparison of the standards we clustered the metadata fields in an iterative process. First we clustered the elements of ID3 in semantically related clusters. Next, we tried to map Music Ontology onto the ID3 clustering. This resulted in a small adjustment in the clusters. The same procedure was used to add the remaining metadata standards. The final clustering defines semantically related elements, independent of metadata schemas or metadata fields. The following clusters are defined:

- *Musical info:* musical properties of the audio signal, e.g. bpm, duration, key.
- *Classifiers:* categorizers for music organization, e.g. genre.
- *Performance:* descriptions of the people that are involved in a musical performance, e.g. artists, conductor, engineers.
- *Versioning:* descriptions for different versions of the musical work and relationships between the involved people and the related works, e.g. original artists, contains sample from, remix of.
- *Descriptor:* describing the musical work, e.g. title.
- *Rights & ownership:* information about intellectual property management, e.g. license, owner.
- *Playback rendition:* information useful for rendering of the music file, e.g. relative volume adjustment.
- *Lyrics:* text of the musical work and related information, e.g. translated lyrics, synched lyrics.
- *Grouping & referencing:* data clustering structures and linking to web sites and other works, e.g. wikipedia link, album listing structure.
- *Identifiers:* identification keys, e.g. ISRC-code.
- *Record-info:* information about the recording and the involved people, e.g. recording type, album name.
- *Instrumentation & arrangement:* about the used instruments and orchestration, e.g. instrument type.
- *Sound & carrier:* information about the available media of a musical work, e.g. media type.
- *Event:* information about public performances of the musical work, e.g. festival program, place.
- *Time-modeling:* metadata structures to express points in time, e.g. interval, at duration.
- *Musical notation:* everything for a symbolic representation of music, e.g. clefs, measures, notes.
- *Attention-metadata & usage:* information about the user's attention to music, e.g. times played.
- *Publishing:* information about the publishing process, e.g. record label.
- *Composition:* information about the composition process, e.g. movement, composer.
- *Production:* information about the creative process of the production and its actors, e.g. producer.
- *Meta-metadata:* information about the metadata, e.g. who annotated the metadata.

2.4. Comparison

In this subsection we determine how the clusters, the schemas and the application domains relate to one another. As said in 2.1, we are mainly interested in determining which metadata schema to use for some desired functionality. We use the level of abstraction of the metadata clusters to join metadata schemas and application domains.

2.4.1. Application domains vs. metadata field clusters

In table 1 we compared the music metadata fields with the application domains. When assigning music metadata field clusters to application domains, the different domains are considered from the perspective of the provider of the functionality. We used the results of survey [4] on how users search for the library and encyclopedia domain. For the personal music collection management domain, we used a combination of the survey presented in [5] and we looked at the applications in the domain. For the other domains, we investigated the metadata usage of different applications in these domains to determine to what degree which clusters are used.

	libr./enc.	pers. coll.	comm.	edit./prod.	playback	recomm.	retrieval	notation
musical info	-	+/-	-	+	+/-	+	+	+
classifiers	+	+	+	-	-	+	-	-
performance	+	+	+	+	+/-	+	+	-
versioning	+	+	+	+/-	-	+	+	-
descriptor	+	+	+	+	+	+	+	+
rights & ownersh.	+	+	+	+	+/-	-	-	-
playback rendition	+/-	+/-	-	-	+/-	-	-	-
lyrics	+	+	-	+	+/-	+	+	+
group. & refs.	+	+	+	+/-	-	+	-	-
identifiers	+	+	+	-	-	+/-	+/-	-
record-info	+	+	+	-	-	+	+/-	-
instr. & arrang.	+	+/-	+/-	-	-	+	-	+
sound & carrier	+	+	+	-	+/-	-	-	-
event	+	+/-	+/-	-	-	+/-	-	-
time-modelling	-	-	-	+	-	-	+/-	+
notation	-	-	-	+/-	-	-	+	+
attention-metadata	-	+	-	-	+	+	+/-	-
publishing	+	+	+	-	-	+/-	-	-
composition	+	+	+	+	-	+	-	+
production	+	+	+	+	-	+	-	-
meta-metadata	+	+	+	+	-	-	-	-

Table 1. Metadata field clusters vs. application domains

Based on Table 1, we can conclude that library and encyclopedia, personal collection management, commercial applications and recommendation make use of the largest set of metadata clusters. The bibliographical nature of libraries strives for completeness and accounts for a wider range of clusters. For collections, recommendation and commercial use, [16] and [4] suggest wide ranges of user goals, often with a focus on enjoyment that require a broad range of information. In the latter case, rights management and information about the carrier is relevant. Information closely describing the actual audio signal found in musical info is not so relevant here as opposed to e.g. playback where descriptive metadata typically is not needed to render an audio file.

Recommendation and retrieval rely both on signal and metadata in order to function. Pachet [16] defines 3 music metadata categories: editorial, cultural and acoustic. Music recommendation techniques can make use of elements out of these 3 categories [17].

Music notation is useful, either during composition, early in the music creation cycle, or while describing already existing music by means of transcription. Information about publishing, producing, recording, etc. are not immediately relevant.

The table can be used as a reference for the metadata needed when building an application.

2.4.2. Metadata standards vs. metadata field clusters

	ID3	freeDB	MusicBrainz	Dublin Core	Music Vocabulary	Music Ontology	MPEG-7	MusicXML
musical info	+	+/-	+/-	-	+/-	+	+	+
classifiers	+/-	+/-	-	+/-	+/-	+/-	+	+/-
performance	+	+/-	+/-	+	+	+	+	+/-
versioning	+	-	-	+	+/-	+	+	-
descriptor	+	+	+	+	+	+	+	+
rights & ownersh.	+	-	-	+	-	+	+	+
playback rendition	+	-	-	-	-	+	+	+
lyrics	+	-	-	-	+	+	+	+
group. & refs.	+	-	-	+	+/-	+	+	+
identifiers	+	+	+	+	-	+	+	-
record-info	+	+	+	-	-	+	+	-
instr. & arrang.	+/-	-	-	-	+	+	+	+
sound & carrier	-	-	-	+	-	+	+	+/-
event	-	-	-	-	+	+	+	-
time-modelling	-	-	-	-	-	+	+	+
notation	-	-	-	-	+/-	+/-	-	+
attention-metadata	+	-	-	-	-	+	+	-
publishing	+	-	-	-	-	+	+	-
composition	+	-	-	+/-	+	+	+	+
production	-	-	-	-	-	+	+	-
meta-metadata	-	+	-	-	-	-	+	-

Table 2. Metadata field clusters vs. metadata standards

By doing the clustering of the metadata fields for each metadata standard (see 2.3), we obtain a table with the fields of each metadata standard in the corresponding clusters. This table enables us to determine how good the metadata clusters are represented in the different metadata standards, as can be seen in table 2.

From table 2, we can clearly see that the scope of Dublin Core is more general than music alone, because the clusters typically related to music are not present. MPEG-7 and Music Ontology are the most versatile schemas that are present in almost all clusters. MusicBrainz and freeDB have similar fields and scope.

Some clusters are better represented in some standards, like the performance cluster in Music Vocabulary, Music Ontology and MPEG-7.

2.4.3. Application domains vs. metadata standards

	libr./enc.	pers. coll.	comm.	edit./prod.	playback	recomm.	retrieval	notation
ID3	70	75	70	67	90	81	81	64
freeDB	30	31	36	29	30	30	38	21
MusicBrainz	21	22	25	19	30	26	38	21
Dublin Core	48	47	57	43	50	41	44	21
Music Vocabulary	45	40	39	50	45	56	56	71
Music Ontology	91	91	89	88	100	96	94	93
MPEG-7	100	100	100	95	100	100	88	86
MusicXML	48	47	43	71	70	52	63	100

Table 3. Application domains vs. metadata standards in %

We use the first two tables to determine which metadata schema is most apt, given the application domain.

To be able to compute the decision table we formalized table 1 and 2. The +, +/- and - signs are converted to 1, 0.5 and 0 respectively. This results in 2 matrices: MCAD (table 1) of dimension (#metadata field clusters (MC) x #application domains) and a matrix MCMS (table 2) of dimension (#metadata field clusters x #metadata standards). The sum of the columns of MCAD results in the maximum score any metadata standard can reach for an application domain, a vector mssum.

$$mssum_j = \sum_{i=1}^{\#MC} MCAD_{i,j} \qquad (1)$$

We now calculate the product of the transposed MCAD and MCMS in (2). This measures how well a metadata standard covers the clusters for each application domain.

$$R_{i,j} = \sum_{x=1}^{\#MC} MCAD^T_{i,x} . MCMS_{x,j} \qquad (2)$$

The mssum vector is used to normalize the values for the different schemas (3) and this results in a matrix MFAD, comparing metadata schemas with application domains, see table 3 for the actual values.

$$MFAD_{i,j} = \frac{R_{i,j}}{mssum_j} \qquad (3)$$

Note that we do not penalize excess information, because most of the metadata fields are not obligatory. If they are not needed, they can be omitted.

3. ANALYSIS OF THE DECISION TABLE

It is clear that MPEG-7 scores well for all application domains. Does this mean that this is the über music metadata standard that should be the only output format for a metadata generation framework? No, though it is possible to store any kind of information in MPEG-7 through the use of description schemas and the description definition language. It might be the case that the more limited set of e.g. Dublin Core is sufficient. The same is valid for MusicOntology.

Some interoperability issues with MPEG-7 have been signaled [15]. Semantically identical data can be represented in multiple ways, e.g by keywords, free text and a structured annotation with labels. Another issue is that the intended semantics of description structures are not always clear in MPEG-7.

FreeDB and MusicBrainz seem to score low overall, but seem best suited for music retrieval. Both are deployed in online services focus on identification of songs/albums. The available information linked to the identifier consists mainly about artist, track and album related metadata. Applications will use that data to fill in other metadata schemas, for example an ID3 tag.

As expected, MusicXML is extremely relevant for notation. This is the main goal of the standard. With the existence of tools that convert MusicXML to MIDI, it can also be used in playback.

Music Vocabulary's goal is to define a vocabulary to describe classical music masterworks and performances/concerts [6]. This limited scope has severe consequences for the general expressiveness of the schema. The fact that it scores relatively high on music notation is mainly because of its elaborate description of composition and orchestration/instrumentation. It could be used together with Music Ontology, to enrich its classical work descriptions.

Dublin core is designed to offer simple cross-domain metadata fields. This leaves very little margin for music domain specific information. However, often a simple, interoperable description is desired for an application. MusicXML, Music Ontology and Music Vocabulary use parts of Dublin Core, for example dc:author.

The creators of ID3 use 'The audience is informed' as their motto. This is visible in the good performance in the playback application domain. It scores equally well for recommendation applications, due to its expressiveness in classifiers, descriptors, performance and versioning.

4. FUTURE WORK

The implementation of a music metadata framework as described in Figure 1 is currently under development. It will be integrated in the SAmgI framework [3], where metadata for learning content is gathered, combined and presented to the end user in multiple formats, e.g. LOM, Dublin Core and human readable. The fundamentals for conflict resolution and management of different generators are already present. Furthermore SAmgI runs as a Java web service that embraces distributed metadata generation. The main challenge is adapting the internal structure for music specific metadata fields.

5. CONCLUSION

In conclusion of this study on the expressiveness and richness of different music metadata schemas and their relation with application domains, we can state that the formalization of these relations offers two practical uses in the context of music metadata generation. Firstly, once an application developer has decided on the functionality, he can determine what parameters (clusters) he should retrieve from a metadata generation framework based on the desired application domains. Secondly, the metadata framework developer can decide on the metadata formats to be returned given the set of requested parameters.

The scope of this work is not to provide a metadata schema decision table to the detail of individual fields of the metadata schemas. However it provides useful insights in whether schemas are relevant for certain application domains. The granularity of functionality can be refined after a first selection, but this requires deep knowledge of the considered application.

6. ACKNOWLEDGEMENT

We gratefully acknowledge the funding by IWT-Vlaanderen under grant: IWT 060237 Music Metadata Generation.

7. REFERENCES

[1] N. Corthaut, S. Govaerts, E. Duval. “Moody Tunes: The Rockanango Project”, *Proceedings of the 7th International Conference on Music Information Retrieval*, Victoria, Canada, 2006, pp. 308-313.

[2] S. Govaerts, N. Corthaut, E. Duval. “Mood-ex-Machina: Towards Automation of Moody Tunes”, *Proceedings of the 8th International Conf. on Music Information Retrieval*, Vienna, 2007, pp. 347-350.

[3] C. Kardinaels, M. Meire, E. Duval. “Automating Metadata Generation: the Simple Indexing Interface“, *Proceedings of the 14th international conference on World Wide Web*, pp. 548-556, 2005.

[4] J. H. Lee, J. S. Downie. “Survey of Music Information Needs, Uses and Seeking Behaviours: Preliminary Findings”, *Proceedings of the 5th International Conference on Music Information Retrieval*, 8 pages, Barcelona, Spain, 2004.

[5] S. J. Cunningham, M. Jones, S. Jones. “Organizing Digital Music for Use: an Examination of Personal Music Collections”, *Proceedings of the 5th International Conference on Music Information Retrieval*, 8 pages, Barcelona, Spain, 2004.

[6] A. Swartz. “MusicBrainz: A Semantic Web Service”, *IEEE Intelligent Systems*, vol.17, no.1, pp.76-77, 2002.

[7] O. Lassila and R. Swick, “Resource description framework (RDF) model and syntax specification”, 1998, W3C, available: citeseer.ist.psu.edu/article/lassila98resource.html

[8] Masahide Kanzaki, Music Vocabulary, available: http://www.schemaweb.info/schema/SchemaDetails.aspx?id=162

[9] J.M. Martinez, R. Koenen, F. Pereira, “MPEG-7: the generic multimedia content description standard, part 1”, *Multimedia, IEEE*, vol.9, no.2, pp.78-87, 2002.

[10] Y. Raimond, A. S. Abdallah, M. Sandler, F. Giasson. “The Music Ontology”, *Proceedings of the 8th International Conference on Music Information Retrieval*, Vienna, Austria, 2007, pp. 417-422.

[11] Dublin Core Metadata Element set, 2008-01-14, http://dublincore.org/documents/dces/

[12] M. Good, “MusicXML: An Internet-Friendly Format for Sheet Music”, *XML Conf. & Expo*, 2001, pp. 12.

[13] M.-F. Plassard. “Functional Requirements for Bibliographic Records - Final Report”, IFLA Study Group on the Functional Requirements for Bibliographic Records, K.G. Saur Verlag GmbH & Co. KG, München 1998, ISBN 3-598-11382-X.

[14] B. Brown, E. Geelhoed, and A. Sellen, “The Use of Conventional and New Music Media: Implications for Future Technologies”, In Hirose and M., editors, *Proceedings Interact'2001*, 67-75.

[15] O. Celma, S. Dasiopoulou, M. Hausenblas, S. Little, C. Tsinaraki, R.Troncy, “MPEG-7 and the Semantic Web”, W3C Incubator Grp. Editor's Draft 08/14/07, http://www.w3.org/2005/Incubator/mmsem/XGR-mpeg7/

[16] F. Pachet. “Knowledge Management and Musical Metadata”, 2005, *Encyclopedia of Knowledge Management*, Schwartz, D. Ed. Idea Group, 9 pages.

[17] O. Celma. “Music Recommendation: a multi-faceted approach”, Doctoral Pre-Thesis Work, 2006, Universitat Pompeu Fabra, Barcelona, 139 pages.

[18] L. Cantara. “METS: The Metadata Encoding and Transmission Standard”, 2005, *Cataloging & Classification Quarterly*, 40 (3-4), 237-253.

[19] R. Iannella. “Open Digital Rights Language”, v1.1, 2002, http://www.odrl.net/1.1/ODRL-11.pdf

[20] B. Furrie. “Understanding MARC Bibliographic: Machine-readable Cataloging”, The Library of Congres, Follet Software Co., 1988.

[21] R. Gartner. “MODS: Metadata Object Description Schema”, Pearson New Media Librarian, Oxford University Library Services, October 2003.