

Music Warehouses: Challenges for the Next Generation of Music Search Engines

François Deliège and Torben Bach Pedersen

Aalborg University

Abstract. Music Information Retrieval has received increasing attention from both the industrial and the research communities in recent years. Many audio extraction techniques providing content-based music information have been developed, sparking the need for intelligent storage and retrieval facilities. This paper proposes to satisfy this need by extending technology from business-oriented *data warehouses* to so-called *music warehouses* that integrate a large variety of music-related information, including both low-level features and high-level musical information. Music warehouses thus help to close the “semantic gap” by supporting integrated querying of these two kinds of music data. This paper presents a number of new challenges for the database community that must be taken up to meet the particular demands of music warehouses.

1 Introduction

The tremendous growth of digital music available on the Internet has created a high demand for applications able to organize and search in large music databases. Thanks to new digital music formats, the size of personal music collections often reaches up to thousands of songs and large online music stores and online radios are becoming very popular. However, current search tools still remain very limited: popular search engines only provide searches based on external annotations, but to offer truly natural and intuitive information retrieval, search and query into the primary media is required. These needs have given further impulse to the development of new methods for music information retrieval and research on digital music databases.

Companies have always spent a considerable amount of money and efforts to ensure proper storage and management of their business information in order to answer questions about sales, production, or any operation relevant to their particular business concerns. Therefore, in large companies, each operational unit has always gathered, on a regular basis, various pieces of knowledge using a number of systems. Unfortunately, these systems have usually been provided by different vendors over a long period of time and are based on different technologies and terminologies which often make integration a major problem. This integration is, however, needed when it comes to answering questions implying data from different operational units. For example, in order to determine the profitability of a given product, data from sales and production needs to be combined. Another example is trend analysis that requires combining the budget and the performance information over time. To solve this centralization problem, the *data*

warehousing approach integrates data coming from the various operational units into one common data store, referred to as the *data warehouse* (DW), optimized for data analysis purposes.

The data warehousing approach has already demonstrated its strengths in the business context and has been widely used as a solid ground for *On-Line Analytic Processing* (OLAP) systems. OLAP systems allow queries such as calculating the profitability of products categories over the years to be answered “live”. At the same time, such systems, regardless of the database management system used, have commonly adopted the same conceptual *multidimensional view of data*.

In other contexts, however, applications call for more *complex data structures* than the ones proposed in the classical *multidimensional data model*. One such domain is *music classification and retrieval*. Automated classification of song descriptors, computer or manually generated as in the Music Genome Project (<http://www.pandora.com/mgp.shtml>), has already received a lot of attention from the signal processing and machine learning research communities as well as from private companies. Also, the database community has shown an increasing interest in creating new indexes able to search among large amount of complex data such as music content descriptors. However, to the best of the authors’ knowledge, no work has been reported so far concerning the management of musical information using multidimensional models.

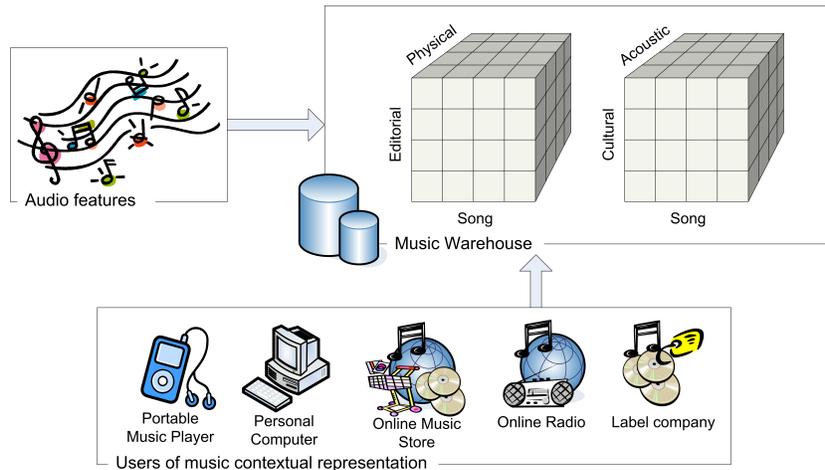


Fig. 1. The key role of an MW

Music Warehouses (MWs) are dedicated DWs optimized for the storage and analysis of music content. They provide the advanced framework necessary to support semantic tools able to close the gap between audio features and contextual representations. In particular, the multidimensional view of data commonly adopted in DWs facilitates the understanding of the mapping between low-level features and high-level representations. Also, the summarization features integrated in multidimensional models present

a prominent advantage. As pictured in Figure 1, MWs will play the key role of centralizing and integrating all music information pieces together. In order to capture the context of a song, MWs will use an advanced data model and its query language. Thanks to specifically designed query optimizations fast responses time will be ensured. The unequaled amount of music information available through MWs will be accessible to a large variety of clients, from personal music players to large music label companies.

The main focus of this paper is to identify and describe the various new challenges to multidimensional database models in the music classification field. The music world requires more powerful data model constructs than the ones offered by traditional multidimensional modeling approaches. However, the issues discussed here are not confined to the music domain but will find applications in other contexts.

2 Related Work in Databases

In [4], Downie et al. describe a secure and collaborative framework for evaluating music information retrieval algorithms. Little attention has been paid so far to the storage issues of audio features in DWs. A more traditional approach is to use classical relational models such as the one proposed by Rubenstein that extends the entity-relationship data model to implement the notion of hierarchical ordering, commonly found in musical data [14]. Through some examples, Rubenstein illustrates how to represent musical notation in a database using the extensions he introduces, but no detailed data types and operations are given. A multimedia data model, following the layered model paradigm that consists of a data definition layer, a data manipulation layer, a data presentation layer, and a control layer, is presented in [17], but no query language is proposed. Finally, a music data model with its algebra and query language is presented in [16]. The data model is able to structure both the musical content and the metadata but does not address performance optimization issues. However, none of these models adopt a multidimensional approach by representing data in cubes, a very convenient structure for performing on-the-fly analysis on large volume of data that has already proved its strengths in DWs [12].

DWs first appeared in the business context to integrate data from different operational units together and provide complete understanding and better coverage of the business matters. Driven by the market, academic research quickly showed interest in the topic. A major focus from both worlds has always been to support OLAP functionalities together with good performance. Research was performed on both conceptual and physical levels and has led to the creation of many different multidimensional models and OLAP systems. Multidimensional models can be divided into 3 categories: *simple cube* models, *structured cube* models and *statistical object* models [13]. OLAP systems have mainly been implemented using two technologies: *Relational OLAP* (ROLAP), based on a *Relational DataBase Management Systems* (RDBMS) [8], and *Multidimensional OLAP* (MOLAP), based on a dedicated *Multidimensional DataBase Management Systems* (MDBMS) [15]. Proper identification of the requirements of music classification systems is a first step to determine which conceptual and physical data warehouse elements are the best suited to take up the challenges offered by MWs.

3 Musical Classification

3.1 Musical Metadata

The music industry needs musical classification. While various classifications exist, no real consensus seems to have emerged. Music retailers, music labels, copyright companies, radio stations, end users, etc., have all designed their own taxonomies. Music retailers taxonomies, for example, that are aimed at guiding consumers in shops, are made up of four levels alphabetically ordered: global musical categories, subcategories, artist names, and album names. Even among the same group of interest, e.g., online music portals, inconsistencies are easy to find. One notable source of inconsistencies is the use of different kinds of *metadata*.

Metadata is commonly used in the research field of audio mining covering areas such as audio classification and retrieval. Literally “data about data”, metadata is defined as information about another set of data. In his work on musical knowledge management [9], Pachet classifies musical metadata into three categories depending on the nature of the source from where the information can be extracted.

In the audio context, metadata elements such as the title, the composer, the performer, the creation date and the publisher of a song are the most commonly used. They are referred to as *editorial metadata* and give authoritative information provided mostly manually by experts. Editorial metadata covers a wide range of information, from administrative to historical facts.

Cultural metadata is defined as knowledge produced by the environment or culture resulting from an analysis of emerging patterns, categories or associations from external sources of documents. Typical methods for generating such information are to use radio station play-lists to find correlations between songs or to crawl music web sites to gather word associations. An example of cultural metadata is the list of the most common terms associated with a given artist. Many online music stores, e.g., Amazon.com, are using cultural metadata based on user recommendations, a well-known collaborative filtering technique.

Acoustic metadata is the third category of music information. Acoustic metadata is defined as purely objective information obtained solely through an analysis of the audio content of the music. However, acoustic metadata remains dependent on the internal primary support of the musical information. While numerous approaches exist on what acoustic features to retain and how to select these features, they can primarily be separated into two classes: symbolic representation (MIDI) and acoustic representation (WAV, MP3). In the symbolic representation, the features usually refer to pitch, rhythm, or their variations, while in the acoustic representation the most common features are produced by time analysis, spectral analysis and wavelet analysis.

Physical metadata, a new fourth category of metadata, is defined as information directly related to the medium holding the music. Contrarily to cultural metadata, physical metadata is not produced by external elements such as the culture but rather provides information on the physical storage characteristics and its related practical constraints. A naive example would be the location of a music file on a computer, a possibly helpful piece of knowledge about the user’s classification. Physical metadata includes information such as the type of medium, e.g., a CD or a vinyl record, the kind of media, e.g.,

a music or a video clip, the format of the source, e.g., PCM or MP3, the compression used, e.g., lossless or lossy compression, etc. Physical metadata contains a lot of useful information in the context of an online music store where, for example, customers, depending on the speed of their internet connection, might want to buy video clips, music with high sound quality or music with lower sound quality.

Together, the four categories of metadata reflect the song context that can only be captured by a large quantity of metadata, possibly of high dimensionality and using heterogeneous units. Along with musical descriptions, methods to uniquely identify pieces of music are needed. Various robust audio fingerprint techniques have been developed to allow audio identification of distorted sources. Such techniques have already been successfully implemented in some systems such as the Moodlogic Music Browser (<http://www.moodlogic.com/>).

3.2 A Case Study of Musical Management

The case study illustrates the special demands of MWs. An ER diagram of the case is shown in Figure 2 using the notation of [5]. It pictures at a conceptual level the data model and is, therefore, not represented using a star-schema.

The *song* is the most important entity type, as indicated by the placement in the center of the diagram. A song is uniquely defined with a song identifier (SID) and has additional attributes such as Title, Length, Format, all of which are considered to be static. Audio fingerprints allow the song to be identified uniquely based on its audio content and independently of its storage format. A song has many relationships with other entities, whose purposes are to describe the song. These other entities might be viewed as *dimensions* and are shared by all songs.

First, a song can be characterized by its editorial information. Each song is authored by one or more composers and can be played by one or more performers. Both composers and performers are artists and are identified using their scene name along with some biographic elements. Performers usually form bands together. Each band is identified with a name and has at least one time interval in which it existed. Performers may join or leave the band without the band being dissolved. Bands are able to dissolve and reunite multiple times. A song is published by a music editor at a given time, either in a single or in album identified by a name, using distribution channels such as web radios, music television channels, online music stores, etc.

Second, using collaborative filtering and user profiles, the cultural context surrounding a song can be depicted. Co-occurrence analysis is performed by tracking user playlists and by crawling the web [3]. Each time a song is played, the previously played song is stored, so that the list of a user's most frequently played songs after a given one can be inferred. A user is uniquely identified using a user identifier (UID). Each user has a location, a date of birth, a language attribute, and possibly a music profile. The music profile stores which descriptors a user values the most. A music profile is defined on a user basis and is composed of a list of weights corresponding to the music descriptors. Based on the user music profiles, groups of similar profiles can be formed. Music profile groups link users that seem to have the "same ear", i.e., using the same criteria. Similar musical profiles, i.e., users identified as having the same musical tastes, can be grouped together into *musical audiences*.

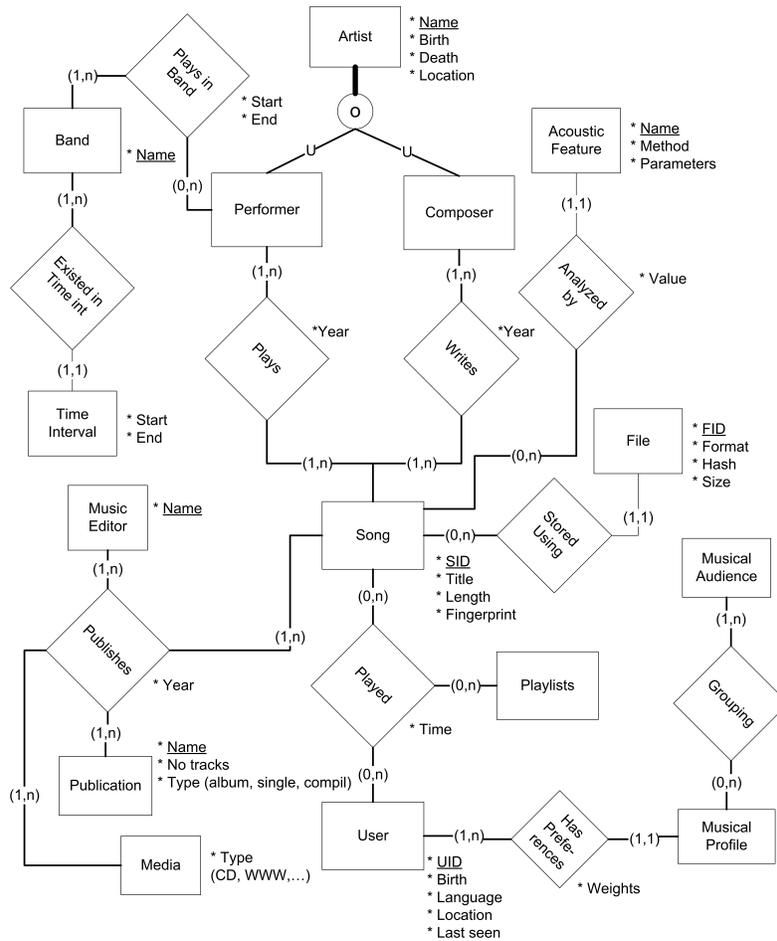


Fig. 2. Case study of an MW

Third, a song is described with acoustic information. For each song, acoustic features can be extracted using an extraction method and its parameters. Each acoustic feature is characterized by a unique pair of an extraction method and its parameters.

Finally, each song is stored using at least one physical medium, e.g., a file where the sound data has previously been encoded in one of the well known encoding formats such as MP3, WMA, or OGG. Each file is given a unique file identifier (FID), and is characterized by an identification tag such as a hash-key that permits to search if a file is already present, an audio format representing the encoding format of the sound, its size, its quality, etc.

Figure 3 presents the music features at a higher abstraction level. The various features capturing the song context are all considered as descriptors regardless of the category of musical metadata they belong. Descriptors are represented in the center of the

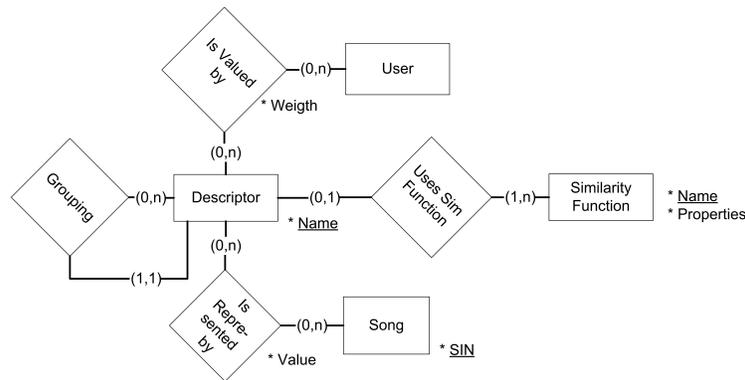


Fig. 3. High level features in an MW

figure as they are the most important entity type. Each descriptor is identified uniquely by its name. Multiple descriptors can be grouped together to form *fused* descriptors. Each song should be represented using as many descriptors as possible. Each descriptor has a weight reflecting its importance to each user. Finally, each descriptor should have at least one similarity function attached to it. Similarity functions allow comparison between values of a given descriptor for different songs. Once the descriptor similarities between two songs have been calculated, they can be computed into a general similarity value using the user weights.

4 Challenges for MW

One of the most prominent demands for MWs is the creation of a data model supporting more complex modeling constructs than classical multidimensional models, while keeping their strengths for decision support, i.e., including the full generality of ER models would be a turn back. The data model should provide integrated semantic support for the demands that follow.

Time series:

Many acoustic descriptors, such as the beat or the pitch, can be represented as multidimensional vectors at successive time points. Unlike typical DW facts, these types of data clearly yield no meaning when summed up. Other standard aggregation operators such as MIN, MAX and AVG do apply, but real demands are for more complex operations, such as standard deviation and other statistical functions that are useful for the similarity functions that underlay music queries. The data model should include operators allowing to cut, add, and compare time series along with aggregation operators enabling modifications of the sampling frequency of the time series. Furthermore, the model should support irregular time series in which samples are separated by non-uniform time intervals. Finally, it should be possible to use the above-mentioned advanced temporal concepts wherever meaningful.

Standards compatibility:

Many different formats are used to store music. While acoustic formats, e.g., MP3,

OGG, WAV, contain information about the audio wave transmitted, symbolic formats, e.g., MusicXML, Humdrum, Guido, represent high level encoding information such as the duration and the intensity of the notes. The current trend for representing audio content description is to use the symbolic MPEG-7 standard in XML format [7]. The MW should be able to integrate a number of different standards such as MPEG-7 and capture data into its multidimensional model.

Data imperfections:

In addition to the editorial, acoustic and cultural metadata, physical metadata, such as sampling frequency and format, could also be integrated in an MW to provide knowledge about the source quality. For example, a statistical measure of correctness could be applied to the title of songs with regards to where the information comes from, e.g., an original CD, a peer-to-peer sharing network, or simply missing information. Furthermore, given the large variety of music formats that support audio content, all automated extraction methods may not always be applicable or may apply with various degrees of precision, creating imperfections into the MW descriptors. Together, physical information and knowledge of imperfections should enable quality-of-service in MWs.

Precision-aware retrieval:

Precision-aware retrieval at the query processing stage is another aspect of quality-of-service in MWs. Indeed, certain queries performed in an MW do not require exact answers. Rather rough approximations would be sufficient. For example, nearest neighbors queries, such as the ranking of the k nearest neighbors of a given song, do not focus on the exact position of each song compared to a given one, but rather on coarser notion of distance, such as very close, close, or far. The exact granularity of the answer should not be fixed but rather determined either implicitly by an appropriate algebra, or explicitly in the query. Queries including the notion of ranking, referred to as Top-K queries, are very frequent in DWs. At the query processing level, optimizations can be performed in order to drastically improve the response time. Operators such as ranked selection and ranked joins use specific algorithms that have already demonstrated their usefulness for relational models. In MWs, however, Top-K queries require ranking at a coarse level of granularity where elements need to be ordered only in subsets, e.g., very close, close and far. A third aspect of quality-of-service in MWs is the response time. Time consuming queries, such as music comparison and nearest neighbor, spark the need for new techniques able to trade fast response time for precision. Query answers need to take the form of streams, updated progressively with more precise and reliable information. For example, asking what the common characteristics of a set of songs are could result in the immediate creation of a stream of music characteristics in their high-level representation, starting with coarse similarities and progressively refining similarities as the query processing continues.

Many-to-many relationships:

In traditional multidimensional models, facts are linked to the base elements of the dimensions using one-to-many relationships. Three classic alternatives exist to encode many-to-many relationships using multidimensional modeling: traditional dimensions, mini-dimensions, and snowflaking. Using traditional dimensions, all the possible combinations of artists are created. Since the number of combinations grows at an exponential rate when adding artists, this solution quickly becomes infeasible. Limiting the

enumeration to only the combinations actually used still leads to a large number of dimension records. Using mini-dimensions with one dimension for each possible artist will lead to a large number of dimensions, causing performance problems. Finally, snowflaking offers no advantage over traditional dimension as the number of basic elements would remain equal. Classical multidimensional models are able to capture the fact that an artist can perform many different songs but not the fact that multiple artists can perform together in a single song. Counting how many titles were performed by either artist A or B, becomes a dreadful task if we consider that songs performed by both artists should only be counted once. Instead, the intended behavior should be directly captured by the schema.

Versioned irregular hierarchies:

An essential step when approaching the music classification field is to understand the many issues related to how culture and sub-groups define musical categories, construct taxonomies and form interrelationships between categories. These issues have been discussed in the work of Fabbri [6], Brackett [2], Pachet and Cazaly [10], Aucouturier and Pachet [1], just to mention a few. From a data warehouse point of view, the taxonomies presented shared common properties. In a multidimensional database, a dimension hierarchy is said to be: *strict*, if all dimension values have no more than one direct parent, *onto*, if the hierarchy is balanced, and *covering*, if no containment path skips a level [11]. It is clear that, e.g., in a genre dimension, the hierarchy would be non-strict, non-onto and non-covering. However, this is not sufficient. Since very little consensus exists between taxonomies, the techniques already existing for slowly changing dimensions in multidimensional databases may not be appropriate. Instead, MWs require support for versioning abilities, mimicking software versioning systems such as CVS (<http://www.nongnu.org/cvs/>) or Subversion (<http://subversion.tigris.org/>), where different hierarchies could coexist and evolve. A versioned genre hierarchy that, for example, defines a classification of the genre dimension for different user profiles, will create a need for new database operators enabling comparison of the hierarchies between users and their evolution over time.

Fuzzy hierarchies:

Non-strict hierarchies, i.e., hierarchies supporting elements having multiple parents, allow different paths to be followed when performing roll-up operations. In the classical example of the time dimension, days can be rolled-up into months and in turn into years. Similarly, days can be rolled-up into weeks by following a different path since there are overlaps between week-month and week-year precision levels. While non-strict hierarchies are useful, they are very artificial in this precise case where a linear path composed of overlaps, day-week-months-years, may seem more intuitive. Fuzzy hierarchies enable children to belong to multiple parents with various degrees of affiliation, e.g. week 5 of year 2006 has a degree of affiliation of $2/7$ to January and $5/7$ to February. While fuzzy hierarchies are not required to handle typical DWs demands, they become unavoidable for MWs in order to represent complex hierarchies such as the genres. Sub-genres would belong to genres to a certain degree, e.g., the genre Jazz-Rock could belong to 60% to Jazz and 40% to Rock, a notion that multidimensional data models have not been able to fully capture so far.

Navigation in n-dimensional space:

The mental representation of songs as objects in an n-dimensional space is not new in the field of music classification. Far from being purely a dream, projects such as MusicMiner (<http://musicminer.sourceforge.net/>) already offer a two-dimensional mapping of personal music collections. It is therefore very tempting to enrich the MW data model with multidimensional navigation features such as notions of neighborhood, intersections, landscape, fuzzy borders, etc. In such a space, a play-list can be seen as a journey from one song to another. Automatic play-list generation could be as like car navigation systems able to recommend some itineraries with notions of primary and secondary roads to reflect the musical tastes of the user.

Aggregates for dimensional reduction:

A very challenging aspect of MWs is the high number of dimensions used, songs can be described using several hundred dimensions, hence, urging the need for efficient ways to aggregate this massive amount of information in useful ways. The traditional multidimensional approach is to reduce dimensionality by using projection, i.e., throwing out dimensions by omitting pieces of available information. Instead, by using *fused dimensions*, many dimensions, such as the rhythm, the notes, and the loudness could be summarized into a more general one, reflecting the overall melody of the songs. Using aggregates for dimensional reduction clearly offers many advantages as the complexity of the data is reduced, while the essence is maintained. The MW should provide efficient techniques to reduce or increase the number of dimensions.

Integration of new data types:

The musical world does not only deal with audio but also embraces a lot of external multimedia content. New bands often aim to increase their audience by creating interactive web sites, video clips, attractive CD covers, etc. MWs should be able to deal with such information, as not including these non-audio additions is neglecting an increasingly important part of the musical experience of the audience. Interactive web sites, for example, often make extensive use of embedded application such as Flash. These applications offer biographies, an agenda of next concerts, rumors and forums to users. MWs should provide users such pieces of information. It should be possible to define specific extractors for the applications and to perform analysis on the extracted features. MWs should be able to handle queries requiring partial integration of the applications, e.g., obtaining the list of Madonna's next concerts.

5 Conclusions

Inspired by the previous successes of DWs in business integration issues and on-the-fly analytical demands, this paper proposes the development of MWs which are centralized data stores based on the data warehousing approach and optimized to answer the fast-search needs of future large music information retrieval systems.

Previous work on musical classification has shown multiple sources of inconsistencies between ontologies. One source of these inconsistencies is the use of different musical facets when describing a song. These facets can be described using musical metadata. Four high-level categories of metadata are identified and briefly described:

editorial, cultural, acoustic and physical metadata. Using these four categories, a case study of musical database management is presented.

Ten exciting challenges offered by MWs for the existing DWs are identified. While these challenges originate from the requirements music classification systems, they are, however, not confined to this area. In particular, data imperfections, precision-aware retrieval using coarse Top-K queries or streams, versioned irregular hierarchies and fuzzy hierarchies are new and relevant to the general database research community.

Work on these challenges will be pursued in order to support the successful integration of DWs in the musical world. In particular, research on fuzzy hierarchies is currently being conducted in order to be able to classify a song in multiple musical genres. At a later stage, the responses to these challenges will find applications in other contexts.

References

1. J.-J. Aucouturier and F. Pachet. Representing musical genre: A state of the art. *Journal of New Music Research*, 32(1):83–93, 2003.
2. D. Brackett. *Interpreting popular music*. Cambridge University Press, 1995.
3. W. W. Cohen and W. Fan. Web-collaborative filtering: recommending music by crawling the Web. *Computer Networks*, 33(1–6):685–698, 2000.
4. J. S. Downie, J. Futrelle, and D. K. Tcheng. The international music information retrieval systems evaluation laboratory: Governance, access and security. In *Proc. of ISMIR*, pages 9–15, 2004.
5. R. Elmasri and S. B. Navathe. *Fundamentals of Database Systems, 4th Edition*. Addison-Wesley Longman Publishing Co., 2003.
6. F. Fabbri. Browsing music spaces: Categories and the musical mind. <http://www.mediamusicstudies.net/tagg/xpdfs/ffabbri9907us.pdf>, 1999.
7. MPEG Requirements Group. MPEG-7 overview. *ISO/IEC JTC1/SC29/WG11N6828*, 2004.
8. R. Kimball, L. Reeves, M. Ross, and W. Thornthwaite. *The Data Warehouse Lifecycle Toolkit*. Wiley, 1996.
9. F. Pachet. Musical metadata and knowledge management. In *Encyclopedia of Knowledge Management*, pages 672–677. Idea Group, 2005.
10. F. Pachet and D. Cazaly. A taxonomy of musical genre. In *Proc. of Content-Based Multimedia Information Access Conference*, volume 2, pages 1238–1245, 2000.
11. T. B. Pedersen and C. S. Jensen. Multidimensional data modeling for complex data. In *Proc. of ICDE*, pages 336–345, 1999.
12. T. B. Pedersen and C. S. Jensen. Multidimensional database technology. *IEEE Computer*, 34(12):40–46, 2001.
13. M. Rafanelli. *Multidimensional Databases: Problems and Solutions*. Idea Group, 2003.
14. W. B. Rubenstein. A database design for musical information. In *Proc. of ACM SIGMOD*, pages 479–490, 1987.
15. E. Thomsen. *OLAP Solution: Building Multidimensional Information Systems*. Wiley, 1997.
16. C. Wang, J. Li, and S. Shi. A music data model and its application. In *Proc. of the International Multimedia Modelling Conference*, pages 79–85, 2004.
17. M. J. Wynblatt and G. A. Schloss. Control layer primitives for the layered multimedia data model. In *Proc. of the ACM International Conference on Multimedia*, pages 167–177, 1995.