This article was downloaded by: [71.225.4.231]
On: 11 July 2011, At: 11:55
Publisher: Routledge
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK

Journal of Library Metadata
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/wjlm20

Music Metadata in a New Key: Metadata and Annotation for Music in a Digital World

James Waters [a] & Robert B. Allen [b]

[a] Department of Information Science and Technology, Cabrini College, Radnor, Pennsylvania, USA

[b] College of Information Science and Technology, Drexel University, Philadelphia, Pennsylvania, USA

Available online: 20 Nov 2010

To cite this article: James Waters & Robert B. Allen (2010): Music Metadata in a New Key: Metadata and Annotation for Music in a Digital World, Journal of Library Metadata, 10:4, 238-256

To link to this article: http://dx.doi.org/10.1080/19386389.2010.524863

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.tandfonline.com/page/terms-and-conditions

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan, sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Journal of Library Metadata, 10:238–256, 2010
Copyright © Taylor & Francis Group, LLC
ISSN: 1938-6389 print / 1937-5034 online
DOI: 10.1080/19386389.2010.524863

Music Metadata in a New Key: Metadata and Annotation for Music in a Digital World

JAMES WATERS
Department of Information Science and Technology, Cabrini College, Radnor, Pennsylvania, USA

ROBERT B. ALLEN
College of Information Science and Technology, Drexel University, Philadelphia, Pennsylvania, USA

Downloaded by [71.225.4.231] at 11:55 11 July 2011

Not only are music recordings now primarily digital and able to be readily produced, edited, and mashed up, but entirely new metadata frameworks, encoding systems, and data models such as XML, FRBR, METS, and MPEG are being developed to cope with the needs of diverse user communities. Moreover, systematic annotations of the structure and details of musical pieces are increasingly common. This paper surveys the current state of knowledge with regard to annotation and metadata related to recorded music.

KEYWORDS *FRBR, METS, MPEG, music, semantic annotation, XML*

INTRODUCTION

Digital technology is changing the way we create, organize, and interact with music. Music in a digital format can be creatively manipulated. Thus, music can be filtered to remove vocal content, allowing for the creation of karaoke track. Karaoke music typically blends music and graphics (text and music alignment data, such as a bouncing ball) using MP3+G or CD+G files. When the graphic component is played back it indicates where specific words are to be sung. Mashups are ad hoc combinations of other works. For instance, the instrumental part of one music track could be overlaid with the vocal track of another. Moreover, digital music can be easily downloaded

Address correspondence to James Waters, Department of Information Science and Technology, Cabrini College, 610 King of Prussia Road, Radnor, PA 19087-3698, USA. E-mail: james.waters@cabrini.edu

Downloaded by [71.225.4.231] at 11:55 11 July 2011

over the Internet, thus music can be legally purchased instantly and without the need for a physical container. Metadata for such downloads includes some of the low-level attributes described in the following section, plus DRM information, creative information, and descriptive information.There are many different attributes for complex musical objects. However, these attributes are often grouped as Descriptive, Structural and Administrative (NISO, 2004). Descriptive metadata describes the work at a conceptual level. Structural metadata defines what components comprise a complex object, for instance a symphony is composed of movements (normally three to five), a movement is made up of bars, and bars are composed of a sequence of notes (Minibayeva & Dunn, 2002).

Administrative metadata incorporates technical, preservation, and rights metadata. For instance, a digital music file might have technical metadata about its size in bytes, the format type (e.g., MP3, CDDA, WAV, FLAC, AAC, ALAC), the compression type (lossy or lossless), and so on. Rights metadata describes who can do what with the object. For instance I can listen to a CD but I am not allowed to make a copy and sell it to a third party. Preservation metadata might describe the hardware and software used to capture the object, the settings used such as resolution, bit-depth, bit-rate, sampling rate, compression, file format, how to get it back (retrieval and playback), and the history of any changes to the digitized capture.

Components of music such as notes and even melodies (Pavia, Mednes, & Cardoso, 2004) can be determined automatically by signal processing. Nonetheless, there may still be a "semantic gap" between the levels of descriptors that can be used to characterize music at the levels of signal (elementary), content (structural), and human (cognition and affect) experience (Celma, Herrera, & Serra, 2006).

FRBR

The Functional Requirements for Bibliographic Records (Vellucci, 2007) organizes information resources in terms of Work, Expression, Manifestation, and Item. This model attempts to capture and define the boundaries between conceptual, logical, instantiated, and physical representations (Figure 1). While FRBR has proven generally very useful, it needs some reinterpretation to be applied to music (Riley, Mullin, Colvard, & Berry, 2008).

According to FRBR, a Work is an abstract creation, and Descriptive Metadata applies at that level. Somewhat controversially, a given performance may be considered an Expression. Then, recordings of that performance are Manifestations and reproductions, such as CDs, of it are Items.

Musical Objects and Musical Information Packages

There are many notations for representing the music itself. While the Western world is most familiar with a 5-line staff divided into bars, there a many

Element	Description	Example
Work	Intellectual or Artistic creation	*Mahler 2nd Symphony*
Expression	Performance	Mahler *2nd Symphony, Vienna Philharmonic, Conductor: Zubin Mehta, 1975*
Manifestation	Physical embodiment	*The recording of Mahler 2nd Symphony Vienna Philharmonic under Zubin Mehta, 1975*
Item	An exemplar of the manifestation	*Mahler 2nd Symphony, Vienna Philharmonic under Zubin Mehta, 1975, Decca Legends, CD*

FIGURE 1 Description of a symphony at several FRBR levels.

Downloaded by [71.225.4.231] at 11:55 11 July 2011

computer notation schemes. Selfridge-Field (1997) lists more than 30 different computer-based notation schemes, many are variations of the Musical Instrument Digital Interface (MIDI), which was originally developed as a note-based representation for music synthesizers. Notations such as MusicXML (Good, 2001) are readily captured in data files (Figure 2), but those need to be distinguished from the files that are digitizations of performances. A Musical Information Package is composed of the musical object and the metadata associated with it.

```
<?xml version=”1.0” encoding=”UTF-8” standalone=”no”?>
<!DOCTYPE score=partwise PUBLIC “-//Recordare//DTD MusicXML 0.5 partwise//EN”
“http://www.musicxml.org/dtds/partwise.dtd”>
  <score-partwise>
    <part-list>
     <score-part id=”P1”> <part-name>Music</part-name> </score-part>
    </part-list>
    <part id=”P1”>
     <measure number=”1”>
      <attributes>
       <divisions>1</divisions>
       <key> <fifths>0</fifths> </key>
       <time> <beats>4</beats> <beat-type>4</beat-type> </time>
       <clef> <sign>G</sign> <line>2</line> </clef>
      </attributes>
      <note>
       <pitch> <step>C</step> <octave>4</octave> </pitch>
       <duration>4</duration>
       <type>whole</type>
      </note>
     </measure>
    </part>
  </score-partwise>
```

FIGURE 2 One note (Middle C) in a composition as expressed in MusicXML (Good, 2001).

Title:	Symphonie fantastique
Creator:	Berlioz
Subject:	LCSH: Music
Description:	Berlioz cast personal experiences in the form of a Beethoven symphony.
Publisher:	EMI/Angel
Contributor:	Roger Norrington, London Classical Players
Date:	1989
Type:	Sound
Format:	CD
Identifier:	CDC 7495412
Language:	EN
Rights:	EMI

FIGURE 3 Example of the basic Dublin Core elements applied to a work of music.

BASIC METADATA FRAMEWORKS

Dublin Core

Dublin Core was developed as a simplified metadata system, the simplicity of which facilitates its application to many types of content. Metadata frameworks such as Dublin Core may be associated with XML descriptions such as MusicXML via an extension of XML known as the Resource Description Framework (RDF). Figure 3 shows basic Dublin Core elements applied to recorded music.

Dublin Core was developed as a flexible, high-level framework, but this generality makes its elements ambiguous. More detail can be added when the elements are qualified. For instance, the Date might be the date created (recorded) or the date issued (published). An alternative to qualifying (i.e., subdividing) the elements is to extend the metadata system. For instance, the Virtual Music Library (Andresen, 2001) proposes extensions to Dublin Core to include Composer, Lyricist, and Conductor to facilitate access to music Web sites, special collections, sound tracks, and printed music.

Music with lyrics presents particular problems both for modeling schemes such as FRBR and encoding schemes like Dublin Core. What is the primary work, is it the music, the text, or both together (Ayres, 2003)? Several proposals have extended Dublin Core to deal with music; the MusicBrainz project (Swartz, 2002) is defining a metadata standard for music recordings based on Dublin Core and with the long-term aim of defining a metadata standard for music—a metadata catalog of all music recordings. At present the total runs about a million items. Attempts to create a workable musical thesaurus have proven difficult (Hemmasi, 1992, 1994, 1995).

MODS

The Metadata Object Description Scheme (MODS) was created by the Library of Congress and primarily focused on bibliographic elements; this

Downloaded by [71.225.4.231] at 11:55 11 July 2011

scheme is an evolution of MARC. MODS elements include titleInfo, name, typeOfResource, genre, originInfo, Language, physicalDescription, abstract, tableOfContents, targetAudience, note, subject, Classification, relatedItem, Identifier, location, accessCondition, part, extension, and recordInfo. Many elements also include sub-elements to allow a greater level of detail to be captured.

ADMINISTRATIVE METADATA

Metadata frameworks such as Dublin Core do not break out the details of Administrative and Structural Metadata. Here, we focus on the components of Administrative Metadata, which as mentioned earlier includes Technical, Preservation, and Rights Metadata.

Technical Metadata

Technical metadata fields describe the digital aspects of a digitized recording (Casey, 2002; Harvard, 2004). Of course, sound recording can be a highly complex process starting with the capture on sound mixing of a performance. That recording may be edited and remixed and, ultimately, sampled and compressed. Ultimately, we may think of metadata as capturing these processing steps, but for now we might simply record attributes such as:

- Engineer
- Date
- Hardware (Microphones, Computer, ADC, Mixer) Settings and Calibration
- Data recording devices
- Software used for digitization
- File format such as WAV, MPEG, FLAC, ALAC, AAC, WMA, OGG, etc.
- Compression
- Data format, i.e., Pulse Code Modulated (PCM), Direct Stream Digital (DSD)
- Codec (Compressor decompressor software, i.e., LAME, Fraunhofer)
- Data rate (Kbits per second), (fixed or variable)
- Bit-depth (8,16,24,32, ...)
- Sampling rate (variable from 11 Khz to 192 Khz)
- Size in samples/frames/time/Bytes
- Checksum, Wordsize, Block size (Bytes), Byte-order (LSB at beginning or end)

Preservation Metadata

Digital preservation is an integral to digital curation. The most common set of preservation metadata, Preservation Metadata Implementation Standard (PREMIS, 2004), specified a set of core metadata elements and a supporting data dictionary and a set of strategies to identify and evaluate alternative

Downloaded by [71.225.4.231] at 11:55 11 July 2011

strategies for encoding, storing, and managing preservation metadata in digital preservation systems (PREMIS, 2008). A notable feature of PREMIS is the *digiProv* element, which specifies the digital provenance. Digital provenance is the history of copies and transformation from one format to another. PREMIS represents Objects, Events, Agents, and Rights. Essentially, this describes the sequence of format changes that were made to the files. Ultimately, the Preservation Metadata and Technical Metadata overlap need to be coordinated.

Rights Metadata

Rights data is increasingly important. It is now easy to "rip" (bit-perfect data extraction) audio from physical media such as CD and DVD and files may thus be quickly distributed worldwide, which may deprive copyright holders of revenue from sales. Thus, copyright holders are contesting even the right to make fair use personal copies. Rights metadata, when attached to or embedded in objects, can limit or prevent certain operations on that object by interacting with suitable equipped applications that enforce systems of rights management. For instance you may own a physical copy of a CD that cannot be played on a personal computer, or you may buy a digital file that cannot be moved from the device it is downloaded to. At the same time, consumers are sometimes uncomfortable with copyright owners' use of digital technologies to prevent copying of their works (Musick, 2004); indeed, the Digital Millennium Copyright Act (DMCA) makes it illegal to create algorithms to circumvent copy protection.

Downloaded by [71.225.4.231] at 11:55 11 July 2011

THE MPEG FAMILY OF STANDARDS

While the bibliographic community has developed standards and frameworks such as FRBR, MODS, and Dublin Core, the multimedia industry has been pursuing a different family of standards. The foremost example is the Motion Pictures Expert Group (MPEG). This was established in 1988 under ISO and is responsible for drafting and developing standards related to the encoding, rights management, and description of audio.

MPEG Encoding Standards

MPEG-1, MPEG-2, and MPEG-4 deal with encoding, while MPEG-7 deals with content description. MPEG-1 and MPEG-2 are generally familiar and we do not discuss them further except to note that the Compact Disk audio encoding is not related to MPEG but that the MP3 (MPEG-1 layer III) standard originated as an audio component of MPEG-2.

MPEG-4 is an evolving set of Audio-Visual standards. These are primarily concerned with compression. Transmitting uncompressed high-quality video requires a huge bandwidth and storing this video requires lots of disk space, but compression techniques can lower this overhead and yet allow a reconstructed signal to have minimal signal loss compared to the original. Similarly, modern portable AV devices have limited storage and the use of compressed files can enable the storage of many hours of high-quality media. Originally intended for such low capacity or low bandwidth devices, MPEG-4 has evolved via refinements such as H.264 to be capable of handling High Definition (HD and Blu-Ray) video streams. MPEG-4 includes a standardized DRM (Digital Rights Management) identification mechanism. MPEG-4 also includes SAOL (Structured Audio Orchestra Language) (Scheirer & Vercoe, 1999). SAOL is an MPEG-4 music-synthesis language model that includes a set of algorithms used to generate sound (Wang, Siao, & Su, 2006) in the form of high-quality low-bit-rate audio.

MPEG-7

MPEG-7 is a comprehensive framework that covers both metadata and annotation, as well as both structure and semantics of Audio-Visual content. The purpose is to allow "interoperable searching, indexing, filtering, and access of audio-visual content by enabling interoperability among devices and applications that deal with AV content description. MPEG-7 aims to achieve this by providing a set of basic descriptors, a descriptor definition language, a set of descriptor schemes, and a coding language. Of most interest here are the descriptor schemes that allow descriptions of structural and conceptual aspects of AV content. The structural descriptions allow the description of an object to be decomposed into spatial, temporal, and spatio-temporal segments. MPEG-7 focuses on a structural decomposition (Gómez, Klapuri, & Meudic, 2003) and proposes a focus on melodic structure and fundamental frequencies (Figure 4).

Downloaded by [71.225.4.231] at 11:55 11 July 2011

Silence D
TimbralTemporal
- Log AttackTime D
- TemporalCentroid D

BasicSpectral
- AudioSpectrumEnvelope D
- AudioSpectrumCentroid D
- AudioSpectrumSpread D
- AudioSpectrumFlatness D

Basic
- AudioWaveform D
- AudioPower D

TimbralSpectral
- HarmonicSpectralCentroid D
- HarmonicSpectralDeviation D
- HarmonicSpectralSpread D
- HarmonicSpectralVariation D
- SpectralCentroid D

SpectralBasis
- AudioSpectrumBasis D
- AudioSpectrumProjection D

SignalParameters
- AudioHarmonicity D
- AudioFundamentalFrequency D

FIGURE 4 MPEG-7 audio framework (JTC1/SC29/WG11, 2004).

The MPEG-7 Multimedia Content Description Interface developed ISO/IEC #15938 to describe the content of multimedia objects and facilitate machine-based searching. The Multimedia Description Schemes (MDS) uses an XML schema syntax with some MPEG-7 extensions (Salembier & Smith, 2001). The basic elements consist of data types, mathematical structures, linking and media localization tools, and elementary Descriptor Schemas. The next level, called the content management and content description, builds on the lowest level. It describes the content from several viewpoints: creation and production, media, usage, structural aspects, and conceptual aspects (Chang, Sikora, & Purl, 2001). The first three elements address primarily information related to the management of the content (content management), while the last two are devoted to the description of perceivable information (content description). In MPEG-7, a set of description schemes provides the complex data types needed for the semantic description of audiovisual content. The core of these schemata is delineation of an artifact into a series of segments within the class hierarchy. An AudioSegment is one type of multimedia segment.

The AudioSegment is merely the top level of a complex descriptor scheme that consists of objective information such as spectra, fundamental frequencies, and key and higher-level features (Pachet, 2003). MPEG-7 can describe music to intricate levels of detail including individual notes, frequencies, beats, scales, meter, contours, tempi, vibrato descriptions, and instrument timbre (Gómez, Gouyon, Herrera, & Amatriain, 2003). Figure 5 illustrates how the MPEG-7 AudioSegment can be used to describe melodic elements.

Downloaded by [71.225.4.231] at 11:55 11 July 2011

STRUCTURAL METADATA AND CONTENT WRAPPERS

Structural attributes of music may include the length of a piece. The internal structure of musical pieces such as the movements of symphonies can also be identified (Minibayeva & Dunn, 2002). As noted at the beginning of this paper, Structural Metadata describes the organization of different parts of a complex digital object. For instance, Web pages often include several different types of content.

For more complex musical digital objects, such as albums, several distinct objects are bundled together and content wrappers provide an overall framework for them. Specifically, MPEG-21 and METS are frameworks we discuss here for describing the relationship of the parts of complex musical objects.

MPEG-21

MPEG-21 describes the constitution of complex digital resources. While MPEG-7 focuses on multimedia content itself, MPEG-21 is centered on the

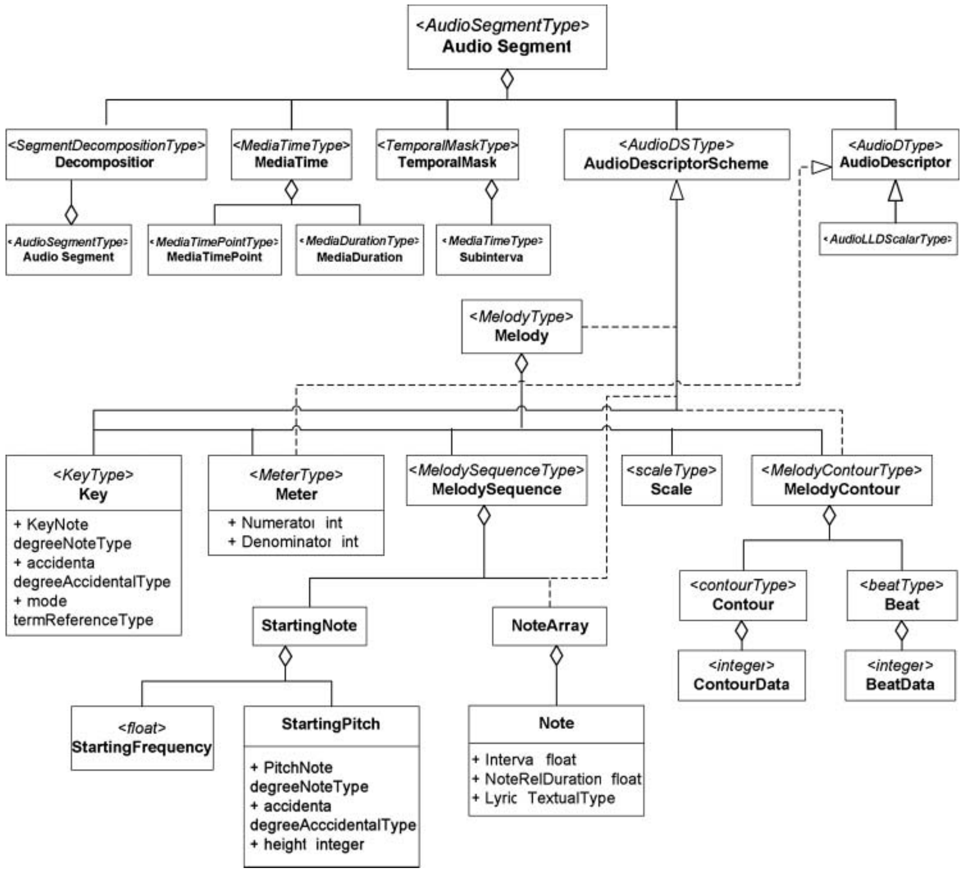

FIGURE 5 MPEG-7 Audio Segment elements (Gómez et al., 2003). This includes the Audio Description Schema, which has musical attributes Key, Meter, MelodySequence, Scale, and Melody Contour. (Reproduced with permission of the Audio Engineering Society).

transparent, augmented, and interoperable delivery and consumption of digital items across multiple platforms (Bormans, Gelissen, & Perkis, 2003; Wactlar & Christel, 2002). The heart of MPEG-21's content description is the hierarchical Digital Item Declaration (DID). A schematic of the DID for an album is shown in Figure 6. The album is composed of sub-Items with distinct units, such as the Cover and Tracks of Music. These sub-Items are composed of distinct Components, but those Components are stand-alone distinct items.

METS

The Multimedia Encoding and Transmission Standard is a content wrapper that was developed at the Library of Congress. It is composed of seven different elements (Figure 7). The StructMap describes the parts or divisions of the complex object. These parts are associated with metadata and are defined in the other sections. An example of the Descriptive Metadata, which

Downloaded by [71.225.4.231] at 11:55 11 July 2011

Item: Album
Descriptor:
Item: Cover
Description
Component: Front cover
Description
Resource
Component: Back Cover
Description
Resource
Item: Track1
Descriptor
Component Music
Description
Resource
Component Lyrics
Description
Resource
Item: Track2
Description
Component Music
Description
Resource
Component Lyrics
Description
Resource

FIGURE 6 A schematic of an MPEG-21 Digital Item Declaration of a complex item in this case for an album.

Downloaded by [71.225.4.231] at 11:55 11 July 2011

in this case is drawn from Dublin Core (Figure 8). The structure description of METS is more flexible than the hierarchical MPEG-21, but this flexibility adds complexity and makes it less intuitive. The Behavior section describes what repository services will be applied to the musical object.

SEMANTIC ANNOTATION

Types of Annotations

The term Semantic Annotation is not well defined and has different meaning to different users (Kiryakov, Popov, Terziev, Manov, & Ognyanoff, 2004). At the simplest level, annotations add content to an underlying informa-

Header
Descriptive metadata
Administrative metadata
File
StructMap
StructLink
Behavior

FIGURE 7 MPEG-21 Elements of a METS record.

```
<dmdSec ID = "dmd002">
  <mdWrap MIMETYPE = "text/xml" MDTYPE = "DC" LABEL = "Dublin Core Metadata">
    <dc:title>Alice's Adventures in Wonderland</dc:title>
    <dc:creator>Lewis Carroll</dc:creator>
    <dc:date>between 1872 and 1890</dc:date>
    <dc:publisher>McCloughlin Brothers</dc:publisher>
    <dc:type>text</dc:type>
  </mdWrap>
</dmdSec>
```

FIGURE 8 METS snippet showing descriptive metadata section (dmdSec). Specifically, this shows a metadata wrapper (mdWrap) for Dublin Core metadata.

tion resource. When RDF is used to add Dublin Core elements to an XML file—annotation is simply another name for metadata.

In many cases the annotations add conceptual or "semantic" knowledge about the content. Indeed, in many cases, such annotations felt to be addressing the Semantic Gap. This information may be contextual background such as the historical, cultural and political context. For instance, Why is Beethoven's third symphony called Eroica and the fifth piano concerto called Emperor? Is there a connection?[1] Why were composers reluctant to create a ninth symphony?[2]

Digital objects make it easy to link annotations to anchors within the digital objects. Indeed, the use of structural metadata to describe the internal structure of music can be thought of as a sort of annotation, but annotations can have many other uses. For instance, there might be a link to background information about a person mentioned in lyrics. Or, the parts of the Grand Canyon Suite could all be linked together as a sort of narrative.

Using "Ontologies" to Annotate Music

Because the term *Semantic Annotation* is often associated with the Semantic Web, the annotations are frequently said to be based on ontologies (Athanasiadis et al., 2005; Bertini, Cucchiara, del Bimbo, & Torniai, 2005; Papadopoulos, Mylonas, Mezaris, Avrithis, & Kompatsiaris, 2006). While perhaps this stretches the formal definition of ontology, it at least implies that there is a formal system behind the descriptive vocabulary.

The relationships between media and concepts are delineated (Alberink, Rutledge, Hardman, & Veenstra, 2004). If we are attaching meaning to musical pieces we can enumerate the size and constituency of the performing personnel (Singers, Instrumentalists, and so on). We can include instructions for playing a piece; this includes who plays which part and the manner in which parts are played (forte, pianissimo, allegro, etc.). Semantic annotation in music can also be used to describe a piece in terms of melody, emotion

Downloaded by [71.225.4.231] at 11:55 11 July 2011

and instrument (Turnbull, Barrington, Torres, & Lanckriet, 2008), rhythm or timbre (Duan, Lu, & Zhang, 2008).

Weyde and Wissmann (2007) focus on harmonic structures as a defining feature of music, and Duan, Lu, and Zhang (2008) use a vocabulary of 50 labels and 10 categories to describe music but allow music to exist is multiple entries within a single category, such as Vocal (male and female). In other cases, the ontologies describe attributes of the music. For instance, Ferrara, Ludovico, Montanelli, Castano, and Haus (2006) develop a system of the genres across many types of music. However, genre definitions are seldom consistent. Patchet and Cazaly (2000) describe a musical genre taxonomy that extends to 378 genres with 800 similarity relationships, while Rauber and Fruhwirth (2001) find that some music providers are limited to 350 genres. They describe a system that can analyze musical clips and create a self-organizing map showing similarity relationships. Gouyon, Dixon, Pampalk, and Widmer (2004) showed how tempi can be used to automatically distinguish between closely related musical genres.

Downloaded by [71.225.4.231] at 11:55 11 July 2011

REPOSITORIES AND DEVICES

End-User Personalization

With the advent of music downloads and compression techniques such as MP3 there has been a great deal of interest in online repositories of information for music that can be used to search for and tag files "ripped" from CD. Databases such as CDDB (Gracenote) and MusicBrainz (Swartz, 2002) have been built from information uploaded by owners of music CDs. This tends to be fairly basic information such as artist, album title, track title, track number, and (sometimes) genre. Tags for digitized music however can be more elaborate and, so, stand-alone tagging tools can now distinguish between Artist/Composer. However, music performances can involve a great number of creators and performers often difficult to represent formally by existing schemes, for instance, in Beethoven's ninth symphony the final movement includes music by Beethoven and words written both by Schiller and Beethoven and when performed needs an orchestra (with leader, normally the first violin), conductor, large choir, choir master, and soloists. Existing schemes require the packaging of much of this information in undifferentiated free text notes elements.

Tag schemes such as FLAC, ID3 (for MP3 files) and MPEG-4 tags can add a wealth of information such as copyright, date of purchase (for online music purchases), encoding data such as bit-rate, publisher, radio station information, series information, length of track, start point within an aggregation such as a CD or tape, modifications, and lyrics. As can be seen in Figure 9, ID3 tags also provide basic creative information about the audio file.

Downloaded by [71.225.4.231] at 11:55 11 July 2011

FIGURE 9 ID3 Header displayed in Winamp (Music player software) note tag versions.

The popular iTunes application allows the playback of digital audio files. The application incorporates the ID3 tagging system, as above; iTunes, however, also allows the user to add other information such as:

- Personal rating
- Beats per minute
- Category
- Sort Elements
- Last Played
- Equalizer (Digital processing applied on playback)

MPEG-A

The MPEG standards we discussed earlier concentrated on digital objects, their encoding, metadata, and interoperability. MPEG-A (the A stands for

	Component						
	JPEG		MPEG				
MAF	JPEG	JPEG2000	MPEG1	MPEG2	MPEG4	MPEG7	MPEG21
Photo Player	x		x		x	x	x
Music Player	x				x	x	
Other Player		x			x	x	x

FIGURE 10 MPEG-A Multimedia Application Formats (MAF) are wrappers for sets of metadata.

Application) by contrast is focused on creating standard Multimedia Application Formats (MAFs) for the applications which will be used to render or manipulate digital objects in devices such as portable digital music players (Diepold, Pereira, & Chang, 2005). An MAF specifies how a media application will embody related MPEG standards; each standard would be handled independently of other standards. Currently there are nine MAFS defined by the ISO. These include Music Players, Photo Viewers, and streaming applications that other MAFs are being defined for—MPEG is drafting MAF standards for portable video devices and interactive music players (Kim, Schreiner, & Diepold, 2008).

Downloaded by [71.225.4.231] at 11:55 11 July 2011

EMERGING CHALLENGES

Documenting the Context and Experience of Music

Metadata has traditionally emphasized the description of works, and the context of those works has generally been captured less formally. We have musical historical context, what went before, what follows, how a piece or composer was influenced or influences others. We have social and cultural historical contexts. Why was a piece outré, why does a certain piece reflect the zeitgeist, what political statement is being made in *Spartacus* or *Nabucco* or in the subtitle of Beethoven's third symphony? Was *The Magic Flute* really all about masons? Why were there riots at the premier of *The Rite of Spring* and Salome (amongst others). Evaluations of music are not fixed; composers and works can *vanish* for centuries before being rediscovered. To date, the only place to see this context is in liner notes or academic treatises. Other mechanisms may simply place composers in a timeline, possibly showing transitions or genetic abnormalities of musical evolution. Surely, we can embed this broader contextual data as part of a metadata/semantic annotation scheme.

Perhaps because of the services provided by the Internet, there is now often a collection of ancillary material for items and works that captures a rich set of contexts. Increasingly, this can be preserved and needs some formal means to connect it to the primary artifact. Context can be very important, for instance, why should we be surprised to hear *Veni Creator Spiritus* (or not) in Mahler's third symphony or *Frère Jacques* in Mahler's first. Whose works were referred to by Thomas Beacham as "I believe I stepped in some once"? It is impossible to re-create the exact experience of an audience listening to a work for the first time. There is a strong analog here with the preservation of other inherently transient cultural artifacts such as video games. Lowood (2004) describes how we will soon lose the ability to directly experience old video games as the original technology becomes obsolete or unavailable or impossible to repair. Similarly, we cannot experience the direct experience of the audience at the premier of *The Rite of Spring*, but we can collect "experiences" via historical documents; and in modern times we can capture interviews with music listeners, and performance reviews can be preserved indefinitely. Yet, we many want to capture that with systematic descriptions.

Downloaded by [71.225.4.231] at 11:55 11 July 2011

Automated Indexing and Retrieval

One of the main purposes of metadata is to support retrieval. However, the richness of music can complicate the general case. Downie (2003) notes that retrieval of music is hampered by its complexity (pitch, tempo, rhythm, harmony, timbre, text, editorial, praxis, and bibliographic elements), the complexity of possible queries, and unresolved issues on query relevance, precision, and recall metrics.

Finding a piece of music for which you know its title, composer, and/or performer is relatively trivial if a recording of that piece exists. However, when your information is limited to what it sounds like, the task is much harder. To address this need, much research has been invested in systems that can take a "hummed" sequence and match it against a database of sound sequence descriptions, that is, query by humming systems (Ghias, Logan, Chamberlin, & Smith, 1995; Kosugi, Nishihara, Sakata, Yamamuro, & Kushima, 2000; Pauws, 2002). Ghias et al. (1995) convert a hummed sound into a sequence of pitch changes and match this against a MIDI database representing similar pitch changes in sequence. Kosugi et al. (2000) use a similarity based system in which beats rather than notes are sampled and compared against a target database; this deals with human inability to hum in tune, keep time, or start at the beginning of a song. Pauws (2002) by contrast uses both pitch changes and musical events (note starts, intervals, and gliding tones) as comparators. MPEG-7 descriptors have been used in QBH systems, Batke et al. (Batke, Eisenberg, Weishaupt, & Sikora, 2004) describes a system in which both fundamental frequency and melody-contour descriptors can be matched against sampled and transcribed hummed queries.

CONCLUSION

We have seen that there is a wide range of approaches to the modeling and usage of metadata and semantic annotation for recorded music. While most media metadata practitioners are attempting to move towards unified or interchangeable schemata, no such impetus exists for communities interested in the application of meta-information to music. Part of this is due to the fact that each community and subcommunity has specific needs and, as such, a unified scheme is unlikely. Even within each subcommunity, individual organizations or persons have their own ways of dealing with music. In music-related online forums a common question is how do you tag your music? A question for which there are increasingly varied answers.

Downloaded by [71.225.4.231] at 11:55 11 July 2011

NOTES

1. No. Beethoven never named his fifth piano concerto—the name came much later.
2. Many composers died after composing nine symphonies (Mahler, Schubert, Beethoven, Bruckner, Vaughan Williams, Dvorak)

REFERENCES

Alberink, M., Rutledge, L., Hardman, L., & Veenstra, M. (2004). *INS-E0409: Clustering semantics for hypermedia presentation. CWI Report*.

Andresen, L. (2001). The use of metadata in Denmark, *International Conference on Dublin Core and Metadata Applications* (pp. 272–277). Tokyo, Japan.

Athanasiadis, T., Tzouvaras, V., Petridis, K., Precioso, F., Avrithis, Y., & Kompatsiaris, Y. (2005). *Using a multimedia ontology infrastructure for semantic annotation of multimedia content*. International workshop on knowledge markup and semantic annotation. Galway, Ireland, November.

Ayres, M. (2003). *MusicAustralia: Experiments with DC.Relation*. Paper presented at the DC-ANZ (Dublin Core in Australia and New Zealand) Conference, Canberra, Australia, February.

Batke, J., Eisenberg, G., Weishaupt, P., & Sikora, T. (2004). *A query by Humming system using MPEG-7 descriptors*. Paper presented at the Audio Engineering Society Convention. Berlin, Germany, May.

Bertini, M., Cucchiara, R., del Bimbo, A., & Torniai, C. (2005). *Video annotation with pictorially enriched ontologies*. Paper presented at the ICME 2005 IEEE International Conference on Multimedia and Expo, July 2005.

Bormans, J., Gelissen, J., & Perkis, A. (2003). MPEG-21: The 21st century multimedia framework. *IEEE Signal Processing Magazine*, *20*(2), 53–62.

Casey, D. (2002). *AUDIOMD: Audio Technical Metadata Extension Schema*. Retrieved from http://lcweb2.loc.gov/mets/Schemas/AMD.xsd

Celma, O., Herrera, P., & Serra, X. (2006). *Bridging the music semantic gap*. Paper presented at the International Conference on Semantics and Digital Media Technology (SAMT), Athens, Greece, December.

Downloaded by [71.225.4.231] at 11:55 11 July 2011

Chang, S., Sikora, T., & Purl, A. (2001). Overview of the MPEG-7 standard. *IEEE Transactions on Circuits and Systems for Video Technology*, *11*(6), 688–695.

Diepold, K., Pereira, F., & Chang, W. (2005). MPEG-A: Multimedia application formats. *IEEE Multimedia*, *12*(4), 34–41.

Downie, S. (2003). *Toward the scientific evaluation of music information retrieval systems*. Paper presented at the 4th International Conference on Music Retrieval (ISMIR, 2003), Baltimore, MD, October.

Duan, Z., Lu, L., & Zhang, C. (2008). *Collective annotation of music from multiple semantic categories*. International Conference on Music Information Retrieval (ISMIR), Philadelphia, PA, September.

Ferrara, A., Ludovico, L., Montanelli, S., Castano, S., & Haus, G. (2006). A semantic web ontology for context-based classification and retrieval of music resources. *ACM Transactions on Multimedia Computing, Communications, and Applications*, *2*(3), 177–198.

Ghias, A., Logan, J., Chamberlin, D., & Smith, B. (1995). Query by humming: Musical information retrieval in an audio database. The Third Annual International Conference on Multimedia. San Francisco, CA, November.

Gómez, E., Gouyon, F., Herrera, P., & Amatriain, X. (2003). *Using and enhancing the current MPEG-7 standard for a music content processing tool*. Paper presented at the Audio Engineering Society Convention, Amsterdam, The Netherlands, March.

Gómez, E., Klapuri, A., & Meudic, B. (2003). Melody description and extraction in the context of music content processing. *Journal of New Music Research*, *23*(1), 23–40.

Good, M. (2001). MusicXML: An Internet-friendly format for sheet music, *Proceedings of XML*, Boston, MA, December 9–14.

Gouyon, F., Dixon, S., Pampalk, E., & Widmer, G. (2004). *Evaluating rhythmic descriptors for musical genre classification*. Paper presented at the AES 25th International Conference, London, UK, June.

Harvard University Library. (2004). *Administrative metadata for digital audio files, internal report*. Cambridge, MA: Harvard University Library. JTC1/SC29/WG11, I. I. (2004). *MPEG-7 Overview (version 10)*.

Hemmasi, H. (1992). ARIS Music Thesaurus: Another view of LCSH. *Library Resources & Technical Services*, *36*(4), 487–503.

Hemmasi, H. (1994). The Music Thesaurus function and foundations. *Notes*, *50*(3), 875–882.

Hemmasi, H. (1995). The Music Thesaurus Project at Rutgers University. *New Jersey Libraries*, *28*(2), 21–23.

Kim, K., Schreiner, F., & Diepold, K. (2008). *MAF overview*. Korea: ISO.

Kiryakov, A., Popov, B., Terziev, I., Manov, D., & Ognyanoff, D. (2004). Semantic annotation, indexing, and retrieval. *Web Semantics: Science, Services and Agents on the World Wide Web*, *2*(1), 49–79.

Kosugi, N., Nishihara, Y., Sakata, T., Yamamuro, M., & Kushima, K. (2000). A practical query-by-humming system for a large music database. Paper presented at a meeting of the *ACM SIGMM*. Marina del Rey, CA, October 30–November 4.

Lowood, H. (2004). *Playing history with games: Steps towards historical archives of computer gaming*. Paper presented at the Annual Meeting of the American Institute for Conservation of Historic and Artistic Works, Portland, OR, November.

Minibayeva, N., & Dunn, J. (July 2002). A digital library data model for music. Paper presented at the *ACM/IEEE-CS joint conference on digital libraries* (pp. 154–155), Portland, OR.

Musick, N. (2004). *Copyright issues in digital media* (report): Congressional Budget Office.

NISO. (2004). Understanding metadata. *NISO Standard*.

Pachet, F. (2003). Content management for electronic music distribution. *Communications of the ACM 46*(4), 71–75.

Pachet, F., & Cazaly, D. (2000). *A taxonomy of musical genres*. Paper presented at the Content-Based Multimedia Information Access Conference (RIAO), Paris, April.

Papadopoulos, G., Mylonas, P., Mezaris, V., Avrithis, Y., & Kompatsiaris, I. (2006). Knowledge-assisted image analysis based on context and spatial optimization. *International Journal on Semantic Web & Information Systems*, *2*(3), 17–36.

Pauws, S. (2002). CubyHum: A fully operational query by humming system. Paper presented at the 3rd International Conference on Music Information Retrieval, Paris, France, October.

Pavia, R. P., Mednes, T., & Cardoso, A. (2004). *An auditory model based approach for melody detection in polyphonic musical recordings*. Paper presented at the Computer Music Modeling and Retrieval, Esbjerg Denmark.

PREMIS. (2004). *Implementing preservation repositories for digital Materials: Current practice and emerging trends in the cultural heritage community*. OCLC/RLG.

PREMIS. (2008). *PREMIS: PREservation Metadata Implementation Strategies: Background*. Retrieved from http://www.oclc.org/research/projects/pmwg/background.htm

Rauber, A., & Frühwirth, M. (2001). Automatically analyzing and organizing music archives. *Lecture Notes In Computer Science*, *2163*, 402–414.

Riley, J., Mullin, C., Colvard, C., & Berry, A. (2008). *Definition of a FRBR-based metadata model for the Indiana University Variations3 project. Phase 2: FRBR Group 283 Entities and FRAD (Report)*. Bloomington, IN: Indiana University.

Salembier, P., & Smith, J. (2001). MPEG-7 multimedia description schemes. *IEEE Transactions on Circuits and Systems for Video Technology*, *11*(6), 748–759.

Scheirer, E., & Vercoe, B. (1999). SAOL: The MPEG-4 Structured Audio Orchestra Language. *Computer Music*, *23*(2), 31–51.

Selfridge-Field, E. (1997). *Beyond MIDI: The handbook of musical codes*. Cambridge, MA: MIT Press.

Swartz, A. (2002). MusicBrainz: A semantic web service. *IEEE Intelligent Systems*, *17*(1), 76–77.

Turnbull, D., Barrington, L., Torres, D., & Lanckriet, G. (2008). Semantic annotation and retrieval of music and sound effects. *IEEE Transactions on Audio, Speech, and Language Processing*, *16*(2), 467–476.

Vellucci, S. (2007). Chapter 12: FRBR and music. In A. Taylor (Ed.), *Understanding FRBR: What it is, and how it will affect our retrieval tools*. Westport, CN: Libraries Unlimited.

Downloaded by [71.225.4.231] at 11:55 11 July 2011

Wactlar, H., & Christel, M. (2002). Digital video archives: Managing through metadata. In *Building a national strategy for preservation: Issues in digital media archiving*. Washington, DC: Council on Library and Information Resources.

Wang, T., Siao, Y., & Su, A. (2006). JavaOL—A Structured Audio Orchestra Language: Tools, Player and Streaming Engine. Paper presented at the Audio Engineering Society Convention, Paris, France, May.

Weyde, T., & Wissmann, J. (2007). Automatic semantic annotation of music with harmonic structure. Paper presented at the Sound and music computing conference, Lefkada, Greece.

Downloaded by [71.225.4.231] at 11:55 11 July 2011