Audio Engineering Society

Convention Paper

Presented at the 133rd Convention
2012 October 26–29 San Francisco, USA

Knowledge representation issues in audio related metadata model design

György Fazekas, and Mark B. Sandler

Centre for Digital Music (C4DM), Queen Mary University of London, London E1 4NS, United Kingdom

Correspondence should be addressed to György Fazekas (gyorgy.fazekas@eecs.qmul.ac.uk)

ABSTRACT

In order for audio applications to interoperate, some agreement on how information is structured and encoded has to be in place within developer and user communities. This agreement can take the form of an industry standard or widely adapted open framework consisting of conceptual data models expressed using formal description languages. There are several viable approaches to conceptualise audio related metadata, and several ways to describe conceptual models, as well as encode and exchange information. While emerging standards have already been proven invaluable in audio information management, it remains difficult to design or choose the model that is most appropriate for an application. This paper facilitates this process by providing an overview, focussing on differences in conceptual models underlying audio metadata schemata.

1. INTRODUCTION

The importance of audio related metadata has been steadily increasing over the last decade due to the proliferation of digital content in audio archives, personal music collections, and the World Wide Web. Similarly, there has been a significant increase in the number of audio applications that are used in the production, distribution, storage and consumption of audio and music. Metadata became a crucial component in most solutions, whether it is for allowing access to stored content, or supporting a production and distribution workflow chain. For instance, ID3 metadata tags embedded in compressed audio files empower online music communities, despite the serious limitations the format exhibits in expressing even the most basic editorial information.

The usefulness of associating audio (and more generally, media) items with additional information is obvious, however the questions of how these associations should be made, and how metadata should be represented and organised present problems that are not easily solved. Numerous organisations pub-

lished constructs ranging from simple file formats, through abstract data models and metadata standards, to complex knowledge representation frameworks. They provide definitions and schemata in a variety of languages and define different methods for encoding information. The result is a large number of different approaches to represent audio related information [26, 24, 31, 32, 8].

1.1. Motivation

In 2010 EDINA, the National Data Centre of the Joint Information Systems Committee (JISC) in the UK conducted a feasibility study of creating an aggregation of metadata about images and time-based media [10], including film and sound. While this study focusses on a specific community and a broader set of media, we can easily argue that the findings are especially applicable to audio, a specific kind of time-based media, and to audio communities at large. The critical points of the final report related to metadata aggregation and model can be summarised as follows:

- Metadata, as well as automatic enrichment and aggregation are particularly important as full text indexing can not be applied to the content.
- The variety of implementations of standard metadata profiles leads to complexity in any harmonisation attempt.
- Deploying a common schema may actually loose information, and so may not provide sufficiently useful metadata.

Metadata formats are typically created for a particular application such as archiving, or with other distinctly selected set of use cases in mind, therefore they commonly address a domain or sub-domain with a specific set of requirements. This results in different conceptualisations in the data model underlying published schemata, which, in many cases creates artificial domain boundaries making interoperability between applications difficult, even if linking or sharing otherwise overlapping information would be useful. These issues can be observed in the areas of music and multimedia information management, which exhibit a number of problems related to both semantic and syntactic interoperability. The differences are most obvious when examining metadata models used in end user applications, archives, or production tools, and when trying to harmonise information related to content and production. It is also very visible when considering music related data repositories exposed on the Web using proprietary Application Programming Interfaces (API), since information returned by distinct API calls typically adhere to unique and distinct metadata models.

The findings of the EDINA report demonstrate an apparent conflict that plague the field of audio information management. Although aggregating or linking metadata is useful, it requires harmonisation between schemata, which is exceedingly difficult, while a common schema may not serve specific tasks or communities well enough.

1.2. Scope

Providing a satisfactory solution to the problems outlined in the previous section is beyond our reach. Instead, on the premise that this may shed some new light on these issues, we examine the conceptualisations underlying some well known metadata formats. Bringing together results of interdisciplinary research, we also examine how the modularity of Semantic Web ontologies facilitate the use of more fundamental (and less task specific) conceptualisations that are valid across broader domains. We discuss how these can be linked to task or application specific models, thus making mapping or linking community, domain or task specific metadata formats feasible, at least in cases where such fundamental models and potential cross domain requirements are observed at the time when new metadata models are drafted.

The paper has three main goals: *i)* Draw attention to different possible conceptualisations of audio related metadata and the consequences of their use. *ii)* Review the conceptualisations underlying frequently used or cited metadata formats, thus provide an entry point to newcomers in the field, without being exhaustive or getting bogged down in domain specific details. *iii)* Raise the awareness of certain *semantic* information management technologies and Semantic Web technologies in particular. Thus, we hope that metadata professionals not yet familiar with these technologies will find this study useful. Finally, this paper complements our most recent technical contribution discussed in [39] focussing on

the use of both editorial and content-based metadata for the control of adaptive digital audio effects.

1.3. Organisation

So far we have used the terms *metadata*, *conceptualisation*, *ontology* or *domain* without definitions. Since the paper aims be useful at an introductory level, we first attempt to make the meaning of "jargon" clearer in this particular context. The rest of the paper is then organised as follows: In Section 2 we review different possible conceptualisations of the audio and music domain, in Section 3 we discuss how knowledge embedded in these conceptualisations may be represented and review the prevailing technologies for encoding data. In Section 4 we enumerate commonly used metadata models and ontologies, and in Section 5 we examine how they are related to the conceptualisations discussed earlier. Finally in Section 6 we draw some conclusions and offer some recommendations.

1.4. Preliminaries

Data and Information Data is a set of symbols with no definite meaning. One may argue that structure can be learned from data which gives rise to meaning. An important question to consider however is that of "sameness of reference" studied by Quine in the context of denotation and truth and semantic agreement [28]. Without a shared conceptual framework and common references, neither data nor emergent pattens carry useful information. In agreement with [1], *information* can then be thought of as data structured within a certain context such that the relations between data are formalised.

Metadata is commonly explained as data about data [21], but its meaning carries ambiguities, inconsistencies, and variation. For one, it may refer to a datum related to another datum such as the musical key of a piece. Metadata may refer to data structures or containers which describe data records, that is, the relations between data items. Metadata can also be associated with other metadata, leading to relational or hierarchical structures.

Domain A domain of interest indicates that one is not interested in modelling the whole world, but rather in modelling just those parts of a specific sphere of activity or knowledge that is relevant to a task.

Conceptualisation represents an abstract model of the concepts we directly admit (or anticipate) in a domain of interest, the arrangement of, and relevant relations between these concepts. Knowledge about entities described in an abstract way is called a conceptual model which will lead to some schema.

Schema describes the structure of some data. It may represent knowledge, describe a model or simply prescribe the syntax of documents.

Ontology represents a branch of Metaphysics in Philosophy, dealing with "a particular theory about the nature of being or the kinds of existence"[1]. Artificial intelligence and computer science borrowed the term from the Greek philosophers and ontologists Socrates and Aristotle who first described classification schemes that bear relevance to modern day data structures. In these fields, ontology is most commonly defined as an "explicit formal specification of a shared conceptualisation of a domain of interest" [33, 13]. This definition implies that an ontology represents consensual knowledge about the entities and their relationships in a domain, preferably expressed using a machine-processable formal language. Every ontology is therefore a schema, but whether the converse is true depends on the degree of formality of an ontology. For instance, the definition given in [7] requires rich axiomatisation and expressive semantics to be present in an ontology that sets it apart from XML Schema[2], topic maps[3], database or classification schema and object models expressed using the Unified Modelling Language (UML).

Knowledge Representation is concerned with methods of structuring information such that automatic interpretation or formal reasoning becomes possible. Semantic Web ontologies are examples of knowledge representations. One particular advantage of ontologies, a direct consequence of reasoning support, is that it is possible to automatically identify entities that represent the same thing even if the definitions differ in the actual terminology, language, spelling or syntax. Thus, it is possible to create mappings between ontologies that cover different but overlapping domains automatically.

[1] http://www.merriam-webster.com/dictionary/ontology
[2] http://www.w3.org/XML/Schema
[3] http://www.topicmaps.org/

2. CONCEPTUALISATION

A conceptualisation requires a model of a domain that is different from schema or ontology and may be represented in them using some degree of formality. A conceptualisation therefore concerns, at the most abstract level, the arrangement of concepts and relationships we admit in our model. In this section, we outline different categories of music related metadata elements that we may wish to model, as well as some common arrangements of entities relevant to these data.

2.1. Metadata categories

Metadata related to music, multimedia and intellectual works in general may be grouped into the following categories:

- **Bibliographic Information**: Editorial type information associating works with *creators*, and *publishers*, simple *subject headers*, and items with *location*, *condition* and other cataloguing information.
- **Cultural Information**: Social information pertaining to *consumption* and *appreciation* of works by different social groups, *relations between works or creators*, *intellectual property rights* and more generally, information that needs to be interpreted in a certain *cultural context*.
- **Content-based Information**: Particular *features* of content that facilitate organisation, search, navigation or similarity calculation. Features may result from *simple transformations* or complex *semantic analysis* resulting in representations that can be used to answer *meaningful queries*.
- **Provenance and Workflow Information**: Detailed description pertaining to the *origin* and *production process* of intellectual works, such as elements of the publishing and production workflow chains.

Unfortunately, the boundaries between these categories are not always clear leading to many different possible conceptualisations. For example, we can envisage a hierarchical system for content-based annotation of audio files blurring the boundaries between automatically extracted low-level features, and higher level musicological terms which often exhibit a strong cultural bias. It can also be argued that rights information forms a distinct group, while it is fair to say that the meaning of rights metadata depends on the jurisdiction bound to a territory and its culture. In the next section, we look at different arrangements from a conceptualisation point of view.

2.2. Hierarchical models

Entities in a conceptual model are often organised hierarchically. The two most common organising principles are taxonomies which express a hierarchy of *is-a* or *type-of* relationships between objects and/or their properties independently, and partonomies which express meronomical or *part-of* relations between objects. A typical example of taxonomical organisation is genre classification, where a tree of genre and sub-genre labels are created and audio items are classified accordingly. Another example involving meronomical organisation would be the relation of a box-set to its constituent albums, or the description of channels present within an audio file or stream. An example of a more complex hierarchical model is the XML-based music encoding initiative (MEI)[4] aimed at the description of music notation documents, providing a structural decomposition down to the actual score. The common feature of hierarchical models is the fact that they are typically centred around the notion of *audio objects* or items, (e.g. albums, files, streams) and enable the classification of these in isolation, largely following the basic principles of how books are commonly organised in a library.

2.3. Conceptual layering

Another possible organising principle might be conceptual layering, typically ranging from abstract to concrete entities. Content annotation is a particular example where many conceptual layers are possible. For instance, we may characterise audio features according to physical quantities (e.g. fundamental frequency of a note), perceptual qualities (e.g. perceived pitch) or musical categories (e.g. musical note). Without further explanation, some additional examples are provided in Table 1.

[4] http://music-encoding.org/

physical	perceptual
sub-symbolic	symbolic
contextual	primary
instantaneous	durational
dense (compact)	sparse (scattered)
signal domain	transform domain

Table 1: Example layers of content-based audio features

A well-known example of a layered representation system is the Functional Requirements for Bibliographic Records (FRBR). It aims at providing a framework that identifies and clearly defines the entities of interest to users of bibliographic records, the attributes of each entity, and the types of relationships that operate between entities [27].

FRBR defines a set of entities related to products, creators and subjects of intellectual works. The first group depicted in Figure 1 describes products and includes entities ranging from abstract to concrete.

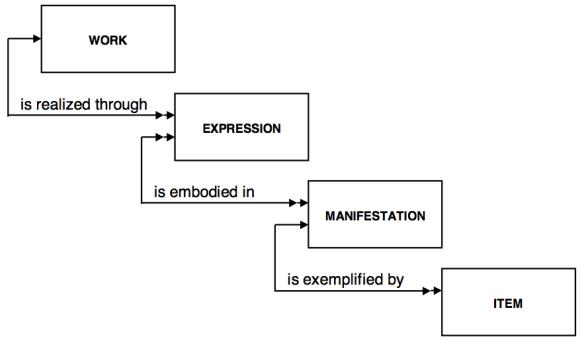

Fig. 1: Entities related to products of intellectual works in the FRBR model [27]. (Double arrows represent 1:N or N:N relationships.)

Work may stand for a poem, the lyrics of a song, or a classical composition. *Expression* represents a particular realisation that remains intangible and reflects artistic qualities, such as a recital of a musical piece, or illustrations in a book. *Manifestation* represents all the physical embodiments of an expression that bear the same characteristics, with respect to both intellectual content and physical form. For example, a book about the Semantic Web published in 2011. *Item* is the only concrete entity in the model, a single exemplar of a manifestation, for instance, a copy of the aforementioned book on my shelf, a compact disc in the collection of the British Library, or an audio file on my computer.

FRBR is particularly interesting since research showed that music libraries would benefit the most from using it, but it also proved valuable in other domains such as fine arts and literature [6]. FRBR provides useful concepts and relationships to describe the production workflow of intellectual works, however, it is not sufficient in itself for this purpose. For example, we cannot express temporal relations between certain stages of a work, which is why process or event based models are important in the audio domain.

2.4. Process, event or workflow-based view

The ABC model [20] is the result of harmonisation between the digital libraries and museum communities. It takes several metadata packages such as the previously mentioned FRBR [35, 27], as well as the CIDOC/CRM [5] models into account. However, rather than attempting to build a universal model, it simply aims to find the most frequently occurring set of entities and relationships (e.g. events, places or people), that can be considered domain independent.

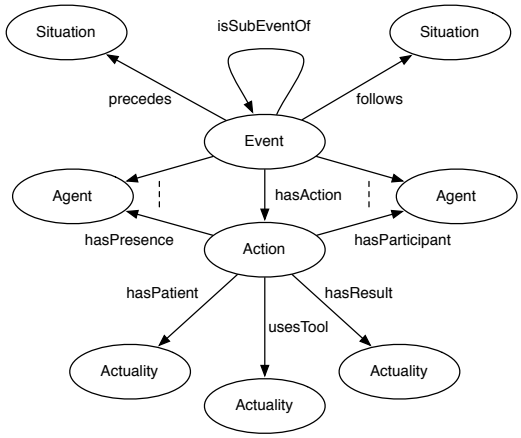

Fig. 2: ABC model showing terms related to situations events and actions (partial).

In the centre of ABC's model are *i)* a layered conceptualisation of the life cycle of intellectual works, and *ii)* an event model centred around three concepts: *situation, event, action*, as shown in Figure 2. ABC's conceptualisation is similar to FRBR. It

includes entities ranging from abstract to concrete such as *work*, *manifestation*, and *item*, however, it leaves the *expression* layer out, which turns out to be of paramount interest in describing music production workflows [29]. ABC considers events as transitions between situations, linked with actions performed on *actualities* (agents or physical artefacts). This can be used to describe the transformations of an item. Using the property `hasPresence`, events or actions may be linked with agents that are simply present during an event. Agents that actively participate are linked using `hasParticipant`, a subproperty of `hasPresence`. ABC is an example of a model which may be used to harmonise metadata schemata. This is demonstrated in [16], fusing a museum ontology (CIDOC/CRM), multimedia content description (MPEG-7), rights management (MPEG-21), and a biomedical ontology (ON9.3) under ABC's common framework, based on principles of an event-aware integration model described in [17].

2.5. Provenance

The notion of provenance is important in the music domain, particularly in the context of rights management and process history. Provenance may simply mean *'source of origin'* (of an artefact). A more complex interpretation, often used in computing as well as by historians and museums, is *'chain of custody'* or origin to present. The e-Science community [34] in turn is typically interested in *'process provenance'* — that is, the full execution history of processes which were utilised to compute some data. In this sense, provenance encoding is a specific type of workflow encoding, and a crucial component of workflow systems. The Open Provenance Model [23], is a prime example focussing on process provenance.

Similarly to other provenance models, OPM defines three core concepts: artefacts, processes and agents. An *artefact* is defined as an *immutable piece of state* which may refer to an actual physical object, a digital representation or some digital data. A *process* represents an action that creates artefacts, either by acting on an existing artefact or by creating a new one. *Agent* describes an entity involved in a process by enabling or controlling its execution. The basic relationships between the concepts defined in OPM is shown in Figure 3.

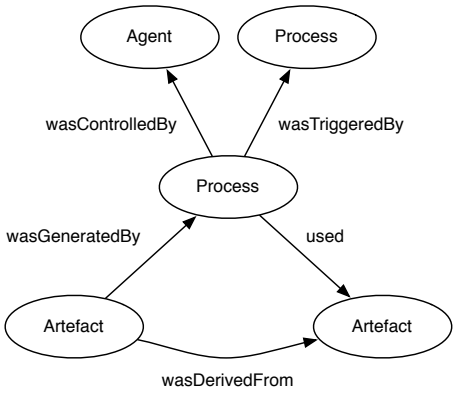

Fig. 3: Entities in the Open Provenance Model

3. REPRESENTATION

In the previous section, we outlined some important concepts and models for representing metadata related to audio and music. In this section, we present a brief overview of the options available for the representation of conceptual models, as well as information expressed using these models.

3.1. XML Schema

Perhaps the most common way of representing a conceptual model is still the provision of a detailed description in natural language. However, in order to facilitate the use of these models in software applications, it is necessary to provide formal, machine-readable definitions. This enables for instance the validation of some data against a model.

A widely used method for expressing certain types of models is the use of XML Schema. Similarly to its predecessor the Document Type Definition (DTD) language, its goal is to define the structure of XML documents. Using XML Schema we can define a set of elements corresponding to valid tags in a document. Elements may be complex or simple types, depending on whether they are allowed to have attributes and include other elements, while attributes are simple types that may be associated with a complex element. We may also define a set of restrictions on the acceptable values for XML elements and attributes.

Albeit XML-based technologies are very commonly used, they are designed primarily to define document structure, therefore they are rather limited in

capturing the semantics related to entities in a conceptual model. For instance, there is no built in support for encoding hierarchical relationships between types. Therefore, when implementing XML-based metadata models, it is always necessary to look at the associated specification. Detailed comparison between XML and its alternatives is beyond the scope of this paper, however, there is a rich literature discussing how XML Schema is related to ontologies and more recent W3C standards associated with Semantic Web technologies [12, 19].

3.2. Semantic Web technologies

The basic idea behind the Semantic Web [3] is to ease information seeking tasks by facilitating the aggregation of Web content, as well as using automated reasoning over Web content. Representing the heterogeneous information on the Web however is a difficult task from a knowledge representation point of view. Therefore, this became an active area of research. The set of techniques, often termed Semantic Web technologies, amalgamate different methods for representing and linking information. The Uniform Resource Identifier (URI) provides a conventional unique naming scheme for ontological concepts and relationships. In the context of the Web these are called *resources*. The Hypertext Transfer Protocol (HTTP), which provides basic methods for obtaining resources identified by HTTP URIs, is the fundamental linking mechanism used on the Web, and on top of these we find the Resource Description Framework (RDF)[5] and various Semantic Web ontologies built on the foundations of this framework.

3.3. Resource Description Framework

Rather than describing document structure, the Resource Description Framework was designed to provide a simple data model that is based on idea of making statements about resources in the form of subject predicate object (s,p,o) triples. A collection of triples can be seen as a graph, with nodes representing subjects and objects, and edges representing predicates. A large set of statements form a complex network of semantic relationships. All elements of a triple may be named by a URI, which enables RDF graphs to be linked and distributed across the Web. The RDF model decouples syntax from semantics, hence several serialisations formats exists including XML-based and concise and human readable non-XML based formats such as N-Triples (see Fig. 4) and Turtle.

[5] http://www.w3.org/RDF/

3.4. RDF Schema and OWL

Although RDF provides a fundamental data model, it does not have the facilities for expressing complex relationships required for modelling a domain. In order to precisely communicate information using RDF statements, we have to be able to define and later refer to concepts such as a specific algorithm we use for audio processing, its concrete implementation and its parameters. We also need a vocabulary of well defined relationships in an application.

There is a hierarchy of W3C recommendations for this purpose. This includes the RDF Schema Language (RDFS)[6] for defining classes and properties of RDF resources, and the OWL Web Ontology Language (OWL)[7] for making RDF semantics more explicit. OWL-1 has three layers, corresponding to three levels of expressiveness. OWL-DL is the most commonly used layer, closely linked to Description Logics (DL), although other OWL flavours also have DL equivalents. Using this language we can impose restrictions on the range and domain types of properties, or constraints on cardinality, the number of individuals linked by a property. An updated language OWL-2[8] has been released to increases the expressiveness of OWL. This language remains backward compatible with the previous version.

4. MODELS AND STANDARDS

Having outlined some fundamental conceptual data models and technologies for representing them, we now briefly review some commonly used metadata models, standards and frameworks, including relevant ontologies, but without being exhaustive.

4.1. Dublin Core

The Dublin Core Metadata Initiative (DCMI) maintains a set of ISO standardised metadata elements. Dublin Core (DC) is a minimal set[9] for describing resources (information sources), documents or digital artefacts. It includes 15 core elements : title,

[6] http://www.w3.org/TR/rdf-schema/
[7] http://www.w3.org/TR/owl-semantics/
[8] http://www.w3.org/TR/owl2-overview/
[9] http://dublincore.org/documents/dces/

Fig. 4: Graph of an N-Triples statement with URI references: In a common graphical formalism resources are depicted using ovals, while literals may be represented by ovals or rectangular boxes.

creator, subject (topic), description, publisher, contributor, date, type, format, identifier, source (a resource the described resource was derived from), language, relation (to other resources), coverage (spacial, temporal, jurisdictional), and rights (held in or over the resource).

DC is not specific to any media, rather, it provides a cross-domain way of describing a wide range of resources. Dublin Core recognises that one particular syntax specification would not fit all applications, hence its specification remains purely conceptual. However, most implementations are based on RDF, which brings the widest range of possible uses forth through its resource linking mechanism. DC has a fundamentally object based conceptualisation, that is, it's aimed to attach metadata to a single item or object. It is often reused in larger frameworks, including some recent EBU and AES metadata standards.

4.1.1. ID3

Perhaps the most commonly used metadata format for encoding bibliographic information in the music domain is ID3[10]. It defines a set of metadata containers (tags) originally designed to be included in MP3 audio files, but later adopted by other formats such as AIFF, WMA and MP4. Its first version ID3v1 provides a limited set of tags such as *artist*, *album*, *title*, *year* and *genre*, each able to hold only a fixed number of characters. The second, incompatible version ID3v2 relaxes this limitation by allowing variable length frames, and extends the type of information that can be included.

Unfortunately it is still very common that the ID3v1 tag set is the only metadata available in a music collection. ID3v1 has a strong bias for representing information about popular music, for example, it has no tags to describe a composer and a performing artist separately, making it highly unsuitable for describing classical music. Although ID3v2 introduces new elements for this purpose, the specification remains ambiguous. For example, the TPE2 frame may stand for *band*, *orchestra* or *accompaniment*. For these reasons ID3 is an example of a format which lacks coherence and extensibility. It is also limited to one level of description strictly related to audio items as identified in [29] — that is, we cannot use it to provide information about a band or an album associated with a song other than their name or title.

4.2. iXML

In the field of professional broadcast audio, iXML[11] is the most commonly used format to embed metadata into Broadcast Wave Format (BWF) files. It serves similar needs to the previously used BEXT format, however it allows to include standard XML data using the iXML specification. Its primary aim is to include workflow related and location information associated with (typically multichannel) audio content. For instance, it can relate a file to a project, a specific scene, or a take, but it can also describe individual tracks and sync-points as well as limited file history. The conceptualisation of iXML is hierarchical and centred around audio objects.

4.3. AES60 and EBU Core

The AES recently published a minimum set of metadata elements for describing audio resources. The AES Core standard[12] has been co-published with the EBU Core standard to facilitate interoperability across industries. While AES60 focusses on audio metadata, EBU Core[13] is broader and includes elements that are relevant in broadcasting, for instance, to associate media resources with events, places, a target audience or rating. Both standards have a hierarchical, object based conceptualisation, and rely

[10] http://www.id3.org/

[11] http://www.ixml.info/

[12] http://www.aes.org/publications/standards/

[13] http://tech.ebu.ch/lang/en/MetadataEbuCore

primarily on Dublin Core. While AES60 is published in text with an accompanying XML Schema, EBU Core is also available as an OWL ontology.

4.4. AES57

Another recent AES standard, AES57, focusses on describing audio files and physical formats alike for the purposes of archival storage and preservation. The conceptualisation of this standard is also based on audio objects, which is the top level entity in its schema. This standard is particularly well suited for describing details relevant to audio file encoding, or the physical properties of different media including tapes and disks. AES57 is accompanied by and XML Schema specification.

4.5. DDEX

DDEX is an industry consortium that develops standards for media, music licensing, and digital service delivery. The DDEX standard most relevant to audio related metadata concerns musical works licensing[14]. Due to the complex requirements of this area, this standard includes a wide range of entities related for instance to artists, audio files, and right holders. It has a layered conceptualisation that separates works, manifestations and audio items similarly to FRBR, however, it leaves the expression layer out, which is particularly important in describing production workflows. Although published using an XML Schema, DDEX also includes an element hierarchy by providing its own tools for expressing sub-class, super-class and is-a relations.

4.6. MPEG-7

MPEG-7 [22] is arguably the most well known standard for the annotation of media content (still images, audio, video, audiovisual items), as well as rich multimedia presentations. Providing an XML-based metadata framework for high and low-level content description is its core objective. It also contains elements to encode editorial and production information (content management), as well as data related to storage, transmission, navigation and user interaction. MPEG-7 provides a set of *descriptors* (**D**) to define how individual features are represented in a document, and *descriptor schemes* (**DS**) that specify the structure of descriptors and other descriptor schemes. In its Audio Framework, we find the low-level features including *AudioSpectrumEnvelope* which represent a logarithmic-frequency spectrum, or the *AudioSpectrumBasis* and *AudioSpectrumProjection* descriptors which describe basis functions derived from the singular value decomposition of a normalised power spectrum.

4.7. The Music Ontology Framework

The aim of the Music Ontology framework[15] [30] is to provide a comprehensive, yet easy to use and easily extended domain specific knowledge representation for describing music related information. Integration of music related resources (Web services and data repositories) on the Semantic Web, and facilitation of service integration and data communication in distributed music processing environments are among its existing applications [29]. It has certain properties which make it particularly suitable as basis for a general semantic audio information management framework as well as data collection in recording and production. For instance, it relies on, and extends the full FRBR model, and provides an event based conceptualisation of music production workflows as shown in Figure 5. The Music Ontology can be used to describe the following types of information:

- **Editorial metadata**: Concepts and relationships involving *artists, bands, labels, albums, tracks, audio files* or *downloads* and their identifiers in various databases.
- **Music production workflow**: The life cycle of musical works from *composition* through *performance*, to the produced *sounds and recorded signals* and their *publication*.
- **Event decomposition**: Further details about particular events in the production workflow such as *individual performances* by different musicians in a recording.
- **Content annotation**: Audio *signals* and *temporal annotation* of their content.

The ontology is published as a modular framework and includes extensions to describe chords, symbolic music notation, audio features, and studio production [9].

[14] http://www.ddex.net/works-licensing

[15] http://musicontology.com/

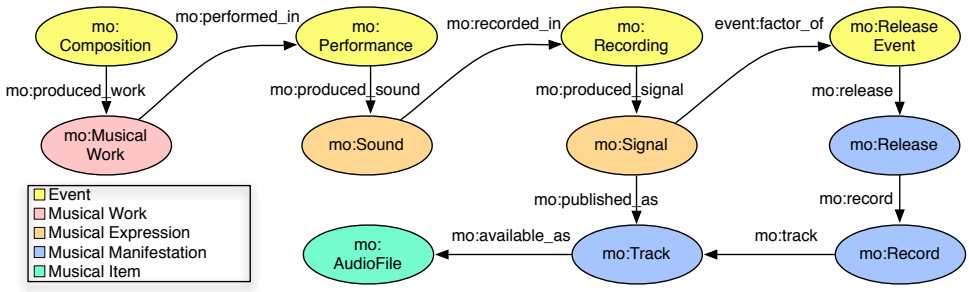

Fig. 5: Music Production Workflow Model: key concepts and selected properties showing how events (top row) make connections between the layers of the FRBR model.

5. ANALYSIS AND DISCUSSION

While our study is far from exhaustive, it covers a large enough area of the domain to draw some useful conclusions.

When examining the different specifications, their respective scope and application domains, it is relatively easy to observe that most models are nearly optimal for their applications. Very few models are designed with sufficient extensibility in mind however, which would be crucial to facilitate harmonisation or cross-domain utilisation. In case of MPEG-7, its usefulness in musical applications has been noted during early developments of the standard [4, 18], yet its long term adaptation remained low owing to a number of issues. The standard defines certain computational steps associated with descriptors, but allows for variability without providing sufficient tools for describing what exactly was computed. This hinders interoperability, and obscures the aim of the standard, making it difficult to decide whether its primary purpose is knowledge representation or data exchange. A number of additional concerns were reported in [15, 26, 37, 11, 24, 38, 29], including the lack of formal semantics, the lack of linking mechanism between certain descriptors types and the content, and the monolithic structure of MPEG-7 documents. Most of these concerns are direct consequences of the use of XML instead of description languages with better knowledge representation support, while only very few of the concerns can be resolved using MPEG-7 profiling [36]. While XML related issues also affect many recent standards, the semantics described in the text specifications are increasingly more precise, and there is a growing tendency for capturing the semantics in machine-processable representations. DDEX is a case in point where this is provided, however, the semantic relations are described using format specific constructs that lack generic tool support.

An interesting observation is that even though FRBR and similar layered models originate from music librarian use cases [6], the full FRBR model is not widely adopted, and currently only the Music Ontology uses all the four layers. We found that the expression layer, involving sounds and signals, is particularly important in describing engineering workflows and process history [9], since we can directly relate performances, and recordings within the workflows as shown in Figure 5. This provides a more detailed and flexible conceptualisation compared to describing these processes using audio objects only (in fact, some engineering decisions can not be expressed using a model restricted to audio objects). This justifies our choice of using the Music Ontology in further developments [39, 9].

6. CONCLUSIONS

We can conclude that XML-based technologies are by far the most common in audio-media metadata management, albeit RDF and OWL are more increasingly used in recently published standards, including EBU Core and the Ontology for Media Resources[16] published by the W3C Media Annotation Working Group [8].

[16] http://www.w3.org/TR/mediaont-10/

The tendency that audio related standards are optimised for small subdomains seems to suggest that Arrows's impossibility theorem [2, 25] is at work in the field, which states that "no ontology can be maximum for all individuals", that is, some group will lose, when one ontology is adopted over some other ontology. Please note that in this context, ontology stands for general conceptual model. There is however a potential solution to this problem, based on lessons learnt from amalgamated research in the database community. If individual communities are allowed to keep their task or domain specific solutions in a federation of loosely coupled systems, the problem of finding a model which suits all needs can be reduced to providing a federated system with sufficient means to interoperate, exploiting similarity between objects across different subdomains [14], thus the impossibility theorem does not apply. This is where the use of Semantic Web ontologies can be highly beneficial. In one scenario, ontologies can be layered upon existing standards, or can be created based on the overlapping elements of domain specific standards. If correspondence between the entities are found, the mapping can be used to provide interoperable services utilising the mapped ontologies. In another scenario, amenable XML-based standards can be published in multiple formats, as is the case with EBU Core. This enables a gradual progress towards more precisely captured ontological semantics, as well as sharing and linking multimedia information on the Web, even if the information is produced by legacy systems.

7. REFERENCES

[1] R. L. Ackoff. From data to wisdom. *Journal of Applied System Analysis*, 16:3–9, 1989.

[2] K. Arrow. Social choice and individual values. *Cowles Foundation Monograph, Wiley, New York, USA*, (12), 1963.

[3] T. Berners-Lee, J. Handler, and O. Lassila. The Semantic Web. *Scientific American*, pages 34–43, 2001.

[4] M. Casey. Musical applications of MPEG-7 audio. *Organised Sound, Cambridge University Press*, 6(2), 2002.

[5] N. Crofts, M. Doerr, T. Gill, S. Stead, and M. Stiff, editors. *Definition of the CIDOC Conceptual Reference Model (Version 5.0.2)*. First published by the ICOM/CIDOC Documentation Standards Group in 2003, continued by the CIDOC CRM Special Interest Group., 2010.

[6] T. J. Dickey. FRBRization of a library catalog: Better collocation of records, leading to enhanced search, retrieval, and display. *Information Technology and Libraries*, 27(1):23–32, 2008.

[7] M. Ehrig, S. Handschuh, A. Hotho, A. Maedche, B. Motik, D. Oberle, C. Schmitz, S. Staab, L. Stojanovic, N. Stojanovic, R. Studer, G. Stumme, Y. Sure, J. Tane, R. Volz, and V. Zacharias. The Karlsruhe view on ontologies. *Technical Report, Institute for Applied Informatics and Formal Description Methods (AIFB), Universität Karlsruhe*, 2003.

[8] J.-P. Evain and T. Bürger. Semantic web, linked data and broadcasting, more in common than you'd think! *EBU Technical Review*, 2011.

[9] G. Fazekas and M. B. Sandler. The Studio Ontology Framework. *in proc. 12th International Society for Music Information Retrieval (ISMIR-11), Miami, Florida, USA.*, 2011.

[10] S. Fraser, L. Halliday, J. Stewart, C. Ingram, and T. Stickland. Aggreagation of metadata about images and time-based media. *EDINA (JISC National Data Centre) Scoping Study Final Report v1.1*, 2010.

[11] R. García and O. Celma. Semantic integration and retrieval of multimedia metadata. *Proceedings of the 5th International Workshop on Knowledge Markup and Semantic Annotation*, 2005.

[12] Y. Gil and V. Ratnakar. A comparison of (semantic) markup languages. *FLAIRS Conference, AAAI Press.*, pages 413–418, 2002.

[13] T. R. Gruber. Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies*, 43:907–928, 1993.

[14] J. Hammer and D. McLeod. An approach to resolving semantic heterogeneity in a federation of aotonomous heterogeneous database systems. *International Journal of Cooperative Information Systems*, 2(1):51–83, 1993.

[15] J. Hunter. Adding multimedia to the Semantic Web - Building an MPEG-7 Ontology. *The first Semantic Web Working Symposium, Stanford University, July 30 - August 1, 2001, California, USA*, 2001.

[16] J. Hunter. Enhancing the semantic interoperability of multimedia through a core ontology. *IEEE Transactions on Circuits and Systems for Video Technology*, 13(1):49–58, 2003.

[17] J. Hunter and D. James. Application of an event-aware metadata model to an online oral history archive. *presented at ECDL2000, Lisbon*, 2000.

[18] H.-G. Kim, N. Moreau, and T. Sikora. *MPEG-7 Audio and Beyond: Audio Content Indexing and Retrieval*. John Wiley & Sons, 2005.

[19] M. Klein, D. Fensel, F. van Harmelen, and I. Horrocks. The relation between ontologies and XML schemas. *Linköping Electronic Articles in Computer and Information Science*, 2001.

[20] C. Lagoze and J. Hunter. The ABC Ontology and Model. *Journal of Digital Information - Special Issue - selected papers from Dublin Core 2001 Conference*, 2(2), 2002.

[21] D. L. Liu and M. T. Özsu. (eds). Encyclopedia of Database Systems. *Springer*, 2009.

[22] J. M. Martínez. Mpeg-7 overview. International Standard, ISO/IEC JTC1/SC29/WG11, Available online: `http://mpeg.chiariglione.org/standards/mpeg-7/mpeg-7.htm`, Retrieved: March 2011., 2004.

[23] L. Moreau, B. Clifford, J. Freire, J. Futrelle, Y. Gil, P. Groth, N. Kwasnikowska, S. Miles, P. Missier, J. Myers, B. Plale, Y. Simmhan, E. Stephan, and J. V. den Bussche. The open provenance model core specification (v1.1). *Future Generation Computer Systems*, 27(6):743–756, 2011.

[24] F. Nack, J. van Ossenbruggen, and L. Hardman. That obscure object of desire: multimedia metadata on the web, part 2. *IEEE Multimedia*, 12(1):54–63, 2005.

[25] D. O'Leary. Impediments in the use of explicit ontologies for KBS development. *International Journal of Human-Computer Studies*, 46(2-3):327–337, 1997.

[26] J. V. Ossenbruggen, F. Nack, and L. Hardman. That obscure object of desire: Multimedia metadata on the web (part i). *IEEE Multimedia*, 12:54–63, 2004.

[27] M.-F. Plassard, editor. *Functional Requirements for Bibliographic Records, Final Report*. International Federation of Library Associations and Institutions Study Group on the Functional Requirements for Bibliographic Records. Approved by the Standing Committee of the IFLA on Cataloguing, K. G. Saur, München, 1998.

[28] W. V. O. Quine. *From Stimulus to Science*. Harward University Press, Cambridge, Massachusetts, London, England, 1995.

[29] Y. Raimond. *A Distributed Music Information System*. PhD Thesis, School of Electronic Engineering and Computer Science, Queen Mary University, London, UK, 2008.

[30] Y. Raimond, S. Abdallah, M. Sandler, and G. Frederick. The Music Ontology. *in Proc. 7th International Conference on Music Information Retrieval (ISMIR 2007), Vienna, Austria*, 2007.

[31] J. R. Smith and P. Schirling. Metadata standards roundup. *IEEE Multimedia, April-June 2006*, 13(2):84–88, 2006.

[32] G. Stamou, J. van Ossenbruggen, J. Z. Pan, and G. Schreiber. Multimedia annotations on the Semantic Web. *IEEE Multimedia*, 13(1):86–90, 2006.

[33] R. Struder, R. Benjamins, and D. Fensel. Knowledge engineering: Principles and methods. *Data and Knowledge Engineering*, 25(1-2):161–198, 1998.

[34] I. Taylor, E. Deelman, D. Gannon, and M. Shields, editors. *Workflows for e-Science: Scientific Workflows for Grids*. Springer Verlag, 2006.

[35] B. Tillett. FRBR: A conceptual model for the bibliographic universe. *Library of Congress Cataloging Distribution Service*, 2004.

[36] R. Troncy, W. Bailer, M. Höffernig, and M. Hausenblas. VAMP: a service for validating MPEG-7 descriptions w.r.t. to formal profile definitions. *Multimedia Tools Appl.*, 46:307–329, 2010.

[37] R. Troncy, J. Carrive, S. Lalande, and J.-P. Poli. A motivating scenario for designing an extensible audio-visual description language. *International Workshop on Multidisciplinary Image, Video, and Audio Retrieval and Mining (CoRIMedia), October 25-26, 2004, Universite de Sherbrooke, Quebec, Canada.*, 2004.

[38] R. Troncy, Ò. Celma, S. Little, R. García, and C. Tsinaraki. Mpeg-7 based multimedia ontologies: Interoperability support or interoperability issue? *1st Workshop on Multimedia Annotation and Retrieval enabled by Shared Ontologies, Genova, Italy. December 5, 2007.*, 2007.

[39] T. Wilmering, G. Fazekas, and M. B. Sandler. High-level semantic metadata for the control of multitrack adaptive digital audio effects. *In proc. of the 133rd convention of the AES, San Francisco, USA.*, 2012.