A Novel Taxonomy for Consumer Metadata

Thomas M. Coughlin, *Senior Member,* IEEE, Scott L. Linfoot, *Senior Member*, IEEE

Abstract — *There is an urgent need for metadata to accompany and describe media data essence. For the growing amount of personal content, user generated descriptors tend to be vague to the point of uselessness. Standardization in the metadata format is needed to allow a full and useful description of content that is interoperable between consumer devices. Manufacturers need to ensure that the metadata generated by a device is complete and understood by other products as well as allow for the creation and use of more subjective metadata. This paper provides a background on metadata layers with the aim to create a standard model of metadata layers in consumer devices*[1].

Index Terms — Metadata, standardization, life log, life recorder, digital storage, physical, physiological, psychological, filter, dimensional extent, operational, textual, semantic, contextual, levels, layers

I. INTRODUCTION

There are now over 200 million hours of media material worldwide. A recent projection [1] has shown that there is a distinct possibility of an almost exponential increase in the amount of digital content that will be available. This growth is due to low-cost consumer devices amassing huge amounts of content through everyday activities. One example is the life-log [2].

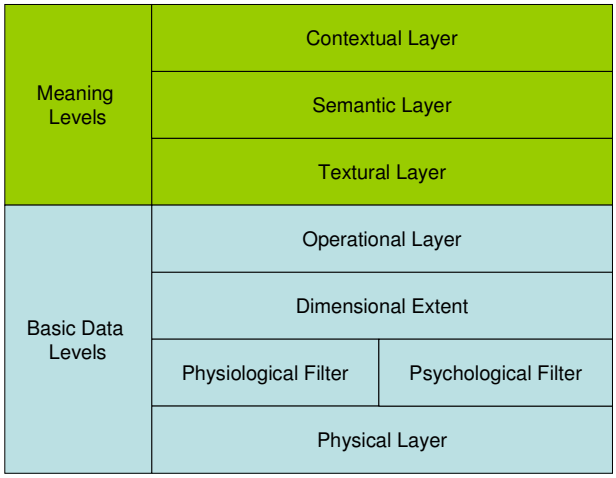
The problem with such devices is that it is becoming difficult to ascertain the media a consumer has and how find and manage this content. Automated metadata generation tools are now coming available but with no specific format, it is unclear how the metadata generated from such products will be useful for consumers.

Linfoot, Coughlin and Cowell presented a summary for the need of standardization of metadata [3] that emphasized four main criteria for a metadata scheme to be adopted by consumers: Flexibility, scalability, upgradability and simplicity. A metadata structure taxonomy is proposed that is comprehensive, can be parsed into any other propriety system and flexible enough to be of general use.

II. A METADATA TAXONOMY

The terminology used here treats metadata as a communications channel and has similarities to and is inspired by the OSI network model [4]. A 7-layer metadata ontology model is proposed where the uppermost layers give a more abstracted level of content metadata (the meaning) while the lower layers provide basic metadata. Figure 1 shows a graphical representation of this metadata model. These metadata layers are described below:

Figure 1. Metadata Layer Model



Layer 1, Physical Layer (*Sensory and source information)*: This is basic information about the content related to the source of the content (where and when) as well as sensory information of various sorts. The sensory information could include sound, sight, touch or smell in some defined fashion.

Layer 2, Physiological and Psychological Filtering: This metadata defines what sort of personal or experiential filtering is applied to the signal. This filtering may relate to the characteristics of the channel used to transmit (physiological) or experience (psychological) the metadata which may differ depending upon the type of data—e.g. speech or music may undergo different psychological filtering. Thus speech can be converted to text, which can be a filtered metadata giving useful and searchable information about what was said but speech to text conversion would do little to pass on the psychological import of a piece of music.

Layer 3, Dimensional Extent: This relates to the complexity of the content described as a set of orthogonal dimensions describing the content. For an image or video this layer may indicate whether it is flat or does it have depth as well. Likewise for audio content this could be used to describe the number of "voices" or the level of presence of the content (e.g. surround sound has more

[1] Tom Coughlin is President of Coughlin Associates, www.tomcoughlin.com, tom@tomcoughlin.com., S. L. Linfoot is with the Engineering Department, De Montfort University, Leicester, UK (e-mail: sllinfoot@dmu.ac.uk).

audio dimensions than a monaural sound). This concept of dimension could be applied to all of the senses with an interesting expansion of our ways of understanding dimensions in touch and smell,

Layer 4, Operational Layer: This level of metadata gives instructions on how to recreate the content in its intended form using defined hardware and software. For instance this level could include information on what operations are performed on the dimensional extent such as the number of frames per second, sampling rate, bit-depth, etc for video content.

Layer 5, Textural Layer: This level can be seen as a subclass of the next level (Semantic). It is metadata describing differences involving constructions built from the lower levels. For instance this could differentiate otherwise identical blue and red cars. A subset of this layer is data about use and interaction of content.

Layer 6, Semantic Layer: This is a concrete definition of the object or experiences in a piece of content based upon generally agreed upon constructions – for example "a tree" - "my friend said..." (as input from an audio byte where it is recognized that it is your friend speaking and he/she said…)

Layer 7, Contextual Layer: This level refers to the description of experience of content by a sensible sentient being. Current computers cannot create true judgmental information for metadata as they cannot look at a scene and define it as "beautiful" or listen to music and define it as "melodical", analyze a smell and refer to it as "pungent". The contextual level is by its nature subjective or personal—specific to the participant. A collection of contextual level metadata from several sensible sentient beings could be represented as providing a sort of temporary consensus on the "meaning" of that content.

The first 4 metadata levels are much more defined or mechanical while the last 3 are increasingly individualistic or subjective. All of these levels of metadata are important in fully describing a thing or experience as perceived by people. Metadata created in these 7 layers can be applied to more effectively search and use content. So, for example, a picture of a tree in a field in a bright sunny day may be defined within the metadata model as:

[Video][Visual][Still][1FPS; 640x480;RGB]{[Green][Tree][Blue][Sky][Green][Grass]}[Peaceful, Calming, Tranquil, Boring]

Another example is an interview between two people FRED and JOHN about the weather at a party

[Audio][Background noise filtered, 50Hz – 4KHz, speech][continuous time][44.1KHz, 16-bit, 10 seconds]{[Fred]["Nice weather today yes?"][George]["Nah, it's raining in Plymouth"][Music]}[Crowded, Noisy]

These 7 metadata levels provide a way that makes it easy to search for content using all possible levels of representation. For example, it may be necessary that a user does a search for "video of Fred talking about weather" and it will look through each layer to find the details:

QUERY : *, *, *, VISUAL, *, VIDEO, *, Fred, weather,

III. CONCLUSIONS

This paper suggests a general way to look at and create metadata and has described a general extended description of 7 layers of metadata content from highly physical and mechanical to those requiring subjective interpretation and evaluation. These 7 metadata layers provide an easy way to search for content using all possible levels of representation and are less complex and other approaches [5]. It is clear that none of the metadata standards that are already in existence meet the necessary criteria for consumer applications because they all fail to meet all four of the specifications of scalability, flexibility, upgradability and simplicity, including provisions for the 7 levels of metadata described earlier.

This model is scalable in that it can deal with unlimited volumes of metadata of increasing complexity. It is flexible in that allows the creation of metadata to describe every conceivable type of content. It is upgradeable in that it can expand its description within the seven layers to meet future metadata requirements. It is simple in that it provides a straightforward breakdown of the various types of information that we can know about in a comprehensible way.

REFERENCES

[1] Thomas M. Coughlin, Digital Storage in Consumer Electronics: The Essential Guide, Newnes Press, March 2008

[2] Frank Nack, "You Must Remember This," *IEEE MultiMedia*, vol. 12, no. 1, pp. 4-7, Jan. 2005

[3] S. Linfoot, T. Coughlin, J. Cowell, "A Need for Improved Standardization of Metadata for Consumer Devices", 14th IEEE International Symposium on Consumer Electronics, Kyoto, 2009

[4] X.200 : Information technology - Open Systems Interconnection - Basic Reference Model: The basic model, International Telecommunications Union

[5] iXML-Product Metadata Interchange, Institute of Broadcast Sound standard 080411 (draft)