Metadata – a matter of process awareness
By Frank Nack, CWI

Over the last 15 years a great deal of multimedia research has been directed towards the development of computer environments that seek to interpret, manipulate or generate audio-visual media either in a manual, semi-automatic, or automatic way. The two major methodologies used in this type of research focussed either on particular intrinsic aspects of a media that the user wished to represent or concentrated on a particular process that can be performed on or with the investigated media. The steady infiltration of those technological advances in everyday media environments, such as image editing tools (Photoshop, Illustrator, GIMP, or Maya), audio systems such as Cubase VST, new media authoring tools, such as Director/Shockwave, Flash, Dreamweaver, Frontpage, and WWW presentation technology, such as HTML and SMIL, deeply changed the social way of exchanging information. The result was a dramatically increasing amount of multimedia-based information, which in turn has shaped the research within multimedia towards focussing on automatic ways to index the available multimedia information in a timely and meaningful way.

Mainly machine-generated metadata, however, turned out to be problematic as it is exclusively organised around the sensory surface structures of media, i.e. the physical features of an image, video or audio stream. Multimedia encoding addresses the conditions for effective perception and recognition, including systems of optional variants (prosodic features) and forms conditions for perception, such as relationships between object and background, contrast in light, or geometrical proportions. The presentations are not perfect, as the obstruction known as the sensory data gap [1] illustrates. The sensory data gap expresses the difference between the human perception of bit-wise different data arrays as the same object, despite differences in lightning, angle of view or shadow, and the automatic analysis of the machine where coded and numerical information stands for one interpretation.

More importantly, however, is the fact that low-level features, which we are able to compute for media indexing and search, merely only grant access to the representation of conceptual items but their semantics. This problem is well known as the semantic gap. In art, for example, we intuitively see the importance of physical features, such as texture, shape, or colour, by using them to identify styles, which is in particular helpful in cases where no metadata for in-depth interpretation is available. Nevertheless, a mere low-level description cannot provide more than an indication of what style type an image might have and it depends on the application up to which certainty it would accept such a retrieved media item.

The need for semantic-aware metadata schemata forced research to explore new ways of content representation. A lot of initiatives developed metadata standards to allow machines as well as humans to access the semantics of media items, such as Dublin Core, the Art and Architecture Thesaurus (AAT) by the J. Paul Getty Trust, the semantic web activity of the W^3C and ISO's MPEG-7 and MPEG-21. Yet, each of these standards still regards the process of attributing metadata to a media item as a "batch process", where the metadata is computed and assigned to the media item and then the complete media repository is published for use. This single-pass process does not reflect the continuous process of interpreting and understanding the concepts expressed through the media item on a syntactic and semantic level. As a result, the large numbers of instantiated static schemata sooner or later add to the problem they once were established to solve – namely to handle the ever faster growing amount of media-based information. In fact, the already provided metadata that allows a machine to access and manage multimedia information shows first signs of similar exhaustion, as the complexity of description schemata is growing and with it the amount of generated data. Moreover, as the aim is a mere automatic annotation and interpretation it seems that the generated metadata is not applicable to human information needs.

Recently, however, it became apparent that a more holistic view on data and metadata is required as we are heading towards a cyberspace as described by William Gibson in his novel *Neuromancer* [2] and envisioned by McLuhan in his work *Understanding Media* [3]. As a result, research is shifting again towards the provision of a knowledge space that facilitates new forms of creativity, knowledge exploration and social relationships, mediated through communication networks (i.e. hypermedia, interactive games, interactive information/experience systems, and so on). Such an interactive, open and multimodal system sustains the activation of articulation powers, which in general represent parts of a semiotic continuum, where verbal, gestical, musical, iconic, graphic, or sculptural expressions form the basis of adaptive discourses.

This direction of multimedia computing, referred to as experiential computing [4,5], aims to facilitate users to gain knowledge by directly using their senses in context-based applications composed out of user adaptable event structures which usually occur in a context.

The essential basis of experiential computing is the recognition that human interpretation of media is time- and context-bound. The ordering and (re-)classification of media items is a never ending task. As a result it is the metadata generation process, rather than the original media item, that needs to be considered as the basis of the process of knowledge accumulation. Changing the view on metadata in this fundamental way results in metadata gaining value by the inclusion of references to the context in and for which it is used. Dynamic metadata, documenting the progress of interpretation and understanding of a concept, leads to the requirement for the support of process-aware metadata. This, in turn, requires novel representation technology, that establishes metadata models where metadata is aware of

- the essential structural elements and their attributes that are required to establish the procedural aspects of an information unite in a given domain, such as event and nested event structures;
- the conditions in which the transformation from information into knowledge takes place, namely transformation in context structure and nested contexts;
- the way we make use of these processes in dynamic, interactive but unpredictable knowledge elicitation environments.

Solving these issues requires that we are able to represent tasks, such as generating, restructuring, representing, resequencing, repurposing or redistributing information, and relate them to process concepts such as task rhetoric, temporality of information and knowledge, interactivity, point of view, etc. Furthermore, we have to understand the representation of context to be able to reflect the domain-related content, the process applied to it, and the environment the process is happening in. Relevant here are the paradigmatic and syntagmatic levels of context relations, which are closely related to the point of view in the representation of events. Such work forms the basis for solutions to model the relationships between context and task processes. Only then we are in the position to understand the relevance of domain and task ontologies for the representation of events. Similarly, we are also able to solve presentational issues, such as the adaptation of dynamic metadata structures to the visualization of work processes, or the relation between event rhetoric and presentation style.

The above research questions provide the means to create process-aware information environments that facilitate an information unit to reflect about its use in various contexts and thus react on a query based on its understanding about its usability to answer the request. The objective is to move the decision about suitable material from the retrieval algorithm to the material itself. The approach also facilitates the support of knowledge manipulation and generation processes in semi-automated environments, such as feedback, creativity, communication and adaptivity. This is possible because the awareness of information units stays with them once being transferred into the information-processing environment.

Communities are only interested in process-aware media-based knowledge spaces, however, if they are robust. In my opinion the main task is, therefore, to provide real world cases that show the applicability of semantic-aware technology, a requirement recently acknowledge through the latest call for participation of the ACM Multimedia conference [6].

Acknowledgement
The author wishes to thank Marc Davis for useful discussion during the development of this work.

References

1. W.M. Smeulders et al. "Content-Based Image Retrieval: The End of the Early Years," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, Dec. 2000, pp. 1349–1380.
2. Gibson, W. (1986). *Neuromancer*. Phantasia Press, 1st Phantasia Press ed. West Bloomfield.
3. McLuhan, M. (2001). Understanding Media. Routledge Classics, London and New York..
4. Jain, R. (2003). *Experiential Computing.* Communications of the ACM **46**(7): 48-55.
5. Sundaram, H. and Rikakis, T. (2003). *An Introduction to Experiential Systems*. Arts Media and Engineering Center, ASU, AME-2003-01, 2003.
6. ACM Multimedia 2004: http://www.mm2004.org/final.pdf