COGNITIVE OBJECT FORMAT

H. Castro and A. P. Alves
INESC Porto, Fac. de Eng., Universidade Porto Campus da FEUP, Rua Dr. Roberto Frias, 378 4200 - 465 Porto, Portugal

Keywords: MPEG-21, MPEG-7, DID, Semantic Web, Internet, Metadata, Information Object, Symbol, Concept, Sensation, Perception, Comprehension, Cognition, Brain.

Abstract: The amount of on-line information content is growing without apparent limits. The lack of a coherent and consistent structure for its expression leads to increasing problems in terms of desired information retrieval and rendering. Multiple initiatives have been undertaken to bring forth such a global coherence. Nonetheless, it is still unattained. The informational landscape is highly fragmented in terms of the formats of the information object (IO) and their semantic interconnection, which is still incipient. This work exploits a loose and common sense based analogy between the Internet and the brain for the development of a new and versatile, MPEG 21 based data structuring format (termed Cognitive Object Format), for the description of information objects, equating them to cerebral memories. The objective is to enable an easier and more pervasive human (machine aided) or automatic interpretation and access to IO and their meanings in order to contribute to the development of a coherent base for their declaration and structuring.

1 INTRODUCTION

The part of living organisms that handles sensing, interpretation and decision on actions upon reality is the nervous system. In the more complex organisms, it includes a brain that is a composed and heterogeneous structure, whose development was performed gradually, in accordance with the dictates of natural evolution. The first nervous systems began as mere decentralized agglomerates of sensory and nervous cells. The present state was only reached with time (Sanes et all, 2006).

In spite of its heterogeneity and concurrent nature, the brain's global operation is coherent and integrated. It is a machine that senses and operates upon the surrounding world, based on an extensive processing and internal exchange of information.

The Internet is a greatly distributed and concurrent system as well. Sensing, interpretations and actions upon reality also take place within the Internet. Still, its integration and coherence level is largely inferior.

The Internet has also endured an evolution. Initially, it was a decentralized set of structures for the exchange of information between peer machines (ARPANET (DARPA), X.25 (ITU-T, 2009), Fidonet (FidoNet Web Site, 2009) and UUCP (Proj. Web site 2009)). It then evolved to attain interoperability between networks (use of TCP/IP) (O'Regan, 2008). This process is the equivalent to the integration of different nervous cell agglomerates.

In its next phase, the DNS system was added in order to provide a scalable way of finding and organizing on-line resources. Later, HTML, a network-based hypertext tool was developed, along with its transfer protocol (HTTP). These progresses may be equated to a joint evolution of a nervous tissue that growingly develops centralized coordinating capabilities.

The continuous increase of computers' and data transmission capabilities are also among the evolutions that the Internet has undergone. They have allowed it to sustain growingly complex interactions and the manipulation and exchange of ever richer information objects (volumes of contextually coherent and interpretable digital information). Thus, just like the nervous system, it also has evolved in the direction of a growing complexity and an increasing capacity of integration and interpretation of reality

For all this, the Internet is comparable to a loosely coupled, distributed cerebral tissue, where each computer is a coherent fragment of it. The data transfer technology is a more voluble equivalent to

Castro H. and P. Alves A. (2009).
COGNITIVE OBJECT FORMAT.
In *Proceedings of the International Conference on Knowledge Engineering and Ontology Development*, pages 351-358

the axons (information delivering slender projection of neurons). The applications running on top of that structure, implement intra or inter-computer interaction patterns, which are equivalent to the different interaction patterns supported by neuronal assemblies (intimately related sets of neurons). The hardware and software provisions interfacing with human users, are equivalent to Internet's "sensory organs".

Under the present analogy, the information objects (IO) exchanged between the different parts of a cerebral tissue may be equated to representations of sensations, or of signifying cerebral connections. The first are IOs that result from the primary storage of sensory stimulae received from the world (e.g. a real (non-synthetic) video file is a record of a sensing event of a specific aspect of reality). The later are IOs that store information with intrinsic meaning within the cerebral system (symbols connecting sensations to concepts of the brain's conceptual "tissue"). They thus endow sensations with meaning.

In face of what has thus far been exposed, the ongoing evolution of the Internet appears to be suggestive of the development of a distributed nervous tissue that progressively acquires superior global capacities for the processing, storage and coherent internal exchange of information, as well as it develops a greater level of integration, central coordination, and sensation/interaction with the outer world. At the present moment, that tissue may be considered to be at a development level comparable to the primordial stages of its biological counterparts. These growing parallels indicate that an approach based on the cerebral-cognitive operation for the description of information objects is advantageous and well prepared to deal with the predictable evolution of the web, paving the way for greater future developments. The fact that the principal interacting agents on the Internet are human beings is also something that is advantageously handled by this approach.

We do not claim that this approach will result in an immediately simpler or faster Internet operation. Oppositely, an immediate extra-burden, on all its entities is to be expected. The advantages are visible only in the greater picture of Internet's overall operation. The continuous increase of its data transport and processing capacities will render this brain-oriented migration possible and even probable, and make negligible the burden of the greater technical responsibilities deposited on the Internet's constituting provisions by our proposed approach.

To contribute to this evolution we develop an analysis of the process through which the brain goes from the sensing of materiality to the detection of patterns and shapes in them, to the interpretation and valuation of the later, to the development of concepts and signification relationships that are transversal to multiple sensations and intertwined in a global sensorial-conceptual tissue. Based on this analysis, a format was created to structure informational objects - Cognitive Object Format (COF).

2 RELATED WORK SCENE

Plenty of work has already been developed on content description and annotation. The initiatives undertaken in this area are generally divided between those oriented towards the semantic web, employing OWL (W3 Org, 2009) or other RDF (W3 Org, 2009) based ontologies, those devoted to the annotation and description (especially low-level) of multimedia content, employing XML based tools (MPEG 7, MPEG 21, etc), and those that attempt, to conciliate the first two.

The main focus of the work described in (Athanasiadis et all, 2005) is the knowledge-based automatic extraction of semantic information from multimedia content. Still, an ontology based structure is used for the expression of content describing metadata. DOLCE (Gangemi et all, 2002) is employed as the core ontology. An MPEG 7 based ontology was used to describe the low-level aspects of media content. Higher level semantics of the content were described using DOLCE based domain specific ontologies.

The works described in (García et all, 2008) and (Vembu et all, 2006) also merge high-level and low-level descriptions, where low-level characteristics of the media record are described employing an MPEG 7 based ontology, and high-level semantic aspects about the content are expressed in an RDF compliant way.

In (Bloehdorn et all, 2005) a work is presented, which was developed in close proximity to that of (Athanasiadis et all, 2005), yet with a greater focus on the formalization of the interrelationship of high and low-level multimedia concept descriptions.

In (Arndt et all, 2007) the COMM tool is defined. It reengineers the most important parts of MPEG 7, (for describing the structure and content of media items) employing the DOLCE foundational ontology.

All these works thus attempt, in an RDF oriented manner, to conciliate the two tendencies in content meta-description, by converting the audiovisual feature describing tools (namely MPEG 7) to RDF based ontologies, and inscribing the entire descriptive metadata (feature and semantic) in a global RDF based ontology.

As argued in (Stamou et all, 2006), the information contained in a multimedia document may be divided into separate layers: the sub-symbolic, symbolic and logical layers. The first represents the raw multimedia information. The second provides a structural layer on top of the binary media stream so that it is possible to further process the information, to what the third is devoted.

The mostly used standards for media descriptions (e.g. Dublin Core, MPEG-7/21, etc) generally operate on the symbolic level. This approach presents a problem as the semantics of the information expressed in such standards are implicit (to its structure and terminology), and only valid within its framework, thus impeding interoperability. This may be handled by replacing the symbolic layer with one composed of formal, machine-processable semantics, typically expressed in the RDF language (Stamou et all, 2006). Broadly, this is the approach taken by most works in the field, including those mentioned above. However, it fails to take advantage of existing XML-based metadata, and ignores the advantages of an XML-based structural layer. A purely RDF based semantic description is very general, open or variable. A tailored RDF based ontology for low-level technical media characteristic description presents an overhead when compared to the existing XML based standards. Furthermore, logically, the structural layer is not at a cognitively semantic level, but more at a perceptual one. An XML based and implicit semantic language, for the structuring of media items (e.g. MPEG 21) and their technical description (e.g. MPEG 7) is therefore more practically and logically appropriate for the symbolic level.

An alternative solution to the implicitness of the structural layer's semantics is thus to add a third layer (the logical abstraction level) that maps the structured information sources to the domain's formal and explicit knowledge representation, thus providing the semantics for the symbolic level.

The work presented here is in line with this later approach. For the middle layer, we employ MPEG 21 for the overall structuring and relating of information objects and MPEG 7 (structurally

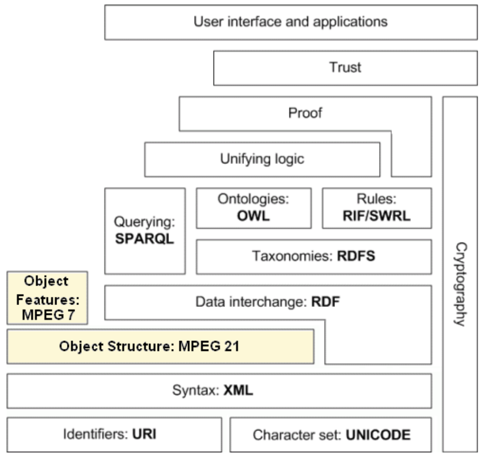

Figure 1: Revised Semantic Web Stack.

contained within the MPEG 21 body), to describe substructures within the media objects and their low-level perceptual characteristics. The logical abstraction level may employ any number of explicit semantics annotation tools (RDF based).

This proposal therefore implies a change in the semantic web stack, (depicted in Figure 11). Its base would effectively become MPEG 21+MPEG 7+ RDF, instead of only RDF. It is a radical but useful change. The relative rigidity and implicit semantics of the symbolic layer tools are an advantage as they provide simplicity and efficiency. All necessary semantic explicitness is added by the third layer. An optimal trade-off is thus achieved between simplicity, logical correctness and accuracy. The approach is taken even further by basing it on a broad view of the human perception structure.

3 COF LOGIC AND STRUCTURE

To develop a logic and a structure for informational objects that is inspired on the cerebral operation, it is necessary to elaborate on the manner in which this structure apprehends reality and coherently stores valuating information about it. This process is still relatively unknown, and fairly beyond the skills of the authors. This analogy is thus based on a present, common sense view of that process.

For the context of this study, the authors considered that the apprehension of reality is divided into three main parts: sensation (sampling of reality by the sensory organs), perception (processing of sensory samples and further structural interpretative elaboration upon them) and comprehension (valuation of reality).

3.1 Sensation Level

The stimulae resulting from the sensing of materiality are passed, in specific formats, to the appropriate cortex, submitted to processing, and storage. Thus, the basic registers are created.

Equivalently, the sensory structures of our reality sampling devices (e.g. camera) also perform an initial capture of aspects of reality, which are subjected to pre-processing, specific encoding and storage. Thus, the base level of the COF information structure is that of the basic and non-signifying sensation registers, the "Sensation Objects" (SO).

Obviously, a video, a sound recording, etc., may not be devoid of symbolic value. Still, in a brain, that signification relationship only exits as an association that follows sensation.

3.2 Perception Level

Perception consists of the cerebral processing that is performed over the sensorial information occurring just above sensation, but not yet at a meaningful level. It includes:

- the basic perception of the functional aspects, of the space-time structuring of sensations, and of the space-time relations between sensations;
- the apprehension of basic features of the sensed materialities;
- the laying of the perceptual foundations for the construction of concepts.

For simplicity, in the context of the COF, it is considered that those activities are functionally isolable from sensations and procedurally posterior to them. Therefore, the Perception Objects (PO) are above the SOs. They are the equivalent to a crystallization of the abovementioned phase of apprehending reality into a static description.

POs consist of one ore more SOs and also one interpretative information carrying object for each of the types of perceptual processing mentioned above.

The Functional and Space-Time Characterization Object (FSTCO) carries the metadata describing the functional aspects of the capture and register of the sensory data and the space-time relations between the SOs.

The Conceptualization Root Object (CRO) carries the metadata that divides the sensation according to the most relevant shapes and patterns (see Figure 2). These delineations of objective

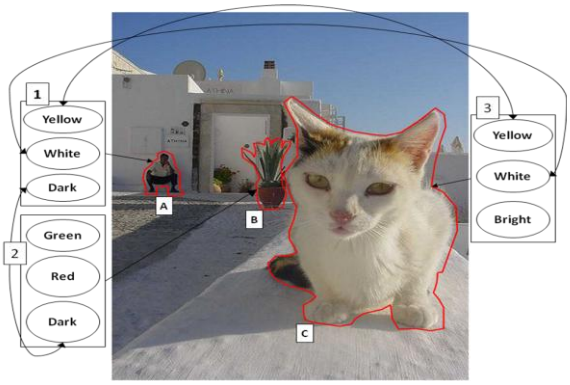

Figure 2: Perception of a Visual Sensation.

bodies are the roots of concepts.

The Base Characteristics Object (BCO) carries metadata identifying the global sensation's relevant characteristics and those of each of its sub-portions (CRO defined). This information is thus bound to the CRO (presented by the linking of tags 1, 2, and 3 to divisions A, B and C in Figure 22), or to its corresponding sub-objects.

The POs are of two different types: simple POs (SPO) containing SOs, FSTCO, CRO and BCO; and composed POs (CPO) containing other PO, FSTCO and CRO.

Given that visual or auditory perception, are different processes in the brain, different types of sensations must be contained in different SPOs.

3.3 Comprehension Level

Comprehension corresponds to the valuation and conceptual-symbolic interpretation of "inferior" sensorial-perceptual constructions. In the COF, the Comprehension Objects (CO) are above the POs. They are the crystallization of the semantic comprehension process into a static description. This description, which is pure meaning in the context of the COF, is contained in the "Semantic Objects" (SmO). The SmO thus expresses relationships between sensation records and concepts through the use of (written) symbols.

Each CO may carry one or multiple POs or sub-Cos, exclusively. They will also contain a CRO and SmOs. The CRO defines the global concept roots, based on those of all PO or CO children. The SmOs play different roles in the COs:

- an SmO may be associated to the CRO (or one of its segments), performing the connection of the POs (whose CROs are pointed to by the global CRO) to the conceptual-symbolic fabric, within a specific context, endowing them with meaning;

- an SmO may be associated to a CO, expressing the "positioning" of the CO in the global conceptual structure. It also expresses the semantic relations between the CO and its sub-objects. This is comparable to a cerebral process of reflection on other "mental objects";
- an SmO may also be associated to another SmO. It performs the contextualization of the concepts expressed in the target SmO.

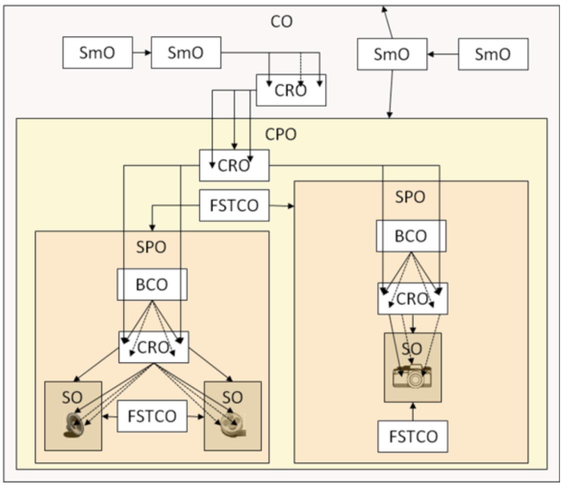

Figure 3: COF Structure Example Overview.

The CO may also contain solely SmOs. These COs correspond to "cerebral IOs" devoid of immediately associated sensation-perceptions. They may be viewed as a "thought" over other IOs (CO).

Information on intellectual rights over info objects may be viewed as such a "thought". For this, in the COF structure, the expression of intellectual rights over IOs will be made with COs carrying only SmOs that contain rights expression metadata, in accordance with a precise standard.

Figure 33 presents an overview of some of the possible structures of COF objects.

4 COF'S STANDARD DATA FORMAT

4.1 Standards and Tools

The selected standards for the construction of the COF structure are MPEG-21 (Chiariglione, 2002) (parts 2 (ISO/IEC FDIS 21000-2, 2005) (DID), 3 (ISO/IEC FDIS 21000-3, 2005) (DII), 5 (ISO/IEC FDIS 21000-5,2006) (REL), 15 (ISO/IEC FDIS 21000-15, 2006) (ER) and 17 (ISO/IEC FDIS 21000-17, 2006) (Fragment ID)) and MPEG-7 (Chiariglione, 2004). MPEG-21 is used in the overall structure of COF objects and for other varied purposes. MPEG-7 tools are used to segment, characterize and provide meaning to the sensation-perception objects.

4.2 Data Format

4.2.1 Sensation Objects

The SOs consist only of MPEG-21 DID and DII metadata and, possibly, of the raw inline media content. Each SO, is represented by a *did:Item* element within a superior structure (the PO). That *did:Item* contains a *did:Descriptor* where a system wide identifier (wrapped in a DII structure) is present, as well as a *did:Component*, that encloses the media content itself or a URL referencing the location of the "sensation data" within its child *did:Resource* element.

4.2.2 Perception Objects

POs encapsulate the SOs. Each PO is represented by a *did:Item* element and contains one or more SO. The PO also carries an FSTCO, a CRO and a BCO.

The FSTCO is represented by a *did:Item* element carrying a series of *did:Annotation* elements. The *did:Annotations* specify (within a *did:Descriptor*), functional information regarding the capturing and recording of the sensory stimulae, as well as information describing the space-time relations between the different sensations carried in the PO. Ergo, in each FSTCO, there is:

- a *did:Annotation* bound to each SO, specifying its encoding format. MPEG-7 part 5 (Multimedia Description Schemes) tools are used for this (MediaFormat D);
- a *did:Annotation* bound to each segment of each SO, which specifies the sensation capture device and any relevant setting related to the capture episode. MPEG-7 part 5 tools are employed for this (Creation DS);
- a *did:Annotation* bound to each fragment of each SO, specifying the positioning of its capture point within an xzy relative (to an arbitrated centre of the overall sensational event) axis system as a function of time. Custom metadata were developed for this purpose.

For the time segmentation of the SO resource, MPEG-7 tools are employed for the definition of media object time segments (VideoSegment DS and AudioSegment DS).

The CRO is represented by a *did:Item* element that carries a series of *did:Annotation* elements. These perform the logical role of attributing information to some target information body. The target body’s (an xml element or portion) ID is specified in the *did:Annotation* target attribute. The *did:Annotations* are logically divided into two levels:

- the base *did:Annotations* bind (using a *did:Anchor* and *did:Fragment* element sequence) an identifying *did:Descriptor* to a specific segment of the sensorial record (data resource). This *did:Descriptor* is logically attributed to the *did:Resource* element of the targeted *did:Component* of a specific *did:Item* element. They are thus bound to perceived “objects” within the SO (represented by *did:Item* elements);
- the top level *did:Annotations* are bound to several base level ones (via the *target* attribute), so as to logically unite different “object” perceptions in different SOs into one single multi-SO “object” perception. The bound information is again identifying data.

Figure 44 exemplifies the employment of the mentioned sequence for the delimitation and identification of a perceived “object” in an image fragment (StillRegion DS tools used). Each *did:Annotation* thus delimits a perceived “object” that plays the role of a concept root.

MPEG-7 part 3 (Visual) tools are used to shape and localize descriptions, if the SO stores a visual resource (StillRegion DS, StillRegion3D DS, MovingRegion DS VideoSegment DS). For audio resources the audio segment defining functionalities of MPEG-7 are used (AudioSegment DS).

The BCO is represented by a *did:Item* element that carries a series of *did:Annotation* elements. Each *did:Annotation* binds low-level feature describing information to a *did:Annotation* contained in the CRO. MPEG-7 part 3 (Visual) tools are used to describe colour texture and motion if the targeted fragment is of visual type. MPEG-7 part 4 (Audio) tools are employed if the targeted fragment is of an audio type.

```
......
<did:Annotation
id="item:perception:conceptroot:objectID_1_ro
otID_1"
target="#item:sensation:objectID_1_componentI
D_1">
   <did:Descriptor>
     <did:Statement mimeType=" text/uri-
        list">
item:sensation:objectID_1_ componentID_1_xyz
     </did:Statement>
   </did:Descriptor>
   <did:Anchor>
      <did:Fragment>
         <mpeg7:Mpeg7>
            <mpeg7:Description
         xsi:type="mpeg7:ContentEntityType">
               <mpeg7:MultimediaContent
                 xsi:type="mpeg7:ImageType">
                  <mpeg7:Image>
                     <mpeg7:SpatialLocator>
                        <mpeg7:Polygon>
                           <mpeg7:Coords
                            mpeg7:dim="2 5">
            5 25 10 20 15 15 10 10 5 15
                           </mpeg7:Coords>
                        </mpeg7:Polygon>
                     </mpeg7:SpatialLocator>
                  </mpeg7:Image>
               </mpeg7:MultimediaContent>
            </mpeg7:Description>
         </mpeg7:Mpeg7>
      </did:Fragment>
   </did:Anchor>
</did:Annotation>
......
```

Figure 4: XML Snippet with Annotation, Anchor and Fragment Sequence.

A number of differences occur if the sensed content is of a textual nature, and already stored as such (not as an image for instance). Such an information object is not as closely related to materiality as its visual or audio counterparts. No low-level features are extracted and thus there is no BCO. The textual content must be marked up according to some XML mark-up language. The CRO specifies the different concept roots similarly to what has been explained above. However, the XML Pointer Language is employed in the *did:Annotations*’ *target* attribute instead of MPEG-7 tools.

The FSTCO only specifies the coding format employed in the record of the textual object. In this case, the fragment specification is carried out with the XML Pointer language within the value of the *did:Annotations*’ target attribute.

A PO may also contain sub-POs instead of SOs. In this case, the PO’s structure is the same as the structure described above, except that it has no BCO. Its FSTCO only specifies the space-time relation between the sub-POs, by binding the previously mentioned data to the appropriate time fragments of the sensorial resources inside the SO of the PO. The CRO will add further conceptual root defining *did:Annotations* to the inferior CROs.

4.2.3 Comprehension and Semantic Objects

The COs encapsulate the POs. Each CO is represented by a *did:Item* element and contains one or more PO and a global CRO object. The CO also carries a series of SmO with different purposes. The CRO object binds concept roots from different PO into joint concept roots.

The SmO are represented by a *did:Item* element that contains a series of *did:Annotations*. These *did:Annotations* bind semantic data to the concept roots defined within the global CRO. There are three types of SmO:

- the first type endows the PO (its concept roots) with human interpretable meaning within some context;
- the second type specifies information and semantic relations about and between different POs or SOs;
- the third type specifies the context in which the signification defined by the first two types of SmOs is valid.

Given the diversity of information that may be expressed in the SmO and the vastness and specificity of the possible contexts for that information, many different standards and protocols may be used to express it. In this work, we propose only a few basic characteristics for that information and some tools to express it.

The first type of SmO should, if contextually possible, answer questions, such as who, which object, which action, where, when, why and how, and add other relevant commentaries. This information is then bound to a concept root. MPEG-7 part 5 (Semantic DS or SemanticBase DS) based ontologies or others may be employed for this descriptive purpose.

The second type of SmO employs a custom ontology. It specifies all the POs, or sub Cos, as the objects of semantic relationships, which are in their turn described as well. For instance, a textual content carrying PO may be declared as a caption or summary for a visual content carrying PO.

The third type of SmO includes the first two types. The tools employed in the SmO of the first type may be used here as well in order to specify the context of the targeted semantic *did:Annotations*.

As previously explained, there are also COs dedicated to the expression of intellectual rights information over other COs. These are also represented by *did:Item* elements containing only SmOs. The SmOs in question are also represented by a *did:Item* element that contains *did:Annotations*. Specifically, each of these CO will have two SmO:

- the first SmO carries MPEG-21 REL metadata to express intellectual rights. It binds such data to a CO via its *did:Annotation's* target attribute;
- the other SmO points to the first SmO.

When the CO carries sub-COs instead of POs, there are no differences in the rest of its internal structure.

Each type of CO will be contained in its own separate MPEG-21 DID.

5 COF USAGE

The COF provides a decoupling between structural and low level technical description and high level semantic valuation of information objects. The provision of such a comprehensive and malleable description base allows a simple and efficient declaration, delineation, base characterization and logical structuring of information objects, with MPEG 21 and MPEG 7. The description of higher level features with RDF oriented tools enables a very versatile and precise way of attributing meaning to information objects.

For all this the COF structure may be employed in the declaration, structuring and semantic describing and interrelating of a myriad of different types of information constructs (audiovisual info objects, combined textual and audio visual, etc), which are to be delivered in an integral manner, as opposed to a fragmented or streamed manner.

6 CONCLUSIONS

This work defines a format to structure and describe information that is inspired, in a broad sense, by the way that the brain apprehends reality.

Other developments have been undertaken in the field of content structuring and annotation, semantic web (Berners-Lee et all, 2001) and web integration and interoperability. Still, the present work merges the content structuring and technically describing information with the semantically describing information, within an integrated approach upon the nature of reality cognition. It builds on an analogy between the Internet and the brain, structuring info objects in a way that, makes it easier for both (computer aided) human and automatic means to understand them, and opens the way for the expression of increasingly complex information objects in a “semantically enabled” manner. That understanding is eased by the structuring, standardization and uniformization effects of the employment of the MPEG 21 protocol at the base of the semantic web stack, as well as by the use of the comprehensive and precise MPEG 7 protocol to describe “pre-semantic” aspects of media objects.

Creating, maintaining and using COF content descriptions may lead to a considerable overhead, when compared to the sole maintenance of raw data objects. Still, if compared to the majority of existing- or under development - description structures, the presented costs are reasonably the same. Furthermore, having the definition of COF being based on two powerful open standards, such as MPEG-21 and MPEG-7, facilitates its employment, expansion and conversion into other protocols.

Some possible future work is the replacement by RDF or OLW based domain specific ontologies, of the custom metadata tools developed for tasks not covered by MPEG 7 (space-time relative positioning of SO objects within a PO, structuring textual SO, expressing semantic functional relations between informational objects (PO and SO) as informational objects, expressing high-level semantic information).

REFERENCES

DARPA Home Page, http://www.darpa.mil, retrieved 05/03/2009.

Sanes, D. H., Reh, T. A., Harris, W. A., 2006. “Development of the Nervous System”. Elsevier Academic Press.

FidoNet Official Web Site, http://www.fidonet.org/, retrieved 05/03/2009.

O'Regan, G., 2008. “A Brief History of Computing”. Springer.

ITU-T’s page on X.25, http://www.itu.int/rec/T-REC-X.25-199610-I/en, retrieved 05/03/2009.

Mosaic Web Browser’s page at NCSA, http://www.ncsa.uiuc.edu/Projects/mosaic.html, retrieved 05/03/2009.

MPEG-7 Web Page at Leonardo Chiariglione’s Web Site, “MPEG-7 Overview”, http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm, 2004, retrieved 08/03/2009.

MPEG-21, 2005. “ISO/IEC FDIS 21000-2:2005(E) MPEG-21 - Part 2: Digital Item Declaration”.

MPEG-21, 2005. “ISO/IEC FDIS 21000-3:2005(E) MPEG-21 - Part 3: Digital Item Identification”.

MPEG-21, 2006, “ISO/IEC FDIS 21000-5:2006(E) MPEG-21 - Part 5: Rights Expression Language”.

MPEG-21, 2006. “ISO/IEC FDIS 21000-15:2006(E) MPEG-21 - Part 15: Event Reporting”.

MPEG-21, 2006. “ISO/IEC FDIS 21000-17:2006(E) MPEG-21 - Part 17: Fragment Identification of MPEG Resources”.

MPEG-21 Web Page at Leonardo Chiariglione’s Web Site, “MPEG-21 Overview v.5”, http://www.chiariglione.org/mpeg/standards/mpeg-21/mpeg-21.htm, 2002, retrieved 08/03/2009.

Berners-Lee, T., Hendler, J., and Lassila, O., 2001. “The semantic web: A new form of web that is meaningful to computers will unleash a revolution of new possibilities”. Scientific American.

UUCP Project Web Site, http://www.uucp.org/info.shtml, retrieved 05/03/2009.

W3C’s OWL specification, http://www.w3.org/TR/owl-features.

W3C’s RDF specification, http://www.w3.org/RDF.

Athanasiadis, T., Tzouvaras, V., Petridis, K., Precioso, F., Avrithis, Y. and Kompatsiaris, Y., 2005. “Using a Multimedia Ontology Infrastructure for Semantic Annotation of Multimedia Content”. Proc. of 5th International Workshop on Knowledge Markup and Semantic Annotation.

Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., and Schneider, L., 2002. “Sweetening Ontologies with DOLCE”. Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web. Proceedings of the 13th International Conference on Knowledge Acquisition, Modeling and Management.

García, R., Tsinaraki, C., Celma, O., Christodoulakis, S., 2008. "Multimedia Content Description using Semantic Web Languages". In Y. Kompatsiaris, P. Hobson, (Eds.) Semantic Multimedia and Ontologies: Theory and Applications, pp. 17-54. Springer.

Bloehdorn, S., Petridis, K., Saathoff, C., Simou, N., Tzouvaras, V., Avrithis, Y., Handschuh, S., Kompatsiaris, I., Staab, S., and Strintzis, M., 2005. “Semantic annotation of images and videos for multimedia analysis”. Proc. of European Semantic Web Conference.

Vembu, S., Kiesel, M., Sintek, M., and Bauman, S., 2006. "Towards bridging the semantic gap in multimedia annotation and retrieval". Proc. of First International Workshop on Semantic Web Annotations for Multimedia.

Stamou, G., Ossenbruggen, J., Pan, J. Z., Schreiber, G., 2006. ”Multimedia Annotations on the Semantic Web”. IEEE MultiMedia, v.13 n.1, p.86-90.

Arndt, R., Troncy, R., Staab, S., Hardman, L., and Vacura, M., 2007. “COMM: Designing a Well-Founded Multimedia Ontology for the Web”. In 6th International Semantic Web Conference.