



AsIsKnown

**A semantic-based knowledge flow system for
the European home textiles industry**

Work package 5: Multimedia ontology

**Deliverable D16 “Report on the state of the art for the design
methods of multimedia ontology”**

Lead participant: IPP-BAS
Nature: Report
Dissemination level: PU
Delivery date: 12 PM



This document has been produced in the context of the AsIsKnown Project. The AsIsKnown project is part of the European Community's Sixth Framework Program for research and development and is as such funded by the European Commission. All information in this document is provided "as is" and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability. For the avoidance of all doubts, the European Commission has no liability in respect of this document, which is merely representing the authors view.

Context

WP 5	Multimedia ontology
Task 5.1	Analysis of state-of-the-art
Dependencies	This deliverable reuses user requirement input from WP 3

Contributors: Kiril Simov (IPP-BAS) Petya Osenova (IPP-BAS) Anelia Tincheva (IPP-BAS)	Reviewers: Kiril Simov (IPP-BAS)
---	--

Approved by: Kiril Simov, Bulgaria as WP5 Head

Executive Summary

This report focuses on the state-of-the-art of design methods and requirements in the area of multimedia ontologies. Such a task is closely related to the AsIsKnown Home Textile Ontology, because it encodes other representational views of the same concepts, and it adheres to the same ontology requirements. For the aims of the project our main target objects will be the static pictures. It means that we will cover two top-level hierarchical types: images and 3D graphics.

The annotation of these multimedia objects is envisaged to be performed at two levels: lower and higher. The former refers to the picture components (e.g. pixels), and the latter views the pictures as a whole.

The role of the multimedia ontology is described with the help of three specific use cases - Visual counterpart of the concept in the domain ontology; Interior Designing Search of Image and Trend Analyzer Project Data Mining. It turns out that the multimedia ontology will be used mainly in three tasks: Annotation, Analysis and Retrieval.

The suggested approach towards the creation of a multimedia ontology follows the methodology, which was established for the AsIsKnown Home Textile Ontology.

Table of Contents

- EXECUTIVE SUMMARY 4**
- TABLE OF CONTENTS 5**
- LIST OF ABBREVIATIONS 6**
- 1 INTRODUCTION AND PROBLEM STATEMENT 7**
- 2 MULTIMEDIA ONTOLOGY USE CASES WITHIN ASISKOWN 8**
- 3 MULTIMEDIA ONTOLOGY REQUIREMENTS 10**
- 4 EXISTING MULTIMEDIA ONTOLOGIES 14**
- 5 CONCLUSIONS 16**
- REFERENCES 17**

List of Abbreviations

AIK-HTO	AsIsKnown Home Textile Ontology
AIK-MO	AsIsKnown Multimedia Ontology
MPEG7	Moving Picture Experts Group - 7
MPEG21	Moving Picture Experts Group - 21
DOLCE	Descriptive Ontology for Linguistic and Cognitive Engineering
OWL	Web Ontology Language
OWL-DL	Web Ontology Language – Description Logic sublanguage
SUMO	Suggested Upper Merged Ontology
XML	Extensible mark-up language

1 Introduction and Problem Statement

In this deliverable we discuss the scope of the AsIsKnown Multimedia Ontology and the methodology for its creation. The AsIsKnown Multimedia Ontology (AIK-MO) will be used as an additional vocabulary within AsIsKnown system which will facilitate the annotation of multimedia documents with content information based on the AsIsKnown Home Textile Ontology (AIK-HTO). Thus the concepts and relations within AIK-MO will be used to describe elements of multimedia images and these elements will be identified as instances of concepts and relations within AIK-HTO. The annotations will be done on two levels – lower level descriptions, where lower level image elements (identified outside AIK-MO and with mechanisms outside AsIsKnown System with the possible exception of Virtual Interior Designer) are identified as an instance of a concept or a relation within AIK-MO; and higher level descriptions where the image objects are not directly annotated in the image representation, but they might be presented in several ways: (1) named and described within AIK-MO related ontology repository; (2) described by a human annotator; (3) with a help of an NLP module which analyzes the text related to the image; (4) by descriptions which chosen among the lower level descriptions. Mixing ontology descriptions from different sources is a possible cause for contradictory information. For the current use within the AsIsKnown system we envisage to keep information about the source of a given ontology description and use this information in the process of description application. Another option that will be provided is as follows: the human intervention resolves the contradictory cases.

As a source of knowledge about the content of the AIK-MO we will use (similarly to AIK-HTO) mainly standardized multimedia descriptions. The main sources of information in this direction are different standards like MPEG7, MPEG21, etc. These standards are already used in various different multimedia ontologies constructed with different tasks in mind. The actual construction of AIK-MO will be done by examining of these sources and combining them. The final version of AIK-MO will be produced on the bases of testing of the earlier versions within AsIsKnown project.

The rest of the report is organized as follows: first, the envisaged usage of the multimedia ontology within AsIsKnown system is discussed; then a short overview of the existing sources of knowledge for the multimedia ontology are presented; finally, a summary of the principles that will be followed in the construction of AIK-MO is provided.

2 Multimedia Ontology Use Cases within AsIsKnown

There are two places where pictures play main role in the AsIsKnown System – Virtual Interior Designer and Trend Analyser. The result of the usage of the Virtual Interior Designer is an interior design which is represented as a set of multimedia documents describing the elements of the design, their features and relations. A crucial part of the multimedia representation of an interior is a set of pictures and plans. Most of the pictures will be stored within the multimedia database of the AsIsKnown system (see Deliverable D4). The role of pictures is important with respect to trend research supported by the Trend Analyser. Here one of the main tasks is the semantic annotation of multimedia documents. This semantic annotation is later used to support data mining for trend analyses. The semantic annotation will comprise semi-automatic semantic annotation of text and manual annotation of graphical information (depending on the source of the multimedia documents). In both cases image information will be retrieved on the basis of content information and spatial relations between elements of interiors. In both cases the multimedia ontology needs to support descriptions of image elements as objects (higher level) with certain properties like rectangular object, oval object, etc.; their location relatively to the picture boundaries and to each other, their properties like orientation, topological features, visibility, etc. Such object descriptions will be connected to concepts and relations from the AIK-HTO. For example, the rectangular object within the bottom part of the picture is a “HandKnottedPileCarpet” from “NaturalFibre” which is partially covered by an oval object. Such descriptions will be done manually by an annotator, or they will be result of the NLP analysis of the surrounding text. Additionally, such object descriptions will be connected with lower level descriptions in the digital representation of the image like a region of pixel, etc. Such lower level descriptions could be produced when the pictures are composed by the Virtual Interior Designer from different components. Also, during the picture composition content descriptions of the multimedia objects will be added when they are predictable from the context. On the basis of all this we could imagine the following three scenarios for usage of the multimedia ontology within AsIsKnown:

Use-Case Title	Visual counterpart of the concept in the domain ontology
Principal Actor	Domain specialist
Rationale	Validation of a concept or need for better comprehension
Precondition	The domain specialist needs to check the content of the concept
Postcondition	Pairs of concepts and its image(s) are presented to the domain specialist
Description	The domain specialist formulates a query about retrieving information about a concept, i.e. formal definition, the textual representation as well as the image view.

Use-Case Title	Interior Designing Search of Image
Principal Actor	Interior Designer
Rationale	New design is under development
Precondition	An image of a particular object (carpet, curtain, etc.) or of a ready interior needs to be incorporated in an interior design
Postcondition	The appropriate image is found in the multimedia database, resized and incorporated in the representation of the design under development.
Description	The interior designer formulates a query about the contents of the image which to be incorporated within the design. The query is

	<p>formulated in terms of selection of items from a menu. The menu items are terms in the lexicon of some natural language and they are connected to concepts and relations in the two ontologies. The terms are substituted by these concepts and relations. Then the query is evaluated in the database and the returned images are shown to the designer.</p> <p>When the designer incorporates an image in the design the ontology description related to the image could be saved as part of the ontology description of the multimedia document included in the design. Thus, later this information could be used by the Trend Analyzer.</p>
--	---

Use-Case Title	Trend Analyser Project Data Mining
Principal Actor	Trend creation specialist
Rationale	A trend analysis project is under development
Precondition	The trend creation specialist needs to cluster information represented in a multimedia document
Postcondition	Clusters of the found information is presented to the user
Description	The trend creation specialist formulates a trend hypothesis and wants to verify it by data mining the available information. Part of this information is multimedia documents annotated with ontological information using AIK-MO and AIK-HTO. The data mining tool returns the information matching the hypothesis definition and presents it to the user.

The three use cases impose some requirements on the content of the AIK-MO and also show the necessary coverage of the ontology. First and more important, the AIK-MO has to cover the representation of static pictures, thus we could exclude from the ontology all the multimedia concepts that are related to audio and video information. Also, one obvious (and mentioned frequently in the literature) distinction is between the image elements and their descriptions as lines, pixels, etc, and the content of the image – the real objects in the real world to which the elements of the image refer. Thus, we have to be careful about the annotation with respect to which of this information is explicated within the semantic annotation. Also, we have to be careful about the search in future, because very often the people refer to the representation of a real object in the same way as to the real object. For instance, “the carpet in the image has to be red” means that the color of the carpet represented by a given object in the image has to be red.

Additionally the two ontologies AIK-MO and AIK-HTO share some ontological information like size, colors, shapes, etc. In the end, such common ontology information will be represented only once.

3 Multimedia Ontology Requirements

In this section we present some general recommendations about the design and creation of multimedia ontologies. Then we list some of the available multimedia ontologies which will be used as a source of information for the creation of AIK-MO. The material in this section is based on [1]. This document is a very good reference point because it reflects the experience within several European projects as well as the usage of several multimedia ontologies. First we will present their requirements:

Scope and Usage

Multimedia ontologies should be designed to serve one or more of the following purposes:

- Annotation – summarization of multimedia content
- Analysis – ontology driven semantic analysis of multimedia content
- Retrieval – context-based retrieval and recommendations
- Reasoning – application of reasoning techniques to multimedia content
- Personalized filtering – delivery of multimedia content according to user preferences
- Meta-Modelling – ontologies used to model multimedia processes, procedures

The AIK-MO will be mainly used for three tasks: Annotation, Analysis and Retrieval. The annotation will be semi-manual or inferred in the process of composition of design. The analysis will be the relation between the annotations of AIK-MO instances in images to the instances in AIK-HTO. Thus we will have at least two levels of annotation of an image: (1) annotation of the instances of the image elements themselves; and (2) corresponding real world instances denoted by the instances of the first level.

Media description – Description of the Information Object

The ontologies need to describe and represent knowledge for either one or even more of the following top-level hierarchical types of multimedia documents:

- Image
- Video
- 3D Graphics
- Audip
- AudioVisual
- Multimedia presentation

As it was mentioned above, AIK-MO will cover Image and 3D Graphics (if they are used in the multimedia database). Thus, our targets are static pictures.

Representation of the Content Structure

- **Multimedia structure representation.** Multimedia ontologies should be able to represent the structure of a multimedia document itself, depending on the type of the

document and the relations between the various structural elements. A hierarchical structure of multimedia segments is thus needed in order to capture all possible types of media decompositions and relations. As it was mentioned above, in AIK-MO we envisage two levels of granularity of the representation of multimedia content: lower level and more abstract level. At the first level regions of an image or individual pixels will be modelled. The second level will identify different objects as elements of a picture and the relations between such elements. The obligatory annotation of the multimedia content will be done by using concepts and relations from the second level of multimedia concepts and relations. The first level of multimedia concepts and relations will be used in order to construct objects at the second level of modelling of the multimedia content.

- **Spatiotemporal Relation Representation.** The isolated description of audiovisual characteristics seems to be an inadequate representation of a multimedia document. Spatial relations are needed to describe how segments are placed and related to each other in 2D space. Topological spatial relations are needed to describe how the spatial boundaries of the segments relate (e.g. touches and overlaps). Temporal relationships among segments or events, providing information about the sequence in time, need to be represented, especially when the multimedia object is of type video. As it was mentioned above, some spatiotemporal relations over multimedia objects are closely related to the content structure of those objects – the real world objects denoted by the multimedia object. Thus, spatiotemporal relations will be shared between AIK-MO and AIK-HTO. Because we will work with static multimedia we do not envisage inclusion of temporal relations, but if it becomes necessary in the process of usage, we will add them later on.
- **Localization.** The regions that correspond to semantic objects need to be described in terms of their location within the multimedia content. In AIK-MO this will be represented on the low-level of granularity via concrete media descriptors and on the abstract level in a relative manner – localization in relation to boundaries of the picture and the general scene.

Representation of the Content Description

- **Low-Level Description Representation.** The ontologies should model concepts and properties that describe audiovisual characteristics of objects, especially audio segments and visualizations in still images and videos, in terms of low-level features. In order to semantically express this kind of characteristics associated with a concept, several different audiovisual attributes (e.g. color, shape, texture, motion, localization etc) need to be represented, depending on the concept. This level will be exactly one from the levels of granularity represented in AIK-MO.
- **Support of Multiple Audiovisual Descriptions.** Audiovisual characteristics of domain concepts cannot be described using one single instance of the audiovisual descriptors in question. For example, while the net of a tennis racket might be described in terms of its texture only once, its shape heavily depends on the viewing angle and occlusions (e.g. by the player in front of the net). The required conceptualisation thus has to provide means for multiple prototypical descriptions of a domain concept. In the case of AIK-MO we will have the necessary concepts to describe each view, but for the moment we do not envisage to represent relations between descriptions of objects that in different images. The only way to express that the objects are the same will be the usage of the same URI for the same object.

- **Alignment to existing metadata standards.** For reasons of interoperability and sharing, the need for alignment to existing metadata standards grows higher. We will follow this requirement 100 % in the creation of AIK-MO. The development of AIK-MO will be based on already existing multimedia ontologies and standards like MPEG-7 ([2]).
- **Support for basic and complex data types representation.** Ontologies aligned to standards built on XML Schema, such as MPEG-7 or others, would have to embody the representation of basic data types, such as numeric types, dates, vector arrays, e.g., numerical vectors for the visual content ontology and so on. The AIK-MO will be represented in OWL-DL which supports a reasonable set of data types. We will not support any specific inference over such data types.

Other Concepts which Need to be Covered by the Ontology

- **Provenance details - Multi-layer annotations.** A multimedia ontology framework should support annotations that distinguish between different modalities, annotations of different sources, annotations produced by different annotators (humans or not), thus enabling their linking and further processing. The annotation of images in AsIsKnown will have metadata attached to it. Thus, the necessary concepts and relations will be added to ontology. Also the annotation will be on two layers: low-level description of multimedia content which will be used to form abstract level descriptions, and the abstract level descriptions themselves which will be used as anchors for the content annotation of the images.

Following an Established Methodology

There is no a single correct methodology for designing and building an ontology. There are, however, a few methodologies that are well established and seem to perform better than others. Such a methodology might be adopted for the development of an ontology. The suggested ones are the OntoKnowledge methodology and the OntoClean methodology. We will adopt as much as possible from the methodology that we have used for the creation of AIK HTO (see deliverable D8). The main steps in our methodology come in order:

- *Processing of the standards and vocabularies in the domain*
In the case of AIK-MO we will start with the available multimedia ontologies and standards like MPEG7.
- *Formalisation of the terms*
In most cases we will have the terms in formal representation, but where necessary we will add our own formalization.
- *Link to an upper ontology*
We will link AIK-MO to DOLCE similarly to AIK-HTO. In this way we will meet the requirements of OntoClean methodology.
- *Evaluation*
The evaluation of the AIK-MO will be done with the help of partners that are developing the Multimedia database and the Trend Analyser. The evaluation will be mainly performed via the annotation of representative multimedia documents.
- *Documentation*

The AIK-MO construction will be supported by the relevant documentation.

This methodology is based on an overview of some well established methodologies (see the deliverable D8) and ensures a relatively wide coverage over the domain. Also, the connection to the upper ontology allows the checking of consistency of the ontology and facilitates the combination with the domain ontology (AIK-HTO in our case). In our opinion, following this methodology we will be able to meet all of the listed above requirements.

4

Existing Multimedia Ontologies

In the process of creation of AIK-MO we will rely on existing multimedia ontologies and some standards in the domain of multimedia. Here is a list of some of the existing multimedia ontologies with which we will start the creation of AIK-MO. Their valuable features include: (1) almost all of them are based on MPEG-7; (2) they support the activities necessary for AsIsKnown – annotation, analysis, retrieval; (3) they are aligned to an upper ontology. Some of the ontologies are freely available. For the other we plan to contact their authors.

1. *aceMedia Ontology Framework* (<http://www.acemedia.org/aceMedia>)

It gives support to a wide range of personalization facilities in concerning multimedia data sources. Scope and Usage: Annotation; Analysis. Media Description Structures: Images and Video. Based on: MPEG-7 (VDO (Visual Descriptor Ontology) and MSO (Multimedia Structure Ontology)) and Spatio-Temporal extensions. Link to: DOLCE.

2. *AIM@SHAPE Ontology* (<http://www.aimatshape.net/>)

An ontology developed for representing, modeling and processing knowledge, which derives from digital shapes. The term digital shape involves every visible object that possesses some space, and that is digitally stored. Scope and Usage: Retrieval of shape models, structures and tools; Analysis of shape models; Annotation of shape models, structures and tools; and Reasoning about shape models, structures and tools. Media Description Structures: Images (2D and 3D); Videos; 3D models of shapes; and Animation of 3D objects. Multimedia domain: Segmentation of 3D shapes; Skeletonization of 3D shapes; H-ANIM, for the description of virtual humans. Link to: AIM@SHAPE common ontology for shapes, tools and workflows - customizable. AIM@SHAPE is a Network of Excellence project within the EC's IST Sixth Framework Program.

3. *Medical Image Domain Ontology* (<http://www.med.univ-rennes1.fr/>)

An ontology developed for multimedia reasoning with medical images in order to interpret mammography images and assist diagnosis. Scope and Usage: Instance Retrieval of images; Text retrieval. Media Description Structures: Image; Text. Multimedia domain: MPEG 7 Descriptors (Color, Shape; Texture); extends current VDO with Shape 3D Descriptors; SphericShape3D; EggShaped3D. Linked to: aceMedia Visual Descriptor Ontology.

4. *NM2 Ontology* (<http://www.joanneum.ac.at/>)

An ontology developed for creating personalized versions of narratives in conjunction with a rule-based component. Scope and Usage: Automatic and manual annotation; Retrieval; Contextual filtering and matching of multimedia content. Media Description Structures: Multimedia programs. Multimedia domain: MPEG-7 (audio, video, images).

5. *MEPCO (Media Presence and Campaign Ontology)* (<http://www.joanneum.ac.at/>)

An ontology developed for cross relating and linking media campaigns. Scope and Usage: Semantic annotation and generally knowledge management. Media Description Structures: Multimedia - TV, Internet. Multimedia domain: MPEG-7 (audio, video, images). Linked to: PROTON upper ontology (<http://proton.semanticweb.org/>)

6. *Salero Ontologies* (<http://www.music.tuc.gr/~chrisa/>)

An upper ontology developed for Media Production. It aims to develop multimedia objects with context - aware behaviour in order to be re-used by artists who want to create complex and compelling media. Scope and Usage: Annotation (basis for creation and manipulation of multimedia object); Retrieval; Reasoning; Cross-platform delivery. Media Description

Deliverable D16 "Report on the state of the art for the design methods of multimedia ontology"

Structures: Images, sounds, videos, characters, movements, behaviours. Multimedia domain: Uses essence internal formats to describe low-level 'physical' features, MPEG-7 to describe low-level semantic features.

7. *MPEG-7 Upper Ontology* (<http://www.music.tuc.gr/~chrisa/>)

an ontology developed for supporting knowledge-based multimedia applications so as to capture all the MPEG - 7 MDS. Scope and Usage: Annotation; Retrieval; Filtering and Analysis (future).

8. *Multimedia Core Ontology* (<http://cidoc.ics.forth.gr/scope.html>)

CIDOC CRM Ontology: is proposed to be considered as part of a core ontology for Multimedia Objects supporting all multimedia objects, especially concerning cultural heritage items and events. Scope and Usage: Annotation; Retrieval. Media Description Structures: 2D Images; 3D scanning & photogrammetry of objects; 3D objects; Recordings; Multispectral images. Multimedia domain: MPEG-7 (audio, video, images).

9. *SWIntO: SmartWeb Integrated Ontology* (http://smartweb.dfki.de/ontology_en.html)

An ontology constructed by merging of several existing ontologies including DOLCE, SUMO, MPEG-7. Scope and Usage: Annotation; Retrieval. Media Description Structures: 2D Images; 3D objects; Video; Text. Multimedia domain: MPEG-7 (audio, video, images). Linked to: DOLCE, SUMO.

5 Conclusions

In this report we identify the scope of AsIsKnown Multimedia Ontology (AIK-MO) on the basis of its usage in the AsIsKnown System. Then we present a common list of requirements to multimedia ontologies and discuss how they reflect on the creation of AIK-MO. Then we list some existing multimedia ontologies which we plan to use in the actual creation of AIK-MO.

References

- [1] Hariklia Eleftherohorinou, Vasiliki Zervaki, Anastasios Gounaris, Vasileios Papastathis, Yiannis Kompatsiaris (ITI) and Paola Hobson (MOTOROLA Labs). 2006. Towards a Common Multimedia Ontology Framework (Analysis of the Contributions to Call for a Common multimedia Ontology Framework Requirements). http://www.acemedia.org/aceMedia/files/multimedia_ontology/cfr/MM-Ontologies-Reqs-v1.3.pdf.
- [2] MPEG-7. Moving Picture Experts Group Home Page. <http://www.chiariglione.org/mpeg/>