Specific targeted research project
IST call 4
FP6-2004-IST-4

Fully Automated multimedia Content management and personalised Synthesis FACS

Date of preparation: 22 March, 2005

List of participants:

1	Consiglio Nazionale delle Ricerche – Istituto di Scienze e Tecnologie dell'Informazione "Alessandro Faedo"	ISTI-CNR	IT	
2	Carnegie Mellon University	CMU	USA	
3	Stichting Centrum voor Wiskunde en Informatica	CWI	NL	
4	Fraunhofer – Gesellschaft zur Foerderung der angewandten Forschung e.V. - Institute for Media Communication	FhG	DE	
5	Technical University of Crete	TUC	GR	
6	Technische Universität Wien	TUWIEN	A	
7	Universiteit van Amsterdam	UVA	NL	

Name of the coordinating person: Pasquale Savino
Coordinator organisation name: ISTI-CNR
Coordinator email: Pasquale.Savino@isti.cnr.it
Coordinator fax: +39-050-315 2810

Table of contents

B1. SCIENTIFIC AND TECHNOLOGICAL OBJECTIVES OF THE PROJECT AND STATE OF THE ART 5

B2. RELEVANCE TO THE OBJECTIVES OF THE IST PRIORITY 10

RELEVANCE TO SPECIFIC OBJECTIVES OF THE CALL 10
RELEVANCE TO STRATEGIC OBJECTIVES OF THE WORK PROGRAMME 11

B3. POTENTIAL IMPACT 13

STRATEGIC IMPACT 13
INNOVATION 13
EXPLOITATION 14
DISSEMINATION 14
EU ADDED VALUE 15
RELATIONSHIPS WITH OTHER NATIONAL AND INTERNATIONAL RESEARCH 15
B3.1. CONTRIBUTIONS TO STANDARDS 16

B4. DESCRIPTION OF THE CONSORTIUM 18

[1] – Consiglio Nazionale delle Ricerche – Istituto di Scienze e Tecnologie dell'Informazione "Alessandro Faedo" (ISTI-CNR) 19
[2] – Carnegie Mellon University (CMU) 20
[3] – Stichting Centrum voor Wiskunde en Informatica (CWI) 20
[4] – Fraunhofer – Gesellschaft zur Foerderung der angewandten Forschung e.V. - Institute for Media Communication (FhG) 21
[5] – Technical University of Crete (TUC) 21
[6] – Technische Universität Wien (TUWIEN) 22
[7] – Universiteit van Amsterdam (UVA) 23
B4.1. OTHER COUNTRIES 23

B5. DESCRIPTION OF PROJECT MANAGEMENT 26

B5.1. PROJECT ORGANISATION 26
B5.2. MANAGEMENT STRUCTURE AND RESPONSIBILITIES 27
B5.2.1. Project Director (PD) 27
B5.2.2. Project Management Board (PMB) 28
B5.2.3. Scientific Advisory Board (SAB) 28
B5.2.4. Work Package Leaders (WPL) 28
B5.2.5. Scientific Executive Committee (SEC) 28
B5.3. MANAGEMENT PROCESSES, MECHANISMS AND PROCEDURES 28
B5.3.1. Project web site and management tools 28
B5.3.2. Communication strategy 29
B5.3.3. Reporting Requirements, Reporting Periods and Project Milestones 29
B5.4. CONSORTIUM AGREEMENT 29

B6. DETAILED IMPLEMENTATION PLAN 30

B6.1. STRUCTURE OF THE IMPLEMENTATION PLAN 30
B6.2. INTRODUCTION 30
B6.3. RTD ACTIVITIES 31
B6.3.1. Modelling 32
B6.3.2. Component-level research 33
B6.3.3. Research activity monitoring 34
B6.3.4. Relationships and dependences of RTD activities 34
B6.4. INNOVATION-RELATED ACTIVITIES 35
B6.5. CONSORTIUM MANAGEMENT 35
B6.6. RISK ANALYSIS AND MANAGEMENT 35
B6.7. PROJECT EXECUTION APPROACH 36
B6.8. DETAILED WORK PACKAGE DESCRIPTION 37
B6.8.1. WP1: Research activity management and monitoring 37
B6.8.2. WP2: Feature extraction 37
B6.8.3. WP3: Semantic and rhetoric annotation 39

B6.8.4. WP4: Modelling and Supporting User Profiles and Contexts ... *40*
B6.8.5. WP5: Intelligent multimedia objects modelling ... *42*
B6.8.6. WP6: Development and support for Semantic Multimedia Content and Query Models ... *43*
B6.8.7. WP7: Management of content, context, and profile descriptions ... *44*
B6.8.8. WP8: Presentations delivery ... *45*
B6.8.9. WP9: Dissemination and Exploitation ... *46*
B6.8.10. WP10: Project Management ... *47*
B6.9. GRAPHICAL PRESENTATION OF COMPONENTS (PERT DIAGRAM) ... 48
B6.10. WORK PLANNING AND TIMETABLE (GANTT CHART) ... 49

B7. OTHER ISSUES ... 70

B7.1. ETHICAL ISSUES ... 70
B7.2. GENDER ISSUES ... 71
B7.2.1. Gender and law recommendations ... *71*

APPENDIX A ... 73

APPENDIX B ... 75

APPENDIX C – REFERENCES ... 77

APPENDIX D – CV OF KEY PROJECT PERSONNEL ... 81

Consiglio Nazionale delle Ricerche – Istituto di Scienze e Tecnologie dell'Informazione "Alessandro Faedo" (ISTI-CNR) ... *81*
Carnegie Mellon University (CMU) ... *83*
Stichting Centrum voor Wiskunde en Informatica – CWI ... *84*
Fraunhofer – Gesellschaft zur Foerderung der angewandten Forschung e.V. - Institute for Media Communication (FhG) ... *86*
Technical University of Crete (TUC) ... *86*
Technische Universität Wien (TUWIEN) ... *87*
University of Amsterdam (UVA) ... *88*

APPENDIX E – LETTERS OF SUPPORT ... 90

Proposal summary page

Title: Fully Automated multimedia Content management and personalised Synthesis

Acronym: FACS

Strategic objectives addressed

IST-2004-2.4.7

Proposal abstract

Within the Strategic Objective 2.4.7, the FACS Project addresses the theme "Exploring and bringing to maturity the intelligent content vision".

The project aims at enabling a new modality for the acquisition of multimedia information which is interesting for users, through the automatic creation of interactive multimedia personalized presentations. These presentations are generated from Intelligent Objects that embed user pertinent clips from multimedia documents and provide the tools that adapt themselves automatically to the user needs (explicitly or implicitly expressed) and generate personalised interactive presentations.

The Project will heavily invest in research to develop the following activities:

- Definition of a model describing multimedia content that enhances and integrates emerging standards, such as MPEG7/21, and domain ontologies, such as OWL.
- Automated generation of media content descriptions, focusing on the acquisition of features and concept classifications, and on the recognition of rhetorical information.
- Development of advanced context models described in knowledge representation structures that can take into account the international standards as well as domain specific ontologies, and user profile models for personalization and contextualization of information and users. User context and profile will be dynamically updated and refined depending on external input including also the user interaction history.
- Management of descriptions of content and context and user profiles in a highly distributed environment based on GRID technology. Definition of new techniques for efficient and effective search and filtering of information by using approximate match paradigm in addition to the exact match paradigm.
- Automated generation of Intelligent and active multimedia presentations. Definition of new techniques to generate professional quality interactive multimedia presentations starting from annotated (portions of) multimedia documents

B1. Scientific and technological objectives of the project and state of the art

Our "vision" of the content scenario by 2010-2012

The 1970's were dominated by the use of large mainframes, in the 80's computing power went to the user's desktop, and with the PC revolution, from the mid 90's the new frontier became the creation of a completely connected world. According to Merrill Lynch, from 2010 (peaking around 2030) we will enter the "Content Centric" era. In this vision, broadband networks will be pervasive and user personal content will be acquired, stored and processed directly on the network. Consider the following futuristic scenario:

> *John Citizen lives in Brussels, holds a degree in economics and works for a multinational company dealing in oil imports. He enjoys travel with emphasis on warm Mediterranean sites with good swimming and fishing. When watching TV his primary interest is international politics, particularly European. During a recent armed conflict he wanted to understand different perspectives on the war, including both relevant historical material as well as future projections from commentators.*
>
> *When he returns home from work, a personalized interactive multimedia program is ready for him, created automatically from various multimedia segments taken from diverse sources including multimedia news feeds, digital libraries, and collected analyst commentaries. The program includes different perspectives on the events, discussions and analysis appropriate for a university graduate. The video program is production quality, including segment transitions and music. Sections of the program allow him to interactively explore analyses of particular relevance to him, namely the impact of war on oil prices in various countries (his business interest), and its potential affect on tourism and accommodation prices across the Mediterranean next summer. Some presentations may be synchronized with a map display which may be accessed interactively. John's behaviour and dialogue with the display are logged along with a record of the information presented to allow the system to better accumulate his state of knowledge and discern his interests in order to better serve him in the future. When John is away from home for business or leisure, he may receive the same personalized information on his mobile device as well, emphasizing information reflecting the neighbourhood of his current Mediterranean location.*

This "vision" has many consequences from both societal and economic perspectives. Societal consequences impact personal information, continuous education, e-government, tourism, leisure, etc.. The economic infrastructure of information provision and dissemination will also change, as new actors will enter the market (mainly providers of the new services to be offered), and existing actors will radically change their mode of operation (think about the new paradigm for the production of the TV programs of the future) and business model.

The objectives of the FACS proposal are derived from this vision.

Project objective

> **FACS will study and develop the models and software components needed to automatically create and deliver personalized and context-aware interactive audio/video presentations.**
>
> **FACS will thus investigate new techniques for the:**
>
> - **Representation of multimedia content descriptions**
> - **Representation of user preferences and context**
> - **Automated content annotation**
> - **Automated presentation authoring**
> - **Distributed retrieval and filtering of content descriptions and user preferences**

The core of the project activity will aim at enabling a new modality for access to information of interest through the automatic creation of interactive, mixed-media, personalised presentations on demand or prescribed. These presentations function as Intelligent multimedia Objects (or *I-Objects)* that embed selected (e.g. novel and pertinent to the user's information need) clips from multimedia documents and provide the tools that adapt the presentations automatically to the user requirements (explicitly or implicitly expressed) that range from content selection to content presentation and rendering.

Automatically synthesized interactive multimedia presentations will go beyond the simple collection of sequences of potentially relevant multimedia segments: I-Objects can generate presentations comparable to professionally edited and composed interactive multimedia programs. The creation of a personalised I-Object, as shown in Figure 1 will be obtained through the following steps:

1. **Automated multimedia document annotation.** Multimedia documents are automatically annotated with metadata expressing the (i) semantic, and (ii) rhetoric/affective/emotional (in the rest of the proposal we will simply refer to these as rhetoric) content of each clip. Semantic annotations identify the conceptual content of the document while rhetoric annotations denote their perspective, e.g. documents may convey an attitude or a bias in their content by

Figure 1: Automated generation of I-Objects

virtue of language, movement, juxtaposition, colour, rhythm, etc. Rhetoric annotations give information like "video sequence with dramatic presentation", "expresses negative opinion vehemently", "evocative music sequence", "visually stimulating picture", etc.

2. **Acquisition of user profiles**. The profile and context of users are created and updated by observing their behaviour when interacting with a presentation through a device, by logging a record of what they have seen, and by allowing them to explicitly express their preferences or modify their profiles.
3. **Creation of presentation templates**. Experts of multimedia presentation authoring will create a set of presentation templates enriched with rhetorical information. A presentation template describes the presentation structure as well as the active/intelligent behaviour of the I-Objects that will be generated, but does not include the media clips to be presented. All components of a presentation template are annotated with specific rhetorical descriptions that indicate what type of clips should be used for that component, from a rhetorical perspective.
4. **Personalized generation of I-Objects**. A personalized I-Object will be an active/intelligent presentation with style and content pertinent to the user interest and prior knowledge. A personalized I-Object will be generated by a) selecting clips of documents having semantic content relevant to the user's requirements, b) selecting the most appropriate presentation template on the base of the user profile, c) binding elements of the selected template to the selected clips, according to their rhetorical annotations. A personalised I-Object will be capable of further refining the final generated presentation by adapting itself to the user's environment, access device and preferences.

Research objectives

The key research issues that will be addressed in the project are the following:

- **Representation of multimedia content descriptions.** We will define a model for describing multimedia content that enhances and integrates emerging standards, such as MPEG7/21, to represent all the video content metadata. We will use MPEG-7 as a knowledge representation language capable of describing semantics in real world applications and we will create a software environment that will allow the description of domain ontologies using MPEG-7. Since MPEG-7 is not really known and used by people that are familiar with the construction and use of ontologies and rather difficult to use for this purpose, we will create an upper layer of interaction with users, which will be based on OWL.
- **Representation of user preferences and context.** We will develop advanced models for personalization and contextualization of information. Models will include those for the context of the original source media as well as those reflecting the situation of the user. User models will include structures to describe personal demographic profiles and preferences, local setting (e.g., time, place, and organization), task assignment and goal, and recursively, other domain specific ontological structures, to enable contextual reasoning. User context and profile will be dynamically updated and refined to reflect external input including the user retrieval and interaction history.

The context models will be described in knowledge representation structures that adhere to the relevant international standards

- **Automated content annotation.** We focus our approach on evolving a cognitive machine understanding of video and audio. We will start with the extraction of low-level, machine recognizable features from the temporal image, and audio files; use these to identify higher-level objects, movements, sounds, words and phrases including names, times and places; combine these to derive actions, activities, and concepts of the subject matter; and finally classify them as to their rhetorical perspectives and affect. All such derived knowledge is coded into a complex database that enables search, correlation and recombination along any dimensions. The goal is to bridge the "semantic gap" between the extraction of the underlying raw feature data and the need for a semantic description of the content and its style. We will build on numerous techniques from probabilistic reasoning - statistical inference, neural networks and machine learning - to enable the underlying detection and classification process. We develop and apply ontologies of visual concepts to understand what is seen; we apply speech recognition and natural language processing to understand what is said; and we develop a machine learning of creative video and audio production methods to understand what may be meant.
- **Automated presentation authoring.** We will employ new techniques to generate professional quality interactive multimedia presentations starting from annotated (portions) of multimedia documents. We will investigate how media and content descriptions can be embedded in I-Objects, jointly with the intelligence to combine them and to autonomously produce end-user interactive multimedia presentations. Intelligent objects will include content metadata, state, behaviour and they will be capable of interoperation with the environment (user interactions, user profiles and contexts). They will be capable of modifying the environment (user profiles and contexts) and being modified by it (object behaviour and presentation). We will create and use prototypes of presentation layouts built according to semantic, rhetorical and presentation optimisation strategies that allow the I-Object to generate the most appropriate presentation. The same I-Object might generate different presentations corresponding to differences in both user profiles (e.g., age and education attributes) and the rhetorical context in which it is delivered, and then be appropriately rendered for the user's output device.
- **Distributed retrieval and filtering of content descriptions and user preferences.** We will represent content, context and user profile descriptions using standardized XML conventions in order to enable import, export, interpretation and real-time customization in a widely distributed, non-uniform, information and user network. As evidenced by recent activity (e.g. RSS and Podcast), we anticipate a plethora of virtually continuous user-generated multimedia content sources in the future. The techniques developed in FACS have the potential to provide a method, apparatus and standards for their automatic, real-time, annotation. Given the high dynamicity, continuous production, and wide distribution of content and descriptors, we will leverage and enhance current GRID technology for their data management. To facilitate global search and filtering, we will develop multi-modal approximate and exact match paradigms. To cope with the issues of unbounded and continuous media production, we will define new and efficient distributed XML data storage concepts, including novel distributed access methods and new query optimisation and processing techniques that utilize the data management capabilities of the GRID environment.

We will utilize the TECVID video collection [67], the ISMIR audio description contests [39], and the existing infrastructure and some of the news and documentary corpora of the Carnegie Mellon University Informedia Digital Video Library [92]. Possessing such-large corpora of data (six terabytes of images, audio, and video with automatically generated metadata and indices for retrieving segments) enables validating component function, testing interoperability, quantifying performance, and measuring progress. This is further critical to measuring correctness of results, demonstrating scalability of methods and enabling user studies and evaluations.

State of the art

Automated generation of media content descriptions

MPEG-7 and MPEG-21 are the default standards for describing multimedia content. Recently it has been shown that these standards can be used as a knowledge representation language and have been used to describe the semantics of the content of sport videos utilizing complex sports ontologies (such as soccer ontologies) [25], [26], [27]. The resulting metadata descriptions are MPEG-7 compliant and can be used by any MPEG-7 application. It has also been shown how to produce MPEG-7 compatible ontologies utilizing OWL as an ontology description language [28], [29].

In this project we will extend the state of the art by providing generalized methodologies for the easy integration of ontologies for various application domains within MPEG-7. The result will be a generalized methodology and tools for creating MPEG-7 compatible models for the semantic description of content in various application domains. These models include connectors, which act essentially as intermediate ontologies (related to application domains) that interconnect the higher level MPEG-7 ontology with the lower level (domain-specific) ontologies.

The fundamental obstacle in automatic annotation is the *semantic gap* between the digital data and their semantic interpretation. Progress is currently being made in known object retrieval [65] [44], while promising results are reported in object category discrimination [62], all based on the invariance paradigm of computer vision. Significant solutions to access the content and knowledge contained in audio/video documents are offered by StreamSage [91] and Infomedia [92]. While the field of content-based retrieval is very active by itself, much is to be achieved by

combination of multiple modalities: data from multiple sources and media (video, images, text) can be connected in meaningful ways to give us deeper insights into the nature of objects and processes.

We will extend the state of the art by addressing the combination of multiple modalities together with novel machine learning techniques. So far multimodal data knowledge mining has mostly been carried out separately on each information channel. Today, however, knowledge sources that marry multiple descriptions are urgently needed to support the analysis and retrieval of mixed-media. The picture–text combination for example is widely considered to be the richest option for information access. In worldwide, task-based retrieval evaluations such as TRECVID, an integrated approach combining text and visual information is the essential ingredient in the most successful systems [24], in combination with the use of machine learning techniques.

Wold et al. [32] presented a system which analyzes sounds based on their pitch, loudness, brightness, and bandwidth over time and tracked the mean, variance, and autocorrelation functions of these properties. Other approaches (e.g. [42], [64], [21]) are based on methods developed in the digital speech processing community using Mel Frequency Cepstral Coefficients (MFCCs) and motivated by perceptual and computational considerations. However, the MFCCs ignore some of the dynamic aspects of music. A different approach is taken by the SOM-enhanced Jukebox system, [14], where characteristics of frequency spectra are extracted and transformed according to psychoacoustic models focusing on the rhythmic characteristics. While classification into semantic categories is a comparatively mature field, having produced a range of approaches and results, annotation according to affective or emotional categories of audio is a relatively young domain, gaining importance recently [89] [90] as conventional genre categories are seen as too limited, inflexible and not applicable to the wide range of individual perceptions.

We will enhance the state of the art in audio feature extraction by developing new feature extraction techniques and/or optimizing existing feature extractors, to cope with the requirements raised by semantic and affective indexing. Recent discussion in the music analysis domain has frequently raised the issue that conventional genre concepts are not sufficient to describe the heterogeneous aspects of music and the resulting interpretations by listeners. We will thus also investigate new concepts for characterizing and describing audio effects and perceived impressions, to allow a multi-layered interpretation of audio-signals that are capable of accommodating the interpretations of different users. This will require new classification models as well as optimized feature sets for the given tasks.

Personalization and Contextualization

McCarthy [70] introduced contexts as formal objects. Context is very useful for localizing the reasoning to a subset of facts known to an agent about a specific problem. Contexts are seen as local, domain-specific, goal-driven theories of the world, and are building blocks of what an agent knows. Contexts have been used extensively in AI to formalize agents which have a representation of the changing beliefs, intentions and goals of other agents involved in dialogue or negotiation [71], [72], [73], [74]. Context is also used in linguistics where it refers to the words in an utterance that are near in the focus of attention and further specify it [75]. In information systems, context has been described in terms of context types [80] that relate to application types. Context types proposed in [81] include organizational, domain-based, personal, and physical context. Each is subdivided into more specific types like workflow and structure, domain ontology, knowledge profiles, usage history, interest profiles, etc. In ubiquitous environments context types including location and time are often used for proximate selection and for context-driven actions [82]. Context OWL (C-OWL) [55] is an extension of OWL that allows the representation of contextual ontologies. It allows the personalization of views of one or more ontologies that encode a single party's understanding. C-OWL allows to contextually relating two ontologies. A context mapping is a set of rules that bridge between the OWL ontology and a target ontology.

Personalization in multimedia content is provided by MPEG-7/21 via user profiles and usage history. However, the personalization of MPEG-7/21 is limited and cannot account for the domain knowledge described in ontologies and thus it cannot exploit combined MPEG-7/domain-specific metadata descriptions.

In this project we will extend the state of the art in the field by providing a generic and expandable model for multimedia information that will integrate the various types of contextual and personalization information with multimedia content description preferences. This model will be compatible with the MPEG-7 semantic metadata descriptions of content and the domain ontology so that matching of content and preferences with MPEG-7/21 semantic content descriptions utilizing domain ontologies is enabled. The current MPEG-7 profiles will be a special case of the new semantic profiles.

Distributed management of XML descriptions

Several XML database management systems have been proposed, to deal natively with large amount of XML data, such as Tamino [9], Xindice[11], Lore [3], and Timber [10]. However, current XML database technology is mainly focused on an *exact match paradigm* (with some text retrieval functionalities).

Distributed data management has been investigated in the GRID field by various initiatives. Among the most well known [1], which considers a file-system level data management; OGSA-DAI [5], which deals with integrated access

to different (relational and XML) databases; OGSA-DQP [50], which offers distributed query processing on multiple wrapped data sources; and GRACE [2], which offers text retrieval and categorisation on GRID enhanced search engines.

In the project we will enhance current state-of-the-art by investigating techniques that consider a highly distributed information source space as a single highly dynamic XML database, where we combine complex exact and approximate match searches. These can be efficiently processed by leveraging and enhancing current GRID technology for data management. Approximate match queries are very useful in contexts where ambiguous, imprecise, descriptions are to be matched.

Automated Generation of Presentations from Intelligent Objects

There are several existing technologies available that can be used to define storage containers for objects. Formats such as MPEG4 [4] provide an *encapsulated object model*, in which one or more instances of a media object can be stored, along with relevant annotation information. MPEG4 supports a multi-layered model in which one or more low-level media encoding are packaged into a composite object as a collection of MPEG streams. While complex stream collections are possible, containing multiple encodings for multiple target environments, such an approach typically leads to an explosion of encodings within a single package, as the final presentation target environment is defined. Practical use of multiple MPEG4 streams is also limited by minimal support for complex stream processing within most MPEG4 codecs. XML languages such as SMIL [84] provide an indirect reference model for defining the multiple media encodings associated with a single abstract media item; SMIL contains an extensible content selection mechanism that be used to define a semantic and syntactic hierarchy to serve as the basis for final form object selection. Neither MPEG4 nor SMIL provide complete solutions to the object encoding problem, since multiple layers of semantic markup will need to be supported, but both technologies provide a valid starting point for investigating presentation component persistent storage.

The existing approaches for encoding presentation style information are less complete than media object storage. The GR*i*NS editor [83] [85] supported the use of a language to specify generic presentation templates for defining abstract layouts from which to generate run-time presentation instances. However, this language supported only physical attributes (e.g., screen size, rendering capabilities, available codecs) when associating media objects with layout elements. Such policies will need to be expanded to define more generic mapping of content to rhetoric, semantic, and physical device properties.

Given a set of presentation style descriptions and candidate media objects, final objects need to be generated subject to knowledge of the target environment and the assets available for presentation. Many text-based systems can make use of style sheet transformations (using XSLT and CSS), but these languages are not rich enough to support a wide range of media and affective content characterizations. Other approaches to solving portions of this generation problem include those of Weitzman et al. [86], who used a method based on relational grammars, and Zhou [87], who used a visual planning model. More recently, several implementations of presentation generation have been defined that are based on the use of rhetorical structure theory (RST) [69], but all of these approaches are tailored to generate text descriptions based on a structured markup at the text encoding level. Significant new effort will be required to provide efficient models and mapping approaches to support the goals of FACS.

B2. Relevance to the objectives of the IST Priority

The FACS project aims at studying, and verifying the feasibility of new ways to create and deliver personalized and context-aware interactive audio/video presentations and at developing all enabling necessary techniques to achieve this goal. The project results will allow users to view personalized interactive programs, automatically synthesized according to their preferences, on different devices for entertainment, personal information, infotainment, etc.

To achieve this goal it is necessary to heavily invest in research in a number of inter-related fields that do not constitute the background of any existing research community. As underlined in other sections of the proposal, the Project will require research contributions from many different areas, such as automated annotation of sound and video; modelling and automatic acquisition of user profiles; modelling of user and data context; modelling of interactive multimedia presentations; modelling of multimedia content description; management and approximate retrieval in a distributed environment of all different types of information and knowledge; automatic synthesis of interactive multimedia presentations and adaptation of the presentations to different delivery devices.

The FACS project will thus combine and coordinate the effort of highly qualified research groups, European or worldwide leaders in these different areas. The creation of this blend of competences is the key factor for the project to be successful.

FACS is positioned within the Strategic Objective "Semantic-based Knowledge and Content Systems" (2.4.7) of the 2005-06 Work Programme for the IST thematic priority. FACS also addresses many of the general and specific objectives stated in the above-mentioned Work Programme.

The overall objective of Strategic Objective 2.4.7 is to

> *"develop semantic-based and context-aware systems to acquire, organise, share and use the knowledge embedded in web and multimedia content, information holdings of all media types, or processes which are driven or supported by ICT systems. "*
>
> *"Moreover, a visionary, strategically important opportunity exists to pioneer intelligent content, which will be self-describing, adaptive to context, and exhibit a seamless interaction with its physical and digital surroundings, and the user. Challenges relating to the ongoing convergence of content and knowledge with metadata based systems and processes for all-digital content creation and management will also be addressed."*

The research issues addressed in the project aim at a significant improvement in methods for

- The **automated generation of semantic and rhetoric** annotations of multimedia material, with particular emphasis on sound and video, as these are the type of data whose management requires the most significant advances in order to achieve a full **holding of all media types.** The project aims at using multimedia content made available from many different sources, from newly created content to data acquired from digital archives or from the web.
- **Organization and sharing** of multimedia content descriptions in an interoperable standard based and ontology based manner among different users.
- Automated **analysis of user behaviour** in order to create user profiles that utilise content description standards and ontologies to select information from different sources and to compose user-specific presentations. This analysis will be done by taking into consideration the **user context**.
- Automated synthesis of **intelligent content** (I-Objects) to produce multimedia interactive presentations.
- **Seamless interaction** of the I-Objects with the **physical and digital surroundings** and **with the user**, in order **to adapt them to different devices and different user needs**.

Relevance to specific objectives of the call

The FACS Project addresses the theme "Exploring and bringing to maturity the intelligent content vision" of the 2.4.7 Strategic Objective. However, some of the research topics of the project are also related to the theme "Knowledge acquisition and modelling" of the same SO.

In particular, as already stated, the major aim of the project is **to explore and bring to maturity the intelligent content vision** by defining and modelling I-Objects, and by providing the necessary research advances that enable their creation.

> *In particular, the project will define a reference architecture that describes the functionality of all components needed to build a system that is able to automatically synthesise interactive multimedia presentations. All theoretical activities (modelling of I-Objects, modelling of user profiles and context, and modelling of data contextual information) and necessary developments of different components (automated semantic annotation*

of sound and video, automatic acquisition of user profiles, support for data and profile search and filtering, automatic generation of I-Objects) will be defined, studied and experimented within the project.

I-Objects will be produced by the automatic synthesis of existing content in order to **seamlessly create new content tailored to user needs**.

One of the main objectives of the project is to provide the user with automatically synthesised personalised interactive presentations obtained through the automatic analysis of multimedia contents to produce descriptions that can be used for retrieval and filtering purposes. The project will investigate and develop appropriate techniques and algorithms for the effective cross analysis of content in order to have higher-level semantically significant content descriptions. In particular, the work carried out in the activity related to ***image and video analysis, and sound analysis, will result in various techniques and models for the description of features of image, video, and sound****. From these features, higher-level descriptions will be extracted capturing the semantic and rhetorical message of the audiovisual content. The extraction will exploit information from cross media analysis of audiovisual content as well as application specific ontologies.*

These objects will **interact with networks and devices** and adapt their behaviour.

In particular, the system will permit: to specify exactly what the user wants, not only in terms of retrieval, but also in terms of synthesis of the personalized programs; automatic tracing of the access patterns and the access history of the users so that the synthesis is produced according to their changing interests and background knowledge; to associate new synthesized programs with others synthesized for the same user in the past, and provide the appropriate reference links whenever needed; to access personalized programs from various multimedia devices like third generation mobile phones, PDAs, and computers, in addition to television sets.

Furthermore, the Project will investigate new techniques **to automatically capture knowledge from raw information and multimedia content in distributed repositories** to make this content usable for the creation of the I-Objects.

In particular the project will allow an automated extraction of feature descriptors from sound and video, and automated semantic and rhetorical annotation of sound and video.

Relevance to strategic objectives of the work programme

From the point of view of the general societal and economic challenges, FACS can lay the grounds for improving competitiveness of both large and small businesses. In fact, the project will define, demonstrate the feasibility, and provide the needed technological advances of a new paradigm for the production, synthesis and delivery of multimedia presentations and programs. This will have consequences from both the economic and the societal point of view. If the project is successful, all existing commercial actors of the content market will be influenced. Moreover, new actors will be encouraged to enter this market, namely service providers specialized in the production of I-Object's templates, and service providers specialized in the automatic synthesis of presentations for commercial sectors. Users will also be strongly influenced, since the paradigm that is behind the project situates the end user in a key position in the production chain: the programs are produced according to the user needs and wants, and not according to what the program producer thinks that the user needs and wants.

The project will also have a number of by products that will have value in themselves, even if the main project goal should not be fully achieved. Indeed, the main project objective will require research in different fields that can be applied in areas such as multimedia Digital Libraries, multimedia search engines, e-learning, etc.

From the scientific point of view, FACS will contribute to the definition of a new vision of intelligent content management, with the definition and modelling of the I-Objects. Moreover, innovative contributions will derive from the definition of models for the description of multimedia annotations, of user behaviour, and user context.

The Project also addresses a significant number of strategic objectives of the IST Work Programme 2005-2006:

- FACS will "**stimulate research and development of ICT so as to master the technologies that will drive future innovation and growth**". In fact, the project will invest on key technologies such as the study of innovative techniques for automated annotation of audio/video, automatic synthesis of interactive presentations, adaptability of interactive presentations to different devices.
- FACS will "**promote the widest and best possible use of ICT-based products and services by all citizens**", by researching the basic technologies that enable the automatic synthesis of personalized and context-aware interactive multimedia programs, in order to provide innovative services in different areas, such as continuous education, e-government, e-entertainment.
- FACS will "**mobilize the research community around high-risk long term goals and enable the development of a European Research Area in IST**". Indeed, the project will involve research groups that are leaders in Europe (and US) in the different fields addressed in the project such as sound, image, and video analysis; machine learning;

knowledge acquisition; user modelling; contextual analysis; efficient retrieval. The ultimate goal of the project requires advances in all these different but related fields that are not available in a single research community, but require their "**cooperation on a global scale":** FACS will build the environment for the effective creation of this multidisciplinary research community.

- The main objective of the FACS project "**Focus(**es**) on** one of **the future generation [of] technologies in which computers and networks will be integrated into everyday environment, rendering accessible a multitude of services and application through easy-to-use human interfaces**".

B3. Potential impact

Strategic impact

Every day we are overwhelmed by information of many types: several TV channels, news feeds, newly published web sites, digital libraries, information that reaches us via the Internet. Users invest much time and effort in finding the information that they really need in this highly dynamic information space. Suppose users interested in the events related to the first democratic elections in Iraq. With today's technology they have a set of options: they can wait for the news broadcast on their preferred TV channel and hope that it will give good coverage of the event, otherwise they can choose a different channel and repeat the same procedure. Alternatively users can search their preferred on-line newspaper or use a search engine. In the first case users have no control on the available information and can just hope that something interesting is broadcast while they are watching TV. In the second case, the quality of the information they are able to find depends on their skill when navigating the web and interacting with search engines.

The FACS project aims at creating the foundations to radically change the way information is prepared, packed, and delivered to the user, without his/her direct intervention. The user will be provided with his/her "Special edition on the Iraqi elections" built using diverse content broadcast, found on the web, or provided by content providers, in the modes he/she prefers (through his TV set-top-box, personal web portal "My-e-news", third generation mobile phone).

The high potential impact expected of the project has already been underlined (see section B2): in the long term, FACS is expected to have **impact on society** as a whole, by changing the way users will access information, **on multimedia content service providers**, and **on the scientific community**.

Impact on society. End-users need access to personalised information in various application areas, such as Entertainment – e.g. to have a personalized view of specific sport events by highlighting interviewees or sport actions, or by providing information on a particular team, Infotainment – e.g. to view political events according to different perspectives, at different levels of detail, etc., Education – e.g. by delivering content that is perfectly tailored to the learning interests and to the know-how of the user.

The paradigm that underlies the project places the end user as a key player in the production chain: the programs are produced according to the user needs and wants, and not according to what the program producer thinks that the user needs and wants. This will influence the user's capability to formulate a personal opinion on everyday occurrences or on historical events, but will also give them more freedom in their choices for entertainment, continuous education, tourism, etc.

Impact on multimedia content service providers. Nowadays the production process requires the contribution of teams of experts working for long periods of time; thus the production of interactive programs personalized for specific users is not feasible at reasonable costs. The research conducted in the project will provide the technological solutions that make the production of presentation fragments economically feasible. This is expected to contribute to the improvement of the competitiveness of the EU content production market, by reducing the production costs, and by improving the quality of the production.

Different types of service providers may benefit from the project results: for example, television companies, will create personalised programs on the set-top-box synthesised from content received from different channels; content providers will be able to sell semantically enriched content to their customers; mobile phone operators will have the possibility to provide their customers with personalised programs to be received on their devices depending on their context; Internet providers will provide a personalised collection of news clips that can be automatically generated using user profiles and gathering information from various multimedia data source. Moreover, new actors may enter this market, namely service providers specialized in the production of I-Object's templates, and service providers specialized in the automatic synthesis of presentations for specific commercial sectors.

Impact on the scientific community. FACS will contribute to the definition of a new vision of intelligent content management, with the definition and modelling of the I-Objects. Moreover, innovative contributions will derive from the definition of models for the description of multimedia annotations, of user behaviour, and user context.

Innovation

A major project goal is to provide innovative methods for the production of interactive presentations: these will include new ways for the definition of presentation fragments, for their storage, representation, search, assemblage and distribution based on the requirements of the viewer.

All scientific and technical activities of the project are expected to produce innovative solutions:

- **Representation of multimedia content descriptions.** Standard based and ontology based descriptions of content will be used in all the aspects of the architecture from content descriptions to profiles and contexts to

personalization and presentation, thus providing an interoperable semantic framework for intelligent multimedia object delivery.

- **Representation of user preferences and context.** Advanced models for personalization and contextualization of information will be developed. Context models will include models for context capturing as well as models for the identification of the context of the user, which will include structures to describe contextual aspects, as well as contextualized domain specific ontological structures. User queries will be expanded with contextual and personalised information. Inference will be used to infer higher-level concepts from lower-level recognized concepts and for contextual reasoning.
- **Automated content annotation.** Innovation is expected from the possibility to automatically annotate sound and video material, with a large ontology composed of more than 1000 concepts. Moreover, annotation will include rhetorical description of the material. The software architecture will facilitate the metadata creation using domain ontologies while remaining compatible with industry standards.
- **Automated presentation authoring.** Techniques to generate professional quality interactive multimedia presentation starting from annotated portions of multimedia documents will be developed. The presentations will be non-linear and will be organized based on the semantic content of the objects retrieved, and they will also integrate intelligence. Models for representing intelligent multimedia objects capable to interact with their environment in an interoperable manner, and in a way that affects their appearance and behaviour and also modifies the environment will also be developed.
- **Distributed retrieval and filtering of content descriptions and user preferences.** We will define a model for expressing the semantics of user queries by exploiting the domain ontologies for metadata descriptions, techniques for query optimization, and for the efficient execution of queries on highly distributed (GRID) architectures.

Exploitation

The know-how and results of the FACS project will enable the definition of tools, methodologies, and guidelines for the smooth migration of the Content and Service Providers' traditional authoring environment towards new solutions for the end user based on automatically synthesised personalised and interactive digital content. The exploitation of project results is both within the FACS consortium and within the EU research community at large.

Exploitation within the consortium. All partners are universities or research institutions, and will strongly benefit from the scientific advances obtained from the cooperation with all other partners in a field that requires the expertise of many research groups.

Exploitation toward the EU research community. The FACS project will address its results exploitation mainly towards technology and service providers. Technology providers will benefit from the research and feasibility assessment in bridging the gap between the current solution architecture and the flexible approach proposed in the FACS project. The FACS unified solution will provide useful indications for the engineering of new functionality into the real distribution environment, both in terms of network capabilities and of service provisioning schema. Service providers will interact with new and different mechanisms coming from the personalisation of services preserving their internal technological solution and with shorter time to market for service supplying. This will be made feasible by the FACS architecture and the specific components developed in the project.

The project exploitation is based on the following strategy:

- Dissemination activities (as outlined in the following) in order to promote the adoption of the project "vision" and to demonstrate the feasibility of the project approach
- All software components and the resources (ontologies, I-Object templates, etc.) developed in the project will be made available as public domain. This approach is intended to support the diffusion and the adoption of project approach and results by the widest possible number of users.

The consortium members will define a detailed exploitation plan explaining in detail how the output of the project would be introduced into common and individual strategies. Particularly, the consortium will have to identify the services and/or the products that could be derived from the project results, based on a scientific/technological survey and on a market survey.

Dissemination

A structured dissemination plan will be carried out during the FACS project in order to support an effective exploitation of the project results towards four main communities: researchers, contents providers, services providers, end-users.

Dissemination activities will be conducted through the following strategy:

- The results of the project, and mainly the reference architecture will result in articles and documentation to be submitted to conferences, workshops and selected high-level magazines (e.g., ACM Multimedia, VLDB Journal, ACM-SIGMOD, ACM-SIGIR, ACM TOIS, IEEE TPAMI, ACM Multimedia Systems Journal, IEEE Transactions on Neural Networks, IEEE Transactions on Signal Processing, etc.).

- The project will regularly organize, in cooperation with the DELOS and MUSCLE NoEs, (annually) a scientific workshop for the international research community, and a series of industrial days (according to the main releases of project results), inviting content and service providers, TV broadcasters, and potential investors.
- Since the FACS project is aimed at developing and experimenting the technologies needed to produce personalised and interactive digital content for entertainment, personal information, and edutainment, a profitable interaction with standardization bodies and user groups – e.g. mainly IETF, Mobile Entertainment Forum, FIAT/IFTA (International Federation of Television Archives), EBU (European Broadcast Union), etc. – is expected.
- A project web site will be created at the beginning of the project and maintained during the entire project lifetime. Through this web site, the public documents produced within the FACS project will be made accessible, along with links to specific sites and documentation about the FACS technological issues.
- A two-monthly FACS Newsletter will be produced and delivered by e-mail to all communities interested to project results. All partners will contribute to the preparation of the newsletter.

The sharing of the expertise held by each partner will be a plus and will result rapidly in collaborative research and publications; in a longer timeframe, thanks to the employment of young researchers and Ph.D. students, it will result in opportunities for the new generation of researchers, both from Western and Eastern Europe.

EU Added Value

The problems that FACS aims to solve must be necessarily tackled at the European level, because significant breadth of complementary strengths and expertise is needed. The multinational FACS Consortium – including universities, and research institutions – reflects the complementary scientific expertise needed. This significantly improves the prospect of a value-adding solution for the European market. A consortium operating at the European level and within a European project has further strength in proposing standards and integrating its work with existing efforts in this area. In addition, carrying out the activities at a European level guarantees that the resulting solution is applicable – after customization – to a wide variety of European small and medium-sized enterprises.

Relationships with other national and international research

The FACS partners cooperate and are well positioned within the respective national and IST scientific committees and communities through ongoing projects, which will facilitate their integration of in FACS. Relations with other European and national research projects in the same area will be maintained at different levels. First, there are a number of projects where the FACS partners are or have been involved (see the table in Appendix A, containing a short description of each project), and there are projects whose results are considered relevant for the project activities (a first list is given in Appendix B, while a more detailed list will be produced during the first state of the art analysis). Here, it is worth mentioning two NoEs, DELOS and MUSCLE that have strong relations with FACS. Cooperation with both NoEs will be established in order to create cross-fertilization and to promote FACS results in a wide research community. A letter of support from coordinators of both networks is attached in Appendix E.

The **DELOS Network of Excellence** (started on January 2004), coordinated by ISTI-CNR, conducts a joint program of activities aimed at integrating and coordinating the ongoing research activities of the major European teams working in Digital Library - related areas with the goal of developing the next generation Digital Library technologies. The objective is to:

- define unifying and comprehensive theories and frameworks over the life-cycle of Digital Library information,
- build interoperable multimodal/multilingual services and integrated content management ranging from the personal to the global for the specialist and the general population. The Network aims at developing generic Digital Library technology to be incorporated into industrial-strength Digital Library Management Systems (DLMSs), offering advanced functionality through reliable and extensible services.

The Network will also disseminate knowledge of Digital Library technologies to many diverse application domains and provides a forum where researchers, practitioners, and representatives of interested applications and industries can exchange ideas and experiences. Furthermore, DELOS has seven research clusters on Digital Library Architectures, Information Access and Personalization, Audio-visual and non-traditional objects, User Interface and visualization, Knowledge Extraction and Semantic Interoperability, Digital Preservation, and Evaluation.

The *Networked Multimedia Information Systems Laboratory (NMIS) of ISTI* that coordinates the FACS project is also the coordinator of DELOS NoE. TUC is the coordinator of the audiovisual cluster of DELOS, which includes 18 European research centres. Furthermore, UVA, and TUWIEN are partners in the network and members of the audiovisual cluster. A strict cooperation will be established between the FACS project and the DELOS NoE, in order to have access to preliminary results of the network and to use the network as a vehicle for the dissemination of the FACS project results. A letter of support of the DELOS NoE Project Manager is attached to the proposal.

The **MUSCLE Network of Excellence** aims at supporting a research network to encourage close collaboration between research groups in multimedia data mining and machine learning. The important goal of the network is to make progress towards several ambitious objectives: (1) moving from modelling to learning by harnessing the full potential of

machine learning and cross-modal interaction for the (semi-) automatic generation of robust meta-data with high semantic value for multimedia documents. (2) improving interoperability and exchangeability of heterogeneous and distributed (meta)data by enabling data descriptions at high semantic levels (e.g. ontologies, XML schemata) and adding inference schemes that can reason about them at the appropriate levels. (3) improve the human-machine interface by exploring how machine learning can invigorate the creation of expressive, context-aware, and human-cantered interfaces that will be able to effectively assist users in the exploration of complex and rich multimedia databases.

In order to achieve these objectives, the network aims at using a wide range of techniques including natural language processing, interfacing technology, learning and inferencing, merging of different modalities, federation of complex meta-data, appropriate representation and interfaces, etc. another important aspect is the application range the network intends to address: surveillance and intrusion detection, face recognition and registration of emotion or affect, automatic analysis of sports videos and movies, etc. The success of these applications depends heavily on the integration and interpretation of various modalities such as vision, audio, and speech.

The research units of UVA and TUWIEN involved in FACS are also partners in the network. A strict cooperation will be established between the FACS project and the MUSCLE NoE, in order to have access to preliminary results of the network and to use the network as a vehicle for the dissemination of the FACS project results. A letter of support of the MUSCLE NoE Project Manager is attached to the proposal.

B3.1. Contributions to standards

In order to maximize protection of the investment and re-use of existing technology, the project will adopt the following official and de facto standards:

- MPEG-4[1]: ISO/IEC standard developed by MPEG (Moving Picture Experts Group),
- MPEG-7[1]: also called "Multimedia Content Description Interface", standardises the description of multimedia content supporting a wide range of application.
- MPEG-21[1]: its general goal is to describe an open framework which allows the integration of all components of a delivery chain necessary to generate, use, manipulate, manage, and deliver multimedia content across a wide range of networks and devices.
- DVB/MHP: ETSI Standard describing the Multimedia Home Platform, enabling the reception and presentation of applications on digital TV receivers (ETSI TS 101 812)
- XML (Extended Mark-up Language)[2]: XML is a set of rules for designing text formats for structuring data. XML is extensible, platform-independent, and supports internationalisation and localisation. XML is fully Unicode-compliant.
- WDSL (Web Service Description Language)[3]: WSDL is an XML format for describing network services as a set of endpoints operating on messages containing either document-oriented or procedure-oriented information.
- SMIL-2.0/2.1: SMIL, the Synchronized Multimedia Integration Language is an XML-based language for describing the logical, temporal and spatial relationships within a multimedia presentation in a presentation-architecture and vendor neutral manner.
- OWL (Web Ontology Language, http://www.w3.org/TR/2004/REC-owl-ref-20040210/) is a language used to describe ontologies that encode shared views of specific knowledge domains, which are common among different parties.

[1] http://www.chiariglione.org/mpeg/index.htm

[2] http://www.xml.org

[3] http://www.xml.org/xml/resources_focus_wsdl.shtml

- C-OWL [55] is a language that extends OWL in order to allow the specification of personalized views of one or more ontologies that encode a single party's understanding (representation of contextualized ontologies).

We will further evaluate three recently defined markup language and annotation standards for time, TimeML [93][94] and opinions/attitudes [95][96] that have been developed and applied in the U.S. natural language question-answering research community.

Furthermore, FACS intends to contribute to the evolution of existing standards. Members of the project consortium are active in several standards areas that relate to the definition of multimedia coordination and distribution infrastructures. CWI (NL) has contributed extensively to the development of the W3C SMIL language, including facilities for core synchronization and presentation transformation support that allow multiple target systems to be supported from within a single document source. We expect that future versions of this language, which has been adopted heavily by the European mobile telephone industry, will be directly influenced by the results of the FACS project. Other project members have been involved in the development of standards for interactive television and for metadata modelling and acquisition. FhG (DE) has had a long history in the support for iTV applications and has contributed to the standardization of structures and protocols for work in the virtual studio domain.

MPEG-7 and MPEG-21 are dominant approaches to standardization of content descriptions. In FACS we focus to extend the functionality of those standards to represent semantic knowledge with the transparent use of domain ontologies. The contributions of FACS are expected both in the areas of semantic Content descriptions and semantic Profile descriptions. We also aim to contribute towards the semantic descriptions of Context.

CWI is a member of the W3C Semantic Web Best Practices group. It is expected that results from FACS will be channelled via this group to the broader Web community.

B4. Description of the consortium

The FACS objective requires a heavy investment in research in a number of inter-related fields that do not constitute the background of any existing research community (neither in Europe not worldwide). We have already pointed out that achievements of the project goals necessitates research contributions in many different areas, namely automated annotation of sound and video; modelling and automatic acquisition of user profiles; modelling of user and data context; modelling of interactive multimedia presentations; modelling of multimedia content description; management and approximate retrieval in a distributed environment of all different types of information and knowledge; automatic synthesis of interactive multimedia presentations and adaptation of the presentations to different delivery devices. These considerations have lead to the forming of the FACS consortium, composed of highly qualified research groups, European or worldwide leaders in these different areas. The creation of this blend of competences is the key factor for the success of the project.

The composition of the Consortium and its dimension is consistent with the scope of the project. The seven partners come from different countries, and have different complementary backgrounds. The EC support will allow their integration in the same Consortium. We have also included Carnegie Mellon University (USA) because of their considerable experience in automated processing, retrieval, filtering, and synthesis of audio and video. They will contribute access to their video processing infrastructure and multi-terabyte corpus for experimentation. The contribution of CMU will also ensure a close link with the US research community for a valuable exchange of knowledge.

Each of the seven partners will employ a project manager with high experience in the conduction of research projects and a scientific leader with extensive research experience. The scientific team of each partner is related to the effort allocated in the project (see the form at the end of the section). All teams will be composed of experienced post-doc researchers and PhD students.

All partners have research laboratories with all the basic equipment needed for the conduction of the project.

At the beginning of the project, each partner will define the team to be allocated to the project. Key personnel will consist of people with permanent position within the different organizations. Each organization either has already the other researchers that will work in the project, or has a precise procedure for the recruitment of new personnel (post-doc researchers, or PhD students). At the beginning of the project, each partner will specify the equipment in their labs and that will be allocated to the project (in part or fully).

Here below, we report the main areas in which the project will operate, together with the partners involved in each area (in bold the partners providing the main contribution), and the type of contribution that will be provided. The partner acronyms are defined in the following subsections.

Partner expertise		**Main contributors**
Scientific		
	Feature extraction from sound and video	UVA, TUWIEN, CMU
	Automated semantic and rhetorical annotation of sound and video	TUWIEN, UVA, CMU
	Retrieval and filtering of multimedia data/annotations	ISTI-CNR, TUC
	User profiling	TUC, ISTI-CNR
	Context and content modelling	TUC, CWI
	Presentation modelling	CWI, FhG
	I-Objects generation	FhG, CWI
Managerial		
	Management of research projects	ISTI-CNR plus all partners
	Technical direction of research projects	ISTI-CNR plus all partners
	Dissemination and Exploitation	All

In the following we will provide a brief description of the competences and role of the FACS partners in the project. Appendix D contains a short CV of key personnel for each partner.

[1] – Consiglio Nazionale delle Ricerche – Istituto di Scienze e Tecnologie dell'Informazione "Alessandro Faedo" (ISTI-CNR)

Role

ISTI-CNR is the coordinator of the FACS project and will provide the support for the project management. ISTI-CNR will lead WP 1 (Research monitoring and delivery) and WP 9 (Dissemination and Exploitation).

With respect to the scientific contribution in the FACS project, ISTI-CNR will lead the work package WP 7 (Management of content, context, and profile descriptions). ISTI-CNR's main contribution will be in the following activities:

- Efficient and effective retrieval and filtering
- Distributed management of user profiles, contexts, and content descriptions
- Modelling descriptions of user needs and automatic acquisition of user profile
- Retrieval and inference language, and query processing

Research Institution Profile

The Institute of Information Science and Technologies (ISTI-CNR, http://www.isti.cnr.it) is an institute of the Italian National Research Council (CNR). The Institute is located in the CNR Research Area of Pisa. ISTI-CNR is the largest Italian Information Science and Technologies research institute. It was constituted in 2002 as a result of a merger between the Istituto CNUCE (CNUCE-CNR) and the Istituto di Elaborazione dell'Informazione (IEI-CNR). The President of CNR has delegated ISTI-CNR to represent the information technology sector in European Research Consortium for Informatics and Mathematics (ERCIM).

The Project will be carried out by the Networked Multimedia Information Systems Laboratory (NMIS) of ISTI-CNR (http://www.isti.cnr.it/ResearchUnits/Labs/nmis-lab/), led by Costantino Thanos, in cooperation with the Knowledge Discovery and Delivery Laboratory (KDD) (http://www.isti.cnr.it/ResearchUnits/Labs/kdd-lab/), led by Fosca Giannotti.

The NMIS Laboratory, over the past 15 years, has gained considerable experience in a number of areas: Design of Object-Oriented Information systems, Information Retrieval, Multimedia Information Retrieval, Access Methods for Similarity Search, User Profile Modelling, Personalisation. The Laboratory is involved or has been involved in many national and EC funded projects, and collaborates with several International Scientific and Research Institutions. Among these it is worth mentioning: FIDE 1 and 2 -ESPRIT Basic Research Action no. 3070 (1989-1995), where the Lab was involved in the design and development of object oriented information systems; ESPRIT HERMES 9141; EuroGatherer IE-8011, where the Lab was involved on the definition of personalised information gathering systems; IST – ECHO (coordinated by NMIS ISTI-CNR), where the Lab was involved in the development of a digital library for audiovisual data, in which metadata are XML encoded; the IST Project Scholnet: A Digital Library Testbed to Support Networked Scholarly Communities (IST-1000-20664) (coordinated by NMIS ISTI-CNR); CLEF: Cross-Language Evaluation Forum (IST-2000-31002) (coordinated by NMIS ISTI-CNR); the D-Lib Competence Center (IST-2001-32587) (coordinated by NMIS ISTI-CNR)**;** the DELOS Network of Excellence on Digital Libraries (coordinated by NMIS ISTI-CNR); the ECD (funded by the Italian government and coordinated by NMIS ISTI-CNR), where the Lab is currently involved in the design and development of efficient XML search engines. Finally, the NMIS Lab is coordinating the IP Project DILIGENT; this project aims at building a Digital Library infrastructure based on the Grid platform.

The Knowledge Discovery and Delivery Laboratory (KDD) (http://www-kdd.isti.cnr.it) is a joint research group of the Institute of Information Science and Technologies (ISTI-CNR) and the Computer Science Department of University of Pisa. In the last decade, the KDD laboratory has pursued fundamental research, strategic applications and higher education in the areas of: Knowledge discovery, Data Mining and Web Mining. A significant activity has been devoted to the automatic extraction of semantics from web resources and their usage pursued within the following three national projects – ECD (funded by Italian Government), MineFaST (co-funded by an Italian Industrial Partner), WEBDIGGER (funded by a regional agency). KDDLAB has been an active partner of the network of Excellence KDNET (2002-2004) aimed at integrating the research and industrial activity in Data Mining at European level. The KDDLAB was responsible for the scientific organization of the 15th European Conference on Machine Learning (ECML) and the 8th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD), which were co-located in Pisa, Italy, September 20-24, 2004 (http://ecmlpkdd.isti.cnr.it).

Key personnel in the project are Pasquale Savino, who will be FACS project director, Giuseppe Amato, that will be the leader of WP 7, and Fosca Giannotti that will coordinate the ISTI-CNR contribution to WP 4. Their CVs are provided in Appendix D.

[2] – Carnegie Mellon University (CMU)

Role

Carnegie Mellon University will mainly contribute to the following activities:

- Multimedia information extraction and indexing
- Capturing knowledge from raw multimedia and web content
- Machine learning applied to image and video feature extraction, classification, searching, and ultimately "understanding"
- Automatic (on the fly) metadata generation as soon as data is available or while data is generated.
- Semi-automatic creation and evolution of ontologies
- Searching and categorization from multiple media types
- Infrastructure for prototyping and system component integration to facilitate experimentation and evaluation

Research Institution Profile

Carnegie Mellon University (CMU) is one of the U.S.'s top private academic research institutions, consistently ranked within the top 25 U.S. colleges and universities overall. Its internationally recognized programs encompass the areas of science, engineering, computer science, liberal arts, fine arts and public and private management. One of CMU's strengths has been its ability to push experimental computer science ideas from theory to practical demonstration, validating new fundamental insights by constructing large prototype systems in which they are embodied. CMU is well known for its interdisciplinary research and centres. In the computer disciplines these include the Software Engineering Institute, The Data Storage Systems Centre, The Information Networking Institute, the Language Technologies Institute, the Human-Computer Interaction Institute, the Entertainment Technology Centre and the Robotics Institute. During the past 35 years it has produced some of the most important breakthroughs in artificial intelligence, operating systems, machine architectures, speech recognition, image understanding, machine translation and digital libraries. CMU has been ranked the top computer science school in the country. The Informedia project at CMU was initiated in 1994 as one of six national Digital Library Initiative (DLI) projects; it has pioneered numerous advances in automated video understanding. The Informedia system automated information extraction from broadcast television, collecting over six terabytes of video with automatically generated metadata and indices for retrieving segments. It applies integrated speech, language, and image processing to both library creation and navigation, including the generation of abstractions and visualizations [97] [98] [99]. Its modularized component structure, subsystem API's and its use of XML and MPEG-7 to communicate video metadata, will facilitate the development of the core FACS subsystems.

Key personnel in the project are Howard Wactlar, who will be FACS project manager, and Dorbin Ng who will be member of the Scientific Executive Committee. Their CVs are provided in Appendix D.

[3] – Stichting Centrum voor Wiskunde en Informatica (CWI)

Role

Regarding the participation in the FACS project, CWI will contribute to the following components:

- Semantic labelling of content related specifically to rhetorical characterisation
- Modelling of I-Object layout and presentation semantics and structure
- Transformation of logical I-Object structures to platform depending representations
- Server-side presentation and interaction control models and implementations
- Client-side presentation and interaction control models and implementations.

Research Institution Profile

Centrum voor Wiskunde en Informatica (CWI) is the national research institute for Mathematics and Computer Science of the Dutch National Science Foundation (NOW). CWI is a private, non-profit organisation. Founded in 1946 (as Mathematisch Centrum), CWI aims at fostering mathematics and computer science research in The Netherlands. CWI receives a basic funding from the Netherlands Organization for Scientific Research NWO, amounting to about 70% of the institute's total income. The remaining 30% is obtained through national research programmes, international programmes and contract research commissioned by industry.

CWI's mission is twofold: to perform frontier research in mathematics and computer science, and to transfer new knowledge in these fields to society in general and trade and industry in particular.

CWI's mission is realized by several means. In addition to the standard ways of disseminating scientific knowledge, for example through publications, presentations at conferences, organization of workshops and exchange of researchers, CWI actively pursues joint projects with external partners, provides consulting services and actively stimulates the creation of spin-off companies. A technology transfer event is organized annually to promote this side of CWI's activities. Also special efforts are made to make research results known to non-specialist circles, ranging from researchers in other disciplines to the public at large. CWI has many contacts with national organizations for applied

research with wide experience in turning research results directly into practical applications. Its researchers are supported by state-of-the-art computing facilities and a library of national importance.

The proposed research will be embedded in two of CWI's scientific themes. The work on I-Object modelling, presentation characterization and server-side/client-side presentation technology will be performed with SEN-5: Convergent Media Infrastructures. The work on rhetorical modelling and characterization will be performed within the theme INS-2: Multimedia and Human-Computer Interaction.

CWI webpage: http://www.cwi.nl/ .

Key personnel in the project are Dick C.A. Bulterman, that will be FACS project manager and the leader of WP 5, and Lynda Hardman. Their CVs are provided in appendix D.

[4] – Fraunhofer – Gesellschaft zur Foerderung der angewandten Forschung e.V. - Institute for Media Communication (FhG)

Role

In this project, FhG will mainly contribute to all aspects of platform independent delivery of I-Objects. For the purpose of WP8 (Delivering presentations), we can provide our experience and the necessary facilities for delivering interactive multimedia to the DVB/MHP and DVB-H Platforms.

On the background of several years of practical experience in multimedia applications targeted towards the internet, interactive TV and (more recently) mobile devices, we are especially interested in developing and evaluating a common framework for multimedia distribution.

Research Institution Profile

The Institute for Media Communication (FhG) of Fraunhofer will participate to the project. The research and development activities of FhG encompass all facets of the new media, including content design, production, distribution, and interaction. The key objectives of the FhG are to expand the range of potential and functionality of the new media, to study their creative and social possibilities, to develop innovative solutions and to open up new fields of application. In cooperation with its partners and clients in business and industry as well as in cultural and educational communities, the FhG implements its scientific knowledge in end-to-end solutions and prototypes tailored to specific user needs and market requirements. Strategic cooperation with renowned national and international universities ensures the high scientific caliber and maintains the cutting edge of FhG technologies. Key topics are virtual environments, interactive TV, media networking, art and technology.

At FhG, computer scientists, engineers, physicists, mathematicians, biologists and geologists work together with media artists and designers as well as experts in law, economics and the humanities. Interdisciplinary teams are often international teams as well since FhG's staff of 120 includes researchers from nearly 20 countries. The FhG has laboratories equipped with the latest in high-end technologies, including a professional virtual studio and a virtual environment laboratory with a CyberStage, a Responsive Workbench and an i-CONE panorama display.

FhG was founded in 1995 as one Institute of the GMD – German National Research Center for Information Technology. FhG became a Fraunhofer Institute through the fusion of GMD and Fraunhofer-Gesellschaft in 2001. The Fraunhofer-Gesellschaft is the leading organization for institutes of applied research in Europe.

The Competence Center for Interactive TV (ITV) of FhG is a leader in developing solutions for digital broadcasting and interactive media applications. With several years of practical experience FhG is a qualified partner for the implementation of interactive TV services. As a strong partner of the public German broadcasters ARD and ZDF FhG.ITV could support the developement of their iTV offers over a decade of years. FhG.ITV is active member of DVB and the Deutsche TV-Plattform. Being a continuous exhibitor on trade shows like IBC, NAB and the German IFA FhG.ITV is heavily engaged in promoting MHP and MHP applications. The JAME iTV Suite, Fraunhofer FhGs authoring and production system for efficient MHP/OCAP applications, completes the offer for broadcaster, content provider and media production companies.

Key personnel in the project are Sepideh Chakaveh, that will be FACS project manager and Joachim Kaeber that will be the leader of WP 8. Their CVs are provided in appendix D.

[5] – Technical University of Crete (TUC)

Role

TUC/MUSIC has extensive experience on all aspects of multimedia document management, multimedia applications and standards. In addition, TUC/MUSIC has deep expertise in semantic modelling, semantic languages, ontologies, as well as user profile and context modelling. In the FACS project TUC/MUSIC will be involved in the modelling of the semantic content of multimedia objects using the international standards (MPEG-7, etc) as well as domain specific

ontologies and context of capturing models. TUC/MUSIC will lead the research effort in the modelling of the user context and personalization, the formulation of user requests, the semantic and syntactic enhancement of those requests with contextual and personalization information, and the matching process. Moreover TUC/MUSIC will contribute to the development of models for intelligent multimedia objects focusing on their behavioural aspects.

Research Institution Profile

The Laboratory of Distributed Multimedia Information Systems and Applications of the Technical University of Crete (TUC/MUSIC **http://www.music.tuc.gr/**) is a centre of research, development and education in the technological fields of Information Systems and their applications in the Information and Knowledge Society and in particular in Technologies, Architectures and Web Services Systems over the Internet, Databases and Knowledge Bases, Information Retrieval, Digital Libraries, Geographic Information Systems, Human Computer Interaction, Multimedia Management Systems, Digital TV Systems and Applications e-Commerce, Tourist, Culture, e-Learning. TUC/MUSIC exists since 1990 and it has participated in more than 40 European and National Projects. In many of these projects TUC/MUSIC was the Prime Contractor and/or Technical Leader. The projects include large Basic Research projects (like the LTR HERMES as Prime Contractor), large projects for the development of innovative software tools and integrated multimedia servers (like the IST UP-TV project in which TUC/MUSIC develops systems for supporting Digital TV services based on the TV-Anytime specification, the ACTS SICMA project in which TUC/MUSIC developed the Kydonia Multimedia Server), large projects for technology transfer (like the STRIDE Multimedia project in which TUC/MUSIC as Prime Contractor has transferred multimedia information systems technology in the major IT Greek companies and academic institutions) and large projects for the development of multimedia applications (like the Cretan TOURnet project, in which the TUC/MUSIC was Technical Leader, that developed a Destination Management System for all the small and medium size tourism related enterprises of Crete). TUC/MUSIC participated in several Excellence Networks of the European Union along with other leading scientific organizations of the Union, as i3-Net in intelligent information interfaces, Kaleidoscope in eLearning, DELOS II in Digital Libraries, and the European Excellence Network IDOMENEUS for the development of multimedia information systems in Europe. TUC/MUSIC participates in the IST project on Digital Business Ecosystems (DBE IP) where it is responsible for the Knowledge Management work packages of the architecture. Knowledge management in DBE develops languages and infrastructures for managing business and service knowledge under the OMG Framework, which are analogous to the Semantic Web languages and infrastructures. DBE develops a p2p architecture for SME cooperation/competition. TUC/MUSIC is member of the Object Management Group (OMG), associate member of the PRO-MPEG Forum and participates actively to the TV-Anytime Forum implementers' workshops.

Key personnel in the project is Stavros Christodoulakis, that will be FACS project manager and the leader of WP 6, and WP 4. His CV is provided in appendix D.

[6] – Technische Universität Wien (TUWIEN)

Department of Software Technology and Interactive Systems (ISIS), at the Vienna University of Technology (TUWIEN)

Role

TUWIEN/ISIS will contribute to the project with its long-standing experience in information space analysis, particularly audio and textual information spaces. As a member of both the DELOS and MUSCLE NoE's TUWIEN/ISIS is deeply rooted within the respective research networks. Our work has focused on feature extraction from a range of data sources, having developed a specific feature set for audio data that focuses on rhythmic characteristics of audio data, which won the Rhythm Track of the ISMIR 2004 Audio Description contest with an average classification accuracy of 82%.

TUWIEN/ISIS will contribute to this project by the development and evaluation of new feature sets for audio indexing, optimized for affective characterization of audio signals. Incorporating its research in the field of text analysis and information retrieval, textual characteristics of audio signals shall be analyzed. These, in combination with visual features developed by UVA, will allow the automatic labelling of multimedia objects according to both their semantic as well as emotional characteristics.

Research Institution Profile

The Department of Software Technology and Interactive Systems at TUWIEN covers research in and development of software and information systems. This includes the broad spectrum of tools and methods which are relevant in the life cycle of software and information systems, beginning from abstract models for problem analysis to the implementation of software products. The department has 3 full professors with a total of 20 full-time faculty positions, with about one third of the faculty being women, as well as about 40 full-time equivalents of third-party funded research positions. Thus, the Institute is engaged in various third-party funded projects with partners from the mobile computing area (e.g. max.mobil), the software area (e.g. Microsoft Research, SAP, Software AG), the banking sector, and others, as well as participating in a national competence centre on e-commerce (EC3).

More specifically, the areas of research include the domains of Data Engineering (Data Warehousing, Data Mining, Non-standard Data Modelling), Information & Knowledge Engineering (Web Semantics, Information Visualisation, Plan Management, E-Learning, Digital Libraries, Text and Audio Mining), Process Engineering (adaptive Workflows, inter-organisational Workflows), Software Engineering (Software Quality Management, Verification and Validation, Security) and Web Engineering (ubiquitous Web Application, Integration of XML & Data Base Systems, E-Commerce).

Our department brings into FACS its long-standing experience in information organization, retrieval, and visualization, both for text as well as specifically audio documents. In the latter domain the department is a member of the EU FP6 Network of Excellence Multimedia Understanding through Semantics, Computation and Learning (MUSCLE), as well as the EU FP6 Network of Excellence on Digital Libraries (DELOS), and it is participating in the respective audio retrieval evaluation competitions of the International Conference on Music information retrieval (ISMIR), having won the Rhythm Classification Contest in 2004.

Key personnel in the project is Andreas Rauber, that will be FACS project manager and the leader of WP 3, His CV is provided in appendix D.

[7] – Universiteit van Amsterdam (UVA)

Role

Regarding the participation in the FACS project, UVA will effectively contribute to:

- content-based retrieval of images and videos
- semantic annotation of images and videos
- advanced human-machine interaction technology for affective annotation and retrieval
- multimedia ontology-based knowledge representation

Research Institution Profile

The University of Amsterdam (UVA) is an academic institution engaged in higher-level education and research. It is one of the largest universities of the Netherlands, both terms of the number of researchers and the students. It is a public institution providing higher-level education to 20,000 students in all fields, covering medicine, humanistic, scientific, and technological areas. UVA will participate in the project through the Informatics Institute of the Faculty of Science. The Institute has a well-established tradition of European program participation on a broad range of topics. On these topics, the institute covers the range from fundamental research to application research in close collaboration with industry. Within the Informatics Institute there are three research labs, one of which is the intelligent systems lab (ISLA). Within ISLA, the research group for Intelligent Autonomous Systems is dealing with representation of geometry, machine learning, robotic and surveillance systems. The research group for Internet Information is doing research in the direction of information retrieval and natural language processing.

The proposed research will be embedded in the ISLA's group on sensory information analysis (ISIS), headed by Prof. Arnold Smeulders. The ISIS group consists of 22 persons, five of which are funded by the University and eighteen from a variety of sources. ISIS has an extensive track record in public-private partnerships. The main research activities are in multimedia information processing, annotation of and retrieval from images and video data repositories, data space visualization, theory of computer vision, color and color invariants in computer vision, multimedia data mining, multimodal human-computer interaction, and cognitive vision.

ISIS webpage: http://www.science.uva.nl/research/isis/isis.html

Key personnel in the project are Nicu Sebe that will be FACS project manager and Arnold W.M. Smeulders that will be the leader of WP 2. Their CVs are provided in appendix D.

B4.1. Other countries

Carnegie Mellon University (CMU) – U.S.A. has been included among the partners in the project, in order to bring to FACS their extensive experience in many research fields of primary importance for the FACS project, such as artificial intelligence, machine learning, speech recognition, image understanding, automatic extraction of low level features and semantic interpretation of audio and video. In particular, the research team participating to the project, is the group that proposed and coordinated the Informedia project at CMU: Informedia was initiated in 1994 as one of six national Digital Library Initiative (DLI) projects, and it has pioneered numerous advances in automated video understanding. Informedia has a modular structure and interfaces which allow for integrating numerous alternative analysis algorithms and metadata type extractions. It also has a multi-faceted user interface that outputs XML, enabling its customization across a broad number of applications and devices. CMU will extend their function and interfaces in ways which render them applicable to the proposed work. Moreover, CMU has done and it will continue in the future seminal research in

the area, mainly on feature extraction, semantic extraction, and interpretation of video, so that a cooperation will greatly help to match the FACS project objectives. CMU will also bring to the project several terabytes worth of (mostly English news) content that will be used to test and validate the new methods of FACS. Finally, CMU will contribute to the dissemination of FACS results. Prof. Howard Wactlar, primary architect and project director of the Informedia Digital Video Library will be a member of the FACS research team.
The budget contribution requested will cover travel expenses for the participation to FACS technical and managerial meetings, and for technology transfer, while no funding is requested for labour effort.
The total budget allocated to CMU is 119070 Euro which counts for the participation to the planned Managerial Board Meetings (two per year), plus at least four technical meetings per year (2 people attending), for a total of 24 person trips in three years. Thus the total is 30 person trips for the entire project. The estimated average cost per person trip is 2700 Euros (all inclusive), which gives a total of 81000 plus 47%, as CMU overheads.

STREP Project Effort Form

Full duration of project

FACS

	CNR	CMU	CWI	FhG	TUC	TUWIEN	UVA	TOTAL PARTICIPANTS
Research/innovation activities								
WP1: Research activity management and monitoring	10.0	1.0	1.5	2.0	4.0	6.0	6.0	30.5
WP2: Feature extraction		8.0	1.0	1.0		35.0	42.0	87.0
WP3: Semantic and rhetoric annotation		10.0				39.0	64.0	113.0
WP4: Modelling and Supporting User Profiles and Contexts	35.0				40.0	6.0	9.0	90.0
WP5: Intelligent multimedia objects modelling			60.0	14.0	22.0	3.0	5.0	104.0
WP6: Development and support for Semantic Multimedia Content and Query Models	30.0		1.0	1.0	40.0	1.0	4,0	77.0
WP7: Management of content, context, and profile descriptions	76.0				15.0			91.0
WP8: Presentations delivery	5.0		72.0	32.0				109.0
WP9: Dissemination and Exploitation	6.0		1.5	1.0	4.0	1.0	1.0	14.5
Total research/innovation	162.0	19.0	137.0	51.0	125.0	91.0	131.0	716.0
Consortium manag. activities								
WP10: Project Management	36	1	1	3	2	2	2	47
Total consortium management	36	1	1	3	2	2	2	47
TOTAL ACTIVITIES	198.0	20.0	138.0	54.0	127.0	93.0	133.0	763.0

B5. Description of Project Management

Monitoring and planning are key issues of the management and co-ordination task, and are a responsibility of the project management structure. The consortium will adopt a simple and effective management system. The main objectives considered when defining the structure and strategy of the project management were to obtain:

- efficiency of the management, internal to the project and with respect to the EC
- high scientific quality of the project approach, activities and results
- high cohesion and cooperation among all partners

Crucial to attain the first objective is a planning and monitoring structure able to promptly track the performance of each ongoing task, identify deviations from the plan and react quickly, defining and implementing corrections and workaround procedures. The second objective requires a high level scientific coordination of all project activities, and the supervision of a highly qualified board that monitors the scientific progress of the project. Harmonization of different research communities toward a common goal is a key and ambitious issue of the project. To facilitate this process, we will organize a project workshop on the project kick-off, with the aim of creating a cooperative environment among all researchers involved in the project and establishing direct contacts among them. The workshop will be repeated every year.

Since information exchange between partners is a key issue for an effective management, a web site containing all project documentation and experimental results and a mailing list will be available to project partners. In addition, regular direct contacts will be established between the Project Director and the members of the Scientific Executive Committee (see Project Organization below). These contacts will be performed via regular information exchanges, e-mails, direct face-to-face meetings, video conferencing and teleconferencing.

B5.1. *Project Organisation*

Three main bodies will oversee the activities in the project:

- the ***Project Management Board*** (**PMB**), for legal/administrative matters,
- the **Scientific Advisory Board (SAB),** to support the PMB in controlling the project research quality,
- the ***Scientific Executive Committee*** (**SEC**), for scientific and technological matters,

The Project Management Board will be supported by an external Scientific Advisory Board, composed of senior researchers. The Project Management Board will be composed by the Project Director (designated by the coordinating partner) and a Project Manager for each of the other partners.

The Scientific Executive Committee will be composed by the Work Package Leaders (WPL) and a representative from all other partners that do not have a Work Package Leader. It is coordinated by the Project Director.

As a punctual monitoring and checking of project results is planned, a strong communication flow is needed. For this reason regular meetings between project participants will be scheduled.

The Scientific Executive Committee will have a video conference call or a direct meeting <u>every month</u> in order to monitor progress and validate activity outputs.

The Project Management Board will meet every 6 months in order to take strategic decisions and monitor project activity.

The Scientific Advisory Board will meet every 6 months immediately before the PMB meetings, in order to provide a report of the scientific progress of the project.

The FACS project management structure is shown in the next figure.

FACS Management Structure

B5.2. Management structure and responsibilities

B5.2.1. Project Director (PD)

The PD will be the official contact point for the European Commission. He will represent the project in relation with the European Commission, report to the European Commission, monitor overall performances of the project, evaluate results according to the work plan and help promote project visibility. He will be responsible for ensuring that all deliverables are sent to the European Commission and will organise review meetings between partners and the European Commission when requested.

The Project Director, with the help of the PMB, will monitor progress against the technical annex, take action if delays occur, monitor and control project costs, distribute information pertinent to the partner consortium, undertake all communications with the European Commission and arrange any necessary ad hoc meetings. Moreover, the Project Director will be supported by a project secretary and project administrator, both hired by ISTI-CNR.

The Project Director will also manage the Scientific Executive Committee.

The Project Director will be Pasquale Savino ISTI-CNR. He brings to the FACS project a powerful combination of technical/scientific expertise and managing capabilities acquired in 20 years of experience at international level. A CV of Pasquale Savino is available in Appendix D.

The Project Director will be fully dedicated to the FACS project.

B5.2.2. Project Management Board (PMB)

The Project Management Board will be set up to deal with all decisions affecting more than one participating organization and all technical decisions that cannot be taken by the Scientific Executive Committee. This Board will include one representative from each partner organisation. The Project Management Board will be responsible for all contractual matters concerning the execution of the contract. It will produce the legal and ethical framework for FACS, resolve problems and ensure that all participants meet their obligations.

If a disagreement occurs within the project, the role of the PMB is to assess the problem and find a solution, if possible by consensus. In the case where no consensus can be found, then a vote among all members of the PMB will decide the conflict. If necessary, the PD will have the casting vote.

Moreover, the PMB will be in charge of overseeing the promotion of gender equality in the project and of science and societal issues related to the research activities conducted within the project.

B5.2.3. Scientific Advisory Board (SAB)

The Scientific Advisory Board is composed senior researches with high reputation in the project research fields. The designated representatives bring to the project their considerable experience in the conduction and coordination of innovative research projects, but who are not directly involved in the project. The SAB will support the PMB by monitoring the scientific progress of all different research activities. In particular, coherent progress towards a common goal will be checked. The SAB will provide guidance with respect to the project strategic orientations and it will identify new research challenges. Finally, the SAB will verify that the project milestones have been matched from the scientific point of view. The SAB members will be appointed during the first PMB meeting.

The SAB will nominate a rapporteur to report the results of SAB meetings to the PMB.

B5.2.4. Work Package Leaders (WPL)

The WP Leader has the responsibility for the contents and timely consignment of the deliverables of his/her work package, as defined in the global project work plan.

The work package leaders will be responsible for the performance of the work packages and will guarantee the accomplishment of the technical objectives. They will provide regularly reports, control the quality and the schedule of the work, and participate actively to technical meetings.

Each WPL will archive the results obtained within his work package (either intermediate or final). This archive will include all working drafts (latest version), the work package deliverables, and minutes of the work package technical meetings.

B5.2.5. Scientific Executive Committee (SEC)

The Scientific Executive Committee is composed of all Work package Leaders plus a representative for those partners that are not leaders of any WP, plus the Project Director. The Scientific Executive Committee will support the Project Director in the supervision of all the activities and all the possible technical matters. In case a consensus is not reached, then a vote among all members of the SEC will decide the conflict. In case of conflict, the PMB will have the casting vote.

B5.3. Management Processes, Mechanisms and Procedures

The entire Management Process will be based on the use of specific tools already tested during the management of similar projects where strong and multiple partnerships were organised and co-ordinated.

B5.3.1. Project web site and management tools

The project web site will be used to exchange information among partners via an internal FORUM, to store information (in the repository section) and to disseminate information to an external public (in the public section of the site).

The consortium will use a well-tested set of management tools, constituted by secure Web Services, associated with the project web site, made available by ISTI-CNR and already used in the context of several ongoing or completed projects.

B5.3.2. Communication strategy

The communication strategy aims at keeping both project partners and the outside world aware of the project status. It is designed to keep all partners fully informed about the project status, planning and all other issues which are important to the partners in order to obtain maximum transparency for all those involved, thus increasing the synergy of co-operation. This will primarily be achieved via the Internet, thanks to a web access established for internal information exchange; in addition communication on specific issues will take place via Videoconference. An external web representation will be set up, that will provide information on the achieved results, such as publications, reports, workshops, and contributions to standard.

The communication strategy will also include liaison with parties outside the consortium such as other relevant EU funded RTDs, and consortiums active in the field.

B5.3.3. Reporting Requirements, Reporting Periods and Project Milestones

The Project Director will be updated by e-mail with technical notes, changes in activities, updates and actual progress against plan. Reports to the European Commission will be as per contract and its annexes, with acceptance by positive notification from European Commission to the Project Director. Documentation will be in English, using MS Word, MS Excel for financial tables, etc. and MS Project for bar/PERT charts, milestone checks, resource allocation and calendar. Documents will be identified with the FACS acronym and logo, contract number, date, unique document name and number. Issue numbers will be recorded in the Master Document Index and will be incremented when (and only when) a change has been made. A change control system will be operated for all controlled documents in accordance with the change management plan and quality plan.

Project reporting depends on the scientific, financial and certification periods for 6FP projects:

- Scientific Periods: every 12 months
- Financial Period: every 6 months
- Certification Periods: every 12 months.

Particular care will be taken to respect the milestones, which are necessary to satisfy the dependencies between the various work-packages and sub-work-packages. In the case of delays, emergency actions may have to be taken by the PMB to put the project back in track.

B5.4. Consortium Agreement

To cover legal issues relating to roles and responsibilities of the project participants, project management, ownership, commercial rights, exploitation, dissemination of the project results, confidentiality and intellectual property rights, a Consortium Agreement will be signed at the beginning of the project by each partner. The specific provisions will comply with the general rules set out in the contract signed with the European Commission. To speed up this phase, the partners have already begun discussing the content of the CA during the project submission phase. It is expected that the final version of the Consortium Agreement will be presented at the project kick-off meeting for the signature of all partners.

The Administrative Co-ordinator will prepare, update and manage the consortium agreement and distribute it to the participants for signature plus a copy to the European Commission.

B6. Detailed Implementation plan

B6.1. Structure of the Implementation plan

B6.2. Introduction

As stated in section B1 "*FACS aims at studying and developing the models and software components needed to automatically create and deliver personalized and context-aware interactive audio/video presentations*". This objective will be achieved by developing techniques that go beyond the current state-of-the-art related to:

- Representation of multimedia content descriptions
- Representation of user preferences and context
- Automated content annotation
- Automated presentation authoring
- Distributed retrieval and filtering of content descriptions and user preferences

These techniques will be developed by pursuing and combining research on the extension of existing models and component-level research, and supporting and exploiting it with management, dissemination and exploitation activities.

The overall structure of the project activities is shown in Figure 2. Activities are grouped as follows:

- *Research and technological development activities,* as discussed in detail in Section B6.3 below, aim at increasing knowledge and advancing the current state-of-the-art in the fields of interest covered by the project. They are structured as follows:
  - *Component-level research* activities will primarily investigate the topics of fully automated content description of media clips, content description matching, fully automated generation and delivery of multimedia interactive presentations.
  - *Modelling* activities will provide strong base to the component-level research. Topics related to user profile and context modelling, presentation structure modelling, and content description and matching rules modelling will be studied. Tools to deal with the defined models will be developed as well.
  - Both component-level research and modelling activities will be supported and guided by a *research monitoring* activity. This activity, besides the specification of a global architecture to be used as a reference for the various research activities, includes continuous monitoring of the state-of-the-art, evaluation of research results, and risk analysis.
- *Innovation-related* activities consist mainly in dissemination and exploitation of the scientific development and results through various means such as scientific publications, participation and organisation of scientific and technology transfer workshops, as discussed in Section B6.4 below.

Figure 2: Structure of the implementation plan

- *Consortium management* activities include project administration and resource monitoring, work plan monitoring, and contacts with the European Community, as explained in Section B6.5.

The project is thus subdivided into the following work packages.

Activity type		Work package
Research and technological development	*Research monitoring*	WP1: Research activity management and monitoring
	Component-level research	WP2: Feature extraction WP3: Semantic and rhetoric annotation WP7: Management of content, context, and profile descriptions WP8: Presentations delivery
	Modelling	WP4: Modelling and Supporting User Profiles and Contexts WP5: Intelligent multimedia objects modelling WP6: Development and support for Semantic Multimedia Content and Query Models
Innovation-related activities		WP9: Dissemination and Exploitation
Consortium Management		WP10: Project Management

B6.3. RTD activities

The automatic generation of personalised and context-aware multimedia presentations (I-Objects) requires, as illustrated in Figure 3, to

1) automatically describe the semantic and rhetoric content of multimedia documents,
2) select relevant media clips that match automatically generated content descriptions with the user needs described in user profiles,
3) automatically compose professional quality presentations with the media selected clips

Distinct component-level activities will be carried out in the project to realize the above functionality. These activities will be supported by modelling activities aimed at formal definitions of the frameworks that the component-level research will rely on. Both component-level and modelling activities will be continuously monitored to ensure awareness of the evolving state-of-the-art, coherence among the various activities, and achievement of the objectives.

Figure 3: Workflow for automatic generation of personalised presentations

B6.3.1. Modelling

Modelling activities will produce a solid framework where component-level research activities can be carried out. All work packages that implement these activities will include a preliminary task devoted to the detailed specification of the research activity carried out.

Three specific modelling aspects will be taken into consideration:

1) Modelling descriptions of user needs and contexts
2) Modelling descriptions of media content
3) Modelling Intelligent Objects (I-Objects)

Modelling descriptions of user needs and contexts

Our aim is to automatically select media clips for specific users. In order to do this, we need to effectively represent user interests so that we can build queries to retrieve/filter the appropriate clips. We will develop advanced models for personalisation and contextualisation, which will be represented using knowledge representation structures, which take into account domain-specific ontologies and international standards. User profiles and contexts will be dynamically adaptable by exploiting external input such as user interaction history.

This modelling activity will be carried out in work package "**WP4:** Modelling and Supporting User Profiles and Contexts". The work package includes specific tasks for

- profile modelling,
- context modelling and dynamic adaptation of user profiles

Modelling descriptions of media content

We will investigate techniques that analyse multimedia documents to automatically describe their content. The descriptions generated will be matched against the user profiles to select media clips for specific users. In order to effectively support this analysis and the matching step, we need to model how the description of content will be represented. We must take into account both types of information that can be automatically inferred by media content analysis and the type of information actually needed to identify relevant material, on the basis of the descriptions that can be represented in the user profiles and contexts. We will also investigate techniques for expressing queries on the content descriptions defined.

This modelling activity will be carried out in work package "WP6: Development and support for Semantic Multimedia Content and Query Models". The work package includes specific tasks for

- defining ontologies and metadata,
- modelling queries on ontologies and metadata
- processing strategies for the high-level queries

Modelling Intelligent Objects (I-Objects)

We want to build personalised professional quality interactive presentations (I-Objects) from media clips selected according to user profiles and contexts. The I-Objects will be built according to templates of presentations corresponding to various presentation styles. This activity will model both the structure of the presentation templates and the I-Objects. The set of presentation templates will be created manually according to a specific application scenario. They will be annotated so that an appropriate template can be chosen in correspondence to a user profile.

In order to create an I-Object from a presentation template and a set of candidate media clips, we need to decide which specific clip should be used in correspondence to each element of the template. To help in this process, when templates are created, their elements are annotated with rhetorical descriptions, which model their intended emotional/affective message. During content analysis of media clips, they are also annotated with such rhetorical information. The rhetorical information is used to decide about the binding of elements of templates with specific media clips. This activity will model the rhetorical descriptions by taking into account, simultaneously, the type of rhetorical information that can be automatically extracted from media clips, and the type of rhetorical information that is useful to describe professional quality presentation templates.

Intelligent objects will have content metadata, state, behaviour and they will be capable of interoperation with the environment (user interactions, user profiles and contexts). They will be capable of modifying the environment (user profiles and contexts) and being modified by it (object behaviour and presentation). The interoperability with the environment will be based on ontologies (both some object states and content metadata may map to user context and preference ontologies). The intelligent objects will have logic to facilitate such import and export interactions. Many types of complex applications in different knowledge domains can be built utilizing such intelligent objects (educational applications in different domains, cultural, etc.).

This set of modelling activities will be carried out in work package "**WP5:** Intelligent multimedia objects modelling". The work package includes specific tasks for

- Modelling and matching rhetorical descriptions
- Modelling presentation templates and matching with the profiles/contexts of users
- Modelling the structure of I-Objects and their active behaviour

B6.3.2. Component-level research

Component-level research activities will focus on investigating technologies beyond the state-of-the-art to support the generation of personalised interactive multimedia presentations. All work packages that implement these activities include a preliminary task, which will produce a detailed specification of the research activity to be carried out, and a final task devoted to the evaluation of the research done in the work packages.

Three challenging topics will be investigated:

1) Automated content description

2) Management, retrieval, and filtering of descriptions

3) Automated generation and delivery of interactive multimedia presentations

Automated content description

We will develop techniques for automated content annotation of image, video, and music media. Content annotation will be performed via an analysis of the visual and musical information. We will not develop new text recognition, or speech to text approaches, given that well-established techniques already exists and approaches derived from text retrieval can be used to index content. Instead, we aim at advancing the technology in the automated generation of annotations of the visual and musical content. Annotations will be generated in terms of features, semantics concepts, and rhetorical concepts. Features refer to objectively measurable characteristics, such as colour histograms, or audio frequencies, or higher-level conceptually significant descriptors that are generated on the basis of advanced analysis of extracted features. Semantic annotations describe "what" is in the media clip. Rhetorical annotations describe the intended emotional/affective message of the media clip. Semantic and rhetoric concepts are obtained by using machine learning techniques on extracted features.

Tests will be carried out on selected datasets that cover a fairly large spectrum of possible cases, as for example documentaries, news, and animations. These datasets will be drawn from public sites that contain open, non-restricted media.

The activities related to the development of techniques for feature extraction will be carried out in work package "**WP2:** Feature extraction". This work package includes specific tasks for

- Selection and collection of content to be used for testing and result validation
- Video and image feature extraction
- Audio and music feature extraction

The activities related to the development of techniques for high-level annotation will be carried out in work package "**WP3:** Semantic and rhetoric annotation". This work package includes specific tasks for

- Automated extraction of semantic annotations
- Automated extraction of rhetorical annotations

The annotations generated will be represented according to the models defined in "**WP6:** Development and support for Semantic Multimedia Content and Query Models", for the semantic annotations, and "**WP5:** Intelligent multimedia objects modelling", for the rhetorical annotations.

Management, retrieval, and filtering of descriptions

FACS results are expected to be used in applications in which the information is highly distributed. Several content providers (including end-users sharing home pictures and videos) continuously produce and offer multimedia material to authorised users. Users have personal profiles that can be maintained locally on their devices, and/or on servers of various service providers, while users subscribe to the services. All distributed information should be transparently accessible, searchable, and filterable. Given the inherently semi-structured nature of the information, and the trends observable in several standardisation initiatives, we will investigate techniques for managing, searching, and filtering XML represented data in highly distributed repositories. In particular, we will design approximate search/filtering algorithms and access methods for highly distributed data, exploiting the GRID service technology.

The activities related to the development of techniques for low-level annotation will be carried out in work package "WP7: Management of content, context, and profile descriptions". The work package includes specific tasks for

- Distributed management of profiles, contexts, and content descriptions
- Distributed approximate match filtering
- Distributed approximate match searching

Figure 4: Relationships among work packages

Generation and delivery of interactive multimedia presentations
After the selection of media clips and of a presentation template that matches a user profile and context, a presentation of professional quality is built. This consists in creating an I-Object by binding the selected media clips with elements of the presentation template, on the basis of their rhetorical descriptions. Depending on the device on which the presentation is going to be played, the presentation will be subject to some transformations that take into account the possible limitations of the end-user players. Specific techniques for client-side rendering of I-Objects will also be investigated.

The activities related to presentation delivery will be carried out in work package "**WP8:** Presentations delivery". The work package includes specific tasks for

- Investigating how to bind media clips with elements of presentation layout
- Performing server-side transformation of the presentation
- Rendering presentations on end-user devices

B6.3.3. Research activity monitoring

The research activities carried out in the various work packages will be constantly monitored. Surveys of the state-of-the-art will be performed at the very beginning of the project and also be done during the project development. A global functional architecture will be also defined at the beginning of the project, and constitutes a precise reference for each research activity. This will also allow the identification of the interaction between the various activities. The monitoring activity will also analyse the results obtained, manage risk detection, and prepare contingency plans to apply appropriate corrective actions.

These activities will be carried out in work package "WP1: Research activity management and monitoring". This work package includes specific tasks for

- Periodic monitoring of the state-of-the-art
- Specification of a functional architecture
- Evaluation of project results
- Risk analysis and contingency plans

B6.3.4. Relationships and dependences of RTD activities

Relationships among work packages are depicted in Figure 4, and in the Pert diagram presented later on. Specifically we have that

- "WP6: Development and support for Semantic Multimedia Content and Query Models" gives input to "WP7: Management of content, context, and profile descriptions" for the type of information that should be managed, with respect to the content descriptions, and the type of queries that should be processed on distributed data. It also gives input to "WP3: Semantic and rhetoric annotationfor the representation of the generated semantic annotations.
- "WP4: Modelling and Supporting User Profiles and Contexts" gives input to "WP7: Management of content, context, and profile descriptions" for the type of information that should be managed with respect to the

profiles and contexts. It also gives input to "WP6: Development and support for Semantic Multimedia Content and Query Models" for the strategies for generating and expanding queries.

- "WP5: Intelligent multimedia objects modelling" gives input to "WP8: Presentations delivery" for the representation of presentation templates and I-Objects. It also provides input to "WP3: Semantic and rhetoric annotation" for the representation of the generated rhetorical annotations.

Component-level research activity mainly produces final results. With the exception of "WP2: Feature extraction" which also produces input to "WP3: Semantic and rhetoric annotation", given that semantic and rhetoric annotations are mainly generated by analysing extracted features, and "WP6: Development and support for Semantic Multimedia Content and Query Models" that also model how to represent features.

"WP1: Research activity management and monitoring" will analyse the results of the project by mainly considering final and progressive results of component-level activities. All component-level work packages periodically evaluate their intermediate results and communicate them to WP1 for the global evaluation of intermediate and final project results, the risks analysis, and the coordination of specific contingency plans. The communication of the intermediate results, produced by the various work packages, will be communicated to WP1 by means of *internal deliverables*, which are separately listed in the work package forms. As a consequence of the result assessment and new emerging state-of-the art techniques being proposed in the literature, work packages may update their research specifications.

B6.4. Innovation-related activities

Any activity related to the promotion of project results internally and externally to the consortium will be carried out here. The achievements and results will be presented into the most appropriate form and publicised via various means (e.g. Web site, journals, conferences, exhibitions, etc-). Specialized workshops will be organized in order to present project results to the scientific and user communities. In particular, a cooperation with the MUSCLE NoE and with the DELOS NoE, where partners of FACS are well represented (consider that ISTI-CNR is also coordinator of the DELOS NoE), is foreseen (see the attached letters of support in Appendix E). Furthermore, a newsletter presenting the most relevant results will be published and distributed by e-mail and RSS feed.

These activities will be carried out in work package "**WP9:** Dissemination and Exploitation". The work package includes specific tasks for

- Planning of the exploitation and dissemination
- Production of promotional material
- Organising thematic workshops related to the project activities
- Exploitation of the scientific knowledge

B6.5. Consortium Management

"WP10: Project Management" guarantees an effective project management organisation, monitoring regularly the project achievements with respect to the established deadlines and funding, co-ordinating partners to ensure that all tasks are efficiently performed and ensure a regularly communication towards the EU Commission. For more details see part B5 and description of "WP10: Project Management".

B6.6. Risk Analysis and Management

This section describes the main risks faced by the project and outlines the approach that will be adopted to recognize potential problem areas at the earliest possible moment so that the necessary actions can be taken and solutions identified. Specific risk management strategies are reported in the description of each work package related to RTD activities in Section B6.8 below.

The main risks in the project can be generated from the following situations:

- **Specific research objectives are not achieved or appear unachievable**. The planned research objectives are ambitious and it is possible that some of the activities that are planned do not fulfil (in total or partially) the objectives stated in the proposal.

- **Research work becomes obsolete**. The field addressed in the project is highly dynamic and many research teams are working in it. Other teams may achieve part of the results planned in the project, before of the project team.

- **Interest in specific research objectives is reduced.** As a result of the research done in the project, or done by other groups, we may realize that some of the objectives of this project have become less significant.

- **Alternative technologies.** The main project objectives, and as a consequence all the scientific objectives, are strictly related to the technologies that are adopted by the TV and Internet production market. Since we are considering a ten-year vision, the risk is that the technologies adopted in these markets evolve in different directions in the meantime.
- **New emerging standards.** Similarly to technological evolutions, the emerging of new standards may result in obsolescence for some of the project objectives.

The FACS Project has several mechanisms to deal with these different eventualities:

- **Continuous monitoring of project research**. Day-by-day monitoring of all research activities will be done by the WP Leaders who will report on the status of the research work every month to the Project Director. WP 1 (Research activity management and monitoring) will analyse the results of the project mainly in consideration of progressive and final results of component-level activities. The Scientific Advisory Board will monitor the quality and the adequateness of research activity. Furthermore, on-going and continual relations will be maintained with other research groups working in the field, through participation in international conferences and through the organization of FACS research workshops. Finally, FACS members will participate in specialized international industrial events and the Project will organize workshops for companies that are potentially interested in the use of the FACS results.
- **Continuous monitoring of the scientific and technological evolutions**. WP 1 (Research activity management and monitoring) will perform a continuous monitoring of the state-of-the-art, in terms of research work done by other research groups, analysis of technological evolutions, and analysis of standards evolution.
- **Organization of research work in multiple phases**. Component-level work packages are organized in two phases; at the end of each phase, the intermediate results are evaluated. Moreover, in WP1 a global evaluation of intermediate and final project results will be made, according to the results provided by different work packages.
- **Project Milestones**. Four milestones have been defined (see section 6.7) in order to periodically check the achievement of planned results.
- **Managerial mechanism for a prompt reshaping of the project**. If deviations from the Technical Annex occur, or if any change in the research program is needed, a specific task of WP 1 will define contingency plans. The Project Director can activate the Scientific Executive Committee (which meets every month) to react to unexpected problems and will either approve the contingency plans prepared within WP 1, or will prepare independent plans. The Project Director will call for an extra ordinary Managerial Board meeting, for prompt approval of the new plans.

B6.7. Project execution approach

All research activities include an initial specification of research objectives. Research work is subdivided into two main parts, each one completed with an evaluation of the results. The second part will be preceded by a revision of the specification, on the basis of the evaluation of the results and of inputs deriving from "**WP1:** Research activity management and monitoring".

At a global level, the project is subdivided into four main phases, each one resulting in a major project milestone.

- *Phase 1:* starts at month 1 and ends at month 6. Produces Milestone **M1**. This phase is devoted to the definition of the basic framework for the project execution, through the assessment of the state-of-the-art, the definition of the global reference architecture and the detailed specification for the various activities.
- *Phase 2:* starts at month 7 and ends at month 18. Produces Milestone **M2**. During this phase both modelling activities and component-level research activity start. Most component-level research activities also produce some preliminary prototypes.
- *Phase 3:* starts at month 19 and ends at month 27. Produces Milestone **M3**. This phase concludes the modelling activity. Most component-level research activities continue their development.
- *Phase 4:* starts at month 28 and ends at month 36. Produces Milestone **M4**. It is mainly devoted to complete the component-level research and to assess the final project results.

A detailed list of the project results expected in correspondence of the various milestones is given below:

Milestone	Expected results
M1: Month 6 **Description:** Definition of the basic framework for project execution	*First assessment of the state-of-the art* *Global functional architecture specification* *Selection of multimedia content for tests* *Detailed research specification for WP2, WP3, WP4, WP5, WP6, WP7*

	Dissemination and exploitation planning *Production of promotion material*
M2: Month 18 **Description:** Production of preliminary prototypes	*Preliminary feature extraction techniques* *Preliminary semantic and rhetoric annotation techniques* *Preliminary distributed repository for content, contexts and profile descriptions* *Evaluation of current results* *Detailed research specification for WP8*
M3: Month 27 **Description:** End of modelling activity	*Models for descriptions of user contexts and profiles* *Model for description of contents* *Ontology integration middleware* *Models for presentation templates* *Model for rhetorical descriptions* *Model for I-Object representations* *Collection of presentation templates*
M4: Month 36 **Description:** End of component-level research and project assessment	*Techniques for binding media with presentation templates* *Distributed repository for content, contexts, and profile descriptions* *Feature extraction techniques* *Final evaluation of project results* *Semantic and rhetoric annotation techniques* *Techniques for interactive generation of contexts and profiles* *Retrieval and inference language and query processing* *Approximate matching and filtering techniques* *Techniques for server-side transformation of I-Objects* *Techniques for client side rendering of I-Objects* *Environment for I-Object creation and access* *Production o promotion material* *Organisation of workshops*

B6.8. Detailed Work package description

B6.8.1. WP1: Research activity management and monitoring

This work package is intended to provide the guidelines for the scientific and technical coordination of the project and to collect and integrate the scientific results obtained in the different work packages.

In order to achieve this goal, initial work will be dedicated to the definition of the global "vision" of the FACS project and to the specification of the research advances that are needed to support this vision. This will constitute the input to all research oriented work packages.

As soon as results on modelling and on component-level research will be available, a reference architecture to support the automated synthesis of personalised and context-aware interactive presentations will be produced. According to the general approach of the project, a preliminary reference architecture will be delivered at month 18. Furthermore, the evaluation of the achievements obtained, and the technological and scientific evolutions occurred, will constitute the input for a revision of the project "vision" and the specification of the research advances. At month 36, a final evaluation of project results and a final reference architecture will be delivered.

The work package will also provide a detailed state of the art on the different research areas addressed in the project. This state of the art will be refined during the project lifetime, while a final version will be delivered at the end of the project. This state of the art will be used inside the project in order to define a common view on different research topics, but it is also intended to be a significant result for the entire research community and for communities of potential users of FACS results.

The continuous monitoring of project research work, and its comparison with the scientific evolution will enable to recognize in the early stage any risk that may arise in the scientific and technical approach – e.g. some planned research is out of date, new standards that may influence the project research, etc. (more details are given in section 6.6, and in the description of different work packages). This will permit to promptly prepare an appropriate contingency plan. Depending on the importance of the problem, the Project Director and the Scientific Executive Committee may be activated.

B6.8.2. WP2: Feature extraction

The objective of this work package is to investigate and develop new feature extraction techniques for feature based searching and filtering of multimedia content. In addition, this work package will compare and contrast the information

loss when the usually high dimensional feature vectors are reduced in dimensions. The features developed in this work package will also be used by the techniques developed in "**WP3**: Semantic and rhetoric annotationto generate semantic and rhetorical annotations of media documents.

A preliminary task of this work package will be selection of multimedia content to be used for the experiments. The selected content will also be updated after the first half of the project. The objective is to gain access to both copyright-free material for external demonstrators as well as potentially copyright-protected material with clearance for research purposes or even serving as a baseline for targeted demonstrators. Resources under consideration are:

- The TRECVID video collection [67] that is an evolving archive of video material. At present the archive consists of about 500 hours of MPG1 video, ranging from public hearings to documentaries and news. The current focus is on broadcast news from English, Chinese, and Arabic sources.
- The Informedia video collection [92]. This is an active digital video library, developed by CMU, which has over 6.000 hours of MPG1 videos associated with automatically generated metadata. The majority are on broadcast news from English, Chinese, and Arabic sources. This collection is made available to the project by CMU.
- Test collections of the ISMIR audio description contests [39]. These will be used for standards-based evaluation with respect to audio indexing, analysis, and retrieval.
- The audio and video holdings of the Internet Archive [38]. This is a repository of old copyright-free material as well as newly produced free music and videos, hosting about 20.000 pieces of music recordings, as well as several thousand movies, ranging from animations, to interviews and documentary material and to free movie productions.
- The material published under the Creative Commons license [30]. This is a machine-readable license specifying to what type of use specific material may be put to. Several open multimedia repositories, such as Soundclick or Morpheus are based on material published via this licensing scheme, and which in many cases will also allow the results produced on top of this material to be published again under the Creative Commons license. Another resource is Legal Torrents [49], a collection of legally downloadable, freely distributable creator-approved files, from electronic/indie music to movies and books.
- A further stream of investigation will include the use of public *podcast* feeds [57] as well as public Internet radio stations to be integrated in media object generation.

Based upon these – and other – open source multimedia repositories, as well as through consultation with copyright holders, such as national archives or national broadcasting corporations, a comprehensive amount of material shall be acquired to allow testing and evaluation of the proposed approaches in this project.

The field of content-based visual and acoustic retrieval is very active [18] [41]. However, information from multiple sources and media (video, images, audio, and text) can be connected in meaningful ways to give us deeper insights into the nature of objects and processes. Fergus' work on automatically learning concepts [62] is inspirational for the techniques that will likely prove to be effective. We plan to use more advanced features, larger data sets [40] and machine learning with very partial annotation [37]. The specific focus provided by directly visual observables, will enable us to automatically identify the so-called frames or conceptual structures which underlie these observables, allowing for high precision information extraction, which in turn provides a novel and high-quality method for populating an ontology. To this aim we will arrange detectors of directly observables in a multimedia ontology. Small textured patches like grass, floor coverings, walls, ceilings, skin, a face, cloth are essential to start the perception-expectation cycle as they permit immediate recognition. They show characteristic textures when sampled with scene-invariant features regardless of context. This is the core of an ontology of directly visual observables, an extension of an early study [48] [17]. Similarly, for audio, directly observable acoustic features, such as the dynamics of audio, rhythmic patterns, energy characteristics will be investigated and combined with more advanced detectors, such as instrument identification. The directly observable detectors may be machine-learned while reducing the amount of annotation on the examples by active learning of one-class classifiers in the knowledge that machine learning provides robustness in real-world data analysis. We will also investigate the integration of the multimedia ontology with a text-based ontology. In a dictionary only a fraction of the words has a visual counterpart. The other words are abstractions. Hence, it is interesting to see how visual and acoustic descriptors relate to standard word ontologies such as WordNet [35].

Specifically we will focus on the acquisition of features from media based on invariant representation of image object and image texture features. Research will be performed on integrating many pieces of information that can be extracted from a video such as sound and image segmentation, speaker segmentation. Several approaches to feature extraction from audio signals will be explored and enhanced, with a specific focus on combinations of specific feature sets, as well as the extraction of features particularly suitable for describing the emotional content of an audio signal. Specifically, feature sets based upon spectral analysis, using Mel Frequency Cepstral Coefficients (MFCCs), Rhythm and Pitch histograms, Roughness, as well as spectral energy features such as spectral flux, roll-off, will be investigated, incorporating psycho-acoustic models. Advanced weak detectors will be applied to the individual media objects to extract direct observables, which in turn will be ontologically connected.

Risk management

We identify a number of significant risks for the task of feature extraction. First of all, it is possible that the features we are implementing do not provide an answer to the intended applications. To address this problem we intend to develop invariant and generic features that will have a far larger applicability. Moreover, the use of machine learning techniques will gear the features set to the specific application domain. Second, there is the problem of computation time. To address this we consider the use of parallel computing paradigms to allow feature computation in acceptable amount of time. Finally, our features may not be sufficiently discriminative and effective when are extracted from low-resolution/low-quality multimedia signals. As a consequence, we will have to impose limits on which our features are still meaningful.

B6.8.3. WP3: Semantic and rhetoric annotation

A major problem in multimedia research has always been the gap between the research in the areas of automatic information extraction and the research in retrieval and presentation of multimedia data. The research in automatic information extraction was concentrating on the (difficult) problems of extracting features from multimedia for indexing purposes. Recent efforts in this field emphasize extracting semantics like recognizing the persons that participate in a picture or video, presence of voice or certain instruments in audio. On the other hand, multimedia retrieval needs retrieval of multimedia parts based on the semantics of the real world described in them (like the events in a soccer game, the views expressed by a person in an interview, or the semantic structure of an art performance) or even in the common characterization of the multimedia pieces as a form of art (like pop music). Additionally, it increasingly also need retrieval based on the feelings and emotions that they evoke (like sad, lively, pleasant). There is a very big gap between those two types of multimedia document descriptions, which we call the **multimedia semantic gap** [63].

Based upon the analysis of features extracted from the multimedia signals, as investigated in **WP2**: Feature extraction", higher-level semantic and affective information will be derived. This can be achieved both by individually analyzing specific feature sets, as well as specifically by combining different feature sets across modalities.

In order to assign respective class labels, machine-learning techniques will be employed to produce weak detectors. Clustering algorithms, specifically topology-preserving mappings such as self-organizing maps [66] will form a basis for the identification of concepts as well as will allow gaining intuitive insight into the feature space and semantic characteristics within. Bayesian Networks and Support Vector Machines [51] as state-of-the-art techniques for classification will be employed to identify and classify multimedia elements into respective categories. This will allow automatically assigning a set of metadata descriptions to the media clips.

These descriptors will include both semantic as well as emotional descriptions, which, in turn will form the basis for rhetorical annotation. Particularly, this will require detailed analysis of the suitability of currently proposed feature sets for different types of multimedia content for classification into affective categories. Current feature extraction techniques focus on content-based retrieval of objects, while tending to put less focus on emotional characteristics of the object in question. It remains to be investigated, in how far these feature sets are capable of discriminating between emotional categories, and which additional characteristics need to be extracted from the object to optimize emotional profiling of the objects.

The affective (rhetorical) content of a video clip or piece of music can be seen as the amount and type of affect (feeling, emotion, and mood) that characterize the multimedia content. As opposed to the semantic annotations extracted in the previous task, the affective analysis steps are required to re-filter the semantic annotations according to the user preferences. A human affective response is usually represented on multiple dimensions or scales (e.g., valence and arousal). Valence (i.e., type of emotion) is typically characterized as a continuous range of affective responses or states extending from pleasant or "positive," to unpleasant or "negative." Arousal (i.e., intensity of emotion) is characterized by affective states ranging on a continuous scale from energized or alert, to calm or peaceful. Typically, arousal and valence values are plotted together to generate a 2D affect curve, which can be seen as the most complete representation of the affective content of a video and can be obtained automatically. In order to obtain the affective content representation, models need to be developed for the arousal and valence time curve. These models fulfil the task of deriving arousal and valence values from the features computed in a video clip or piece of audio.

One interesting visual feature in the context of affective video content is motion. Research results show that motion in a television picture has a significant impact on individual affective responses.

Various effects present in sound may bear broad relations to the affective content of a multimedia object [89] [90]. In terms of affect dimensions, the loudness (signal energy) and speech rate (e.g., faster for fear or joy and slower for disgust or romance) are often being related to the arousal, while the inflection, rhythm, duration of the last syllable of a sentence, voice quality (e.g., breathy or resonant), as well as the pitch-related features (pitch average, pitch range and pitch changes), are commonly related to valence. Varying shot length is also a good example of an editing effect that can be put in relation to the affective content of video.

Our goal is to use the features extracted in "**WP2**: Feature extraction" for the individual modalities to detect and assign semantic and affective information (labels) to the multimedia material. Based on the experiments performed by Hanjalic and Xu [88], we expect to obtain in this way a reliable affective video content representation.

Risk management

Extracting features for affective annotation is a rather new field in video and music indexing, probably calling for totally new, or differently optimized feature extraction techniques. Furthermore, the set of affective categories, their interrelationships, are not clearly defined for unstructured video and audio representations, which is so far focusing mostly on semantic labelling like genre categories, artist identification, or instrument detection.

When developing sets of affective descriptors, we plan to borrow from film and music theory, where different concepts of describing emotional moods of certain movie/musical genres exist, trying to adapt and merge these. Furthermore, intensive user studies will be used to evaluate the derived categories, and to evaluate the resulting performance with different user groups.

B6.8.4. WP4: Modelling and Supporting User Profiles and Contexts

We use the term User Context to denote all the information that is maintained by the system in order to facilitate the retrieval, reasoning, interaction, and presentation of personalized multimedia documents to the user. In FACS contexts may be structured, and a user context may be named, it may have a context-specific functionality, and it may be related to user preferences, user states and ontologies for the context. Contexts are used for enhancing semantic retrieval of information, but also for semantic interaction with the retrieved multimedia objects. Typically, this information is kept in the user station but depending on the particular implementation and business scenario (and with appropriate privacy guarantees) parts of the information will be transmitted to the information providers' site.

We will provide a user environment that supports personalized contextual ontologies which are named representations of knowledge domains that are personalized to the user interests and experiences. They may for example be parts that were extracted from much larger domain ontologies that reflect the interests of the users in the particular domain. In this respect they are a first approximation to user preferences. They may be personalized by renaming concepts, reducing attributes, keeping a subset of the concepts in a rooted hierarchy, adding new concepts that relate to some ontology concepts with various relationships etc. In general, personalized contextual ontologies may map to more than one ontology. We plan to use C-OWL [55] to reflect such mappings. (C-OWL allows expressing generic directional mappings that express the set relationships between the instances of concepts, instances). The personalized contextual ontologies may relate to each other with hierarchical relationships as well. We will design and implement an interactive visual editor that will allow the easy creation and maintenance of personalized contextual ontologies. The users may use concepts, rules or entities from the personalized contextual ontologies in order to formulate queries or preference profiles and the system will automatically convert them to queries or preference profiles that include appropriately weighted concepts from the common ontologies.

Personalized contextual ontologies can be seen as first approximation to user preferences in a particular domain of knowledge and can be used directly by the users with the interactive ontology editor to generate preference profiles. To describe the content metadata for the multimedia objects we will be using MPEG-7 transparently expanded with domain ontologies (see "**WP6**: Development and support for Semantic Multimedia Content and Query Models**"**). Unfortunately, MPEG-7 user profile preferences are very limited and thus cannot be used to express conditions in domain ontologies. We will develop a generic model for expressing the semantics of user queries and user preferences, which will be able to fully exploit the knowledge representation environment of MPEG-7 and domain ontologies that will be created in the work package "**WP6:** Development and support for Semantic Multimedia Content and Query Models" for creating metadata descriptions. The language constructs for the user queries and user profiles will be based on modern ontology query languages (like SPARQL [61] or others, but they will be able to express similarity queries as well). We will also create an interactive visual tool that will allow one to create such queries and preferences profiles. A user may have several such named profiles that refer to different personalized contextual ontologies. These profiles can interact in different ways with the systems of the multimedia information providers. They can for example be sent (periodically) to the providers, they can stay to the user stations monitoring broadcasts, they can interact with the intelligent multimedia objects, etc..

The user context may include many different types of information. We will provide a generic model and an expandable typology for user contexts. We will investigate a range of multimedia applications (such as education, entertainment, learning, professional) to decide the higher levels of the typology of the user contexts. A user context will be associated with a number of user preference profiles, ontologies and functions.

The user context includes device profiles (we will be using the MPEG-21 [15] and CC/PP [47] standard for describing them), personalized contextual ontologies and mappings to common ontologies and user preference profiles. A context model for multimedia consumption environments has been proposed in the 7th part of the MPEG-21 specification, in the UsageEnvironment DS. The model takes into account the user characteristics (including user identification

information, usage preferences, usage history, presentation preferences and accessibility characteristics), the device capabilities, the network characteristics (including network capabilities and network conditions) and the natural environment characteristics (including location, time audiovisual environment features like the noise level).

Other types of context that the context model may contain include user demographics, physical context (time, space), organizational context, goals, tasks, etc. The time and space contexts for example allow for reasoning with the time and space of the user. The model of user goals for example may describe a composite user goal (like learning a foreign language) in terms of sub-goals structured on a graph. The satisfaction of the sub-goals (to a certain degree) contributes to the final goal in some way. Some of the sub-goals are really more important than others and are weighted more towards the final objective. Proceeding through the sub-goals should take into account any natural order (like prerequisite courses) that may exist among sub-goals. The lower level sub-goals have to be satisfied with some reasonable quality before the system proposes higher-level goals. The system will be using the goal graph (which includes weights and satisfaction levels) to determine which kind of material is more appropriate currently to be delivered to the users. The goal graph is updated taking into account user interactions that lead to progress evaluation. At a certain stage, the proposed multimedia material by the system does not satisfy just one goal, but several, in some weighted fashion that depends on the current state of satisfaction of the sub-goals by the user. We will develop a model and algorithms to integrate the user goals into the user context. The user context may also contain information on the tasks that the user has currently in front of him. We will model the user tasks as an activity model similar for example to the one that UML supports. Each activity may be associated with a user personalized contextual ontology or a preference profile that is used to determine what is the related material to be retrieved for the task. The synthesized document will maintain the association of retrieved materials to tasks and will help the users in localizing the relevant material at presentation time (see more below on intelligent document scenarios). We anticipate that contexts that allow the description of tasks and the association of a task with ontologies and preferences related to the task, as well as the communication with automatically synthesized task-related intelligent objects also makes a good metaphor for user interface design in professional applications.

User profiles and contexts will be dynamically updated based on the usage history. The usage history of a user records all the interactions of the user with the various kinds of objects and records the context of those interactions including the metadata of the objects, the time needed for their presentation, the time viewed, and the type of interactions used during viewing, etc. We will develop models that take into account the existing usage statistics per domain of knowledge and their differentials, their existing contextualized ontologies and user preferences, as well as the submitted queries to dynamically update the user profiles. In our vision, it may be possible to mix I-Objects usage and semantic information in order to feed a Usage Mining Algorithms (UMA) module, which extracts useful knowledge models. Then, the patterns produced by the UMA module can in turn be used to refine models at various levels. They can be exploited to refine users' profiles in order to better describe an existing class of users or to create a new class. Usage patterns can also be exploited to refine presentation layouts and the ontology describing users' profiles. Since queries and preference profiles make explicit references to domain ontologies used for metadata production (indexing) we will trace such requests in order to help the information providers to determine the needs for new kinds of domain ontologies to be used for indexing. In order to extract information on user profiles from the way I-Objects are used, UMAs need to access different kinds of input data, namely: some raw representation of the interactions between users and I-Objects, i.e., the list of basic actions performed by the users while viewing the multimedia content; the semantic and rhetorical annotations of the I-Objects content provided to the users; and the profiles of the users who interacted with I-Objects. Such data should be organized in a warehousing system which provides full storage capabilities for both the raw usage data and higher-level information extracted from them. Such information should be obtained by means of a set of pre-processing functionalities provided by the warehousing system, e.g., methods for combining usage data with content semantic annotations in order to obtain semantically enriched views of user sessions. Moreover, the warehousing system should support efficient access methods for all kind of data stored, enabling the application of both simple analysis methods and complex data mining techniques. Semantic annotations of I-Objects and user profiles should be either accessible by the warehouse system through API provided by ad hoc repositories (those involved in the synthesis of I-Objects), or directly represented within the data warehouse.

Risk management

We are aware of the privacy issues regarding profiles of user preferences in an open environment and the tradeoffs involved. Although FACS does not provide new means of enhancing the protection of the privacy of the users we rely on the fact that this issue is adequately considered in other research projects and in commercial implementations. Contextual semantic multimedia retrieval and interaction with intelligent multimedia objects is a new field that creates a complex environment that may become difficult to understand and use. We will rely on extensive consistent modelling and interactive tools to achieve a trade-off between a consistent and easy to understand and use environment and powerful functionality that is usable in different application and knowledge domains.

B6.8.5. WP5: Intelligent multimedia objects modelling

The work package on modelling presentations has two broad components: the development of a persistent object model that contains multimedia segments and various classes of semantic and rhetoric annotations of those segments, and the definition of presentation templates that contain design contexts that can be used to generate presentation fragments based on user requirements and available media. These two aspects are combined within an intelligent multimedia object (or I-Object). The work package will study existing models for characterizing and storing media segments and it will develop new or extended approaches that are appropriate for the non-text information sequences studied by the FACS project.

From the point of view of presentation, intelligent multimedia objects in FACS can be seen as graphs, which specify a presentation sequence and allow alternatives for navigation control. State variables allow user interaction, changes in the flow of control and changes in the appearance of presentation. Content metadata at each retrieved object may be used to define the relationship of the content of the object with other retrieved objects and therefore potential sequences of presentation of the nodes of the graph and/or alternative navigation paths. Constraints on the size and time of presentation as well as on the content themes that must be present will also be used to decide which of the retrieved objects will be used to synthesize the final object. The final object may contain references to other relevant objects that were not included in its construction.

This work package will create the modelling architecture that will allow for the automatic synthesis of intelligent multimedia documents (in WP8) from retrieved multimedia objects (as defined in WP6). This environment will support authoring scenarios. A scenario describes recursively the structure of an intelligent multimedia document based on a scenario type. A scenario type defines the common characteristics of many different scenarios. We view a scenario as a model for representing activities. For example it can indicate parallel execution of sub-activities, joining activities, etc. At each state the execution of activities may be shown on a template of a specific type. For example parallel activities may be shown on a template of type map. User context profiles and queries may include information about the scenario types that the presentations of the retrieved documents will follow.

The first task consists of extending approaches for semantic/rhetorical models of information developed for the text domain. For purely data-centric applications (in which a well-defined object and a pre-defined presentation environment are known), existing notions of which data to display and how to display them are sufficient. For the construction of a more human-centred presentation, some notion of the "message" of the presentation must be defined, as well as a fundamental model of how a particular I-Object represents part of the message and its relationship to other I-Object making up the complete message. Early work [33] in the intelligent multimedia presentation field is based on a single controlling system that drives the construction of a presentation according to high-level goals. This takes no account of the availability of appropriate media items, relying rather on constructing a limited number of media types and modalities. More recent work [43], primarily text based, combines analyses of document structures with a constructive theory of rhetoric (rhetorical structure theory [69]) to drive an automated process of layout using the rhetorical roles of the text fragments or figures. To extend these to multimedia objects, we will need to take into account the less well-known effects of temporal layout on the extraction of meaning by viewers. We can build on the developments within the arts, including the fine arts [20], film [23], music and sound installations [54] which provide means to express and emphasize narratives through the manipulation of time and space, but much less is known about the simultaneous manipulation of time, layout and hyperlinks - all essential ingredients in an interactive multimedia presentation. We also need to investigate the different roles media items are able to play in a multimedia presentation. These roles can be based on rhetorical structure theory, but will need to be extended to take into account the relationships among aspects dependent on the media type and the potential relationships among different media types and modalities [53].

The second thrust of this work package is the design of a framework for document templates, which will provide for the combination of I-Objects into a fluid story sequence. While it is tempting to see an I-Object as an independent entity, any instance of the object has to determine whether it is re-usable in one or more roles in a document template. In turn, a document template needs to state which I-Object roles are admissible. At the various states of presentation the multimedia objects are presented on different state dependent templates. Thus the form of the presentation also depends on the intelligent document state. Templates are objects themselves that have a type (for example a map) and are composed of objects that have presentations, types, and metadata from domain ontologies (for example a specific name of a city on a map). The retrieved multimedia may be modelled with the objects of the templates based on their metadata and presented in the corresponding locations of the templates (or an interaction button indicating their existence is displayed). Special cases of templates are interaction templates and navigation control templates of various types that may be active at various states. The intelligent documents follow a component model and expose all their parameters, composed objects, metadata, state variables, control logic, presentation logic, to the outside world in a standardized manner so that higher level applications (like educational programs) can be easily built, possibly automatically, using them.

There are many presentation formats that can serve as the basis for a logical layout packaging model. We will investigate the integration of concepts from SMIL and MPEG-4 as a basis for defining the temporal-spatial aspects of a

presentation. These models will need to be extended to allow for the expression of the semantics of a coherent story framework and the integration of rhetorical characterizations of the intended presentation context. They will also need to be extended to operate in a dynamic, interactive context.

In this work package we will also model in detail the intelligent object interactions with the user profiles and contexts. The intelligent objects may be seen as structured objects that contain for each of their components content metadata, state behaviour and they are capable for interoperation with the environment, including with user contexts and profiles. The interoperability with the environment may be based on ontologies (both some objects states as well as content metadata may map to user context and preference ontologies). The intelligent objects and the user context environment will have logic to facilitate such import and export interactions. We will study a number of application environments (such as education, learning, entertainment) to understand better, model and support such interactions. For example in learning the state of knowledge of the learner as well as the satisfaction of her/his goals may be affected by such interactions.

The work package will structure its activities in four tasks, each of which is related to the two goals of the work package. The first task, relevant to both goals, is the development of a detailed research specification that defines the functional requirements and interdependencies. The second task is the development of rhetorical modelling and matching frameworks based on the logical structure of the captured information (audio and video/image) and presentation templates. The third task will develop a series of presentation templates that will be used to specify the nature of clips to be selected and to ensure the placement of clips within a coherent story framework. The final task is the detailed modelling of I-Object structures; these will combine the functionality defined in tasks 5.2 and 5.3.

Risk management

There are two risks associated with the modelling of I-Objects. The first is that a general model for non-domain-specific modelling constructs may not be able to be defined. This will result in the restriction that only certain domains will be able to be modelled during the course of the project. The second risk is that a general-purpose model for presentation layout and 'professional' look-and-feel will not be able to be defined that translates to all of the implementation domains supported in WP8. The result of this risk is that some initial artificial bounds will need to be placed on high-level presentation templates to accommodate the need for a complete series of proof-of-concept demonstrators. We expect both classes of risks to be manageable within the scope of the project.

B6.8.6. WP6: Development and support for Semantic Multimedia Content and Query Models

Modelling the way descriptions of document contents are represented is a very important issue. Since the users are not known beforehand and there is no commonly agreed methodology for the description (modelling) of the real world, it is typical that if different people were given the task of indexing they would end up modelling the real world concepts that appear in the multimedia content in different ways and also using different terminology for describing the same concepts. We call this problem the **multimedia modelling gap**. Common models for the multimedia content would also help to relieve the semantic gap problem and focus the future research in this important area, since information extraction researchers would have a concrete target of what information (concepts, attributes, relationships) they try to extract and how they are represented, and the same concrete targets would be shared by the researchers working for the semantic content representation retrieval and inference, as well as by the users to express their queries. The problem of the lack of a common model for describing concepts, relationships among concepts, and instances in an open environment is similar to the problem of the existence of different and unknown schemata by the peers in an open P2P network of independent peers [36]. The unknown schemata of the peers prohibit querying and transforming of the answers of the servers of the peers. The most promising solutions proposed recently in the field are based on the use of common ontologies known by every peer and local ontologies representing the schema of the peer (contextual ontologies) as well as mapping between them [19][45][46][22].

The closest to a commonly agreed model for the representation of multimedia information are standards for content descriptions like MPEG-7 and domain ontologies for various real life domains including domains that describe such things as “feelings”, “artforms” and “privacy”. We focus in this project in the use of multimedia content standards (higher ontologies) and domain ontologies as a common denominator in all the stages of the personalized multimedia document production: in the extraction process, the metadata derivation process, the search process, the personalization process, and the document synthesis process.

We will use MPEG-7/21 to represent all the video content metadata. Since MPEG-7/21 is represented with XML, this representation will be directly managed by the XML database and used by all multimedia applications that understand MPEG-7. We have recently shown that MPEG-7 can also be used as a knowledge representation language capable of describing semantics in some real world applications (like a goal that takes place at particular minute of the match in which different actors participate with roles like goalkeeper, player who scores in this particular event etc.)

[25][26][27][28][29]. We will generalize the methodology and we will create a software environment that will allow the description of domain ontologies using MPEG-7 and the use of MPEG-7 higher level ontologies and domain ontologies for the derivation of the metadata that describe the multimedia content. Thus, the metadata descriptions will be completely within MPEG-7 (and therefore used transparently by any application that understands only MPEG-7) and at the same time these descriptions will be "standardized" descriptions based on MPEG-7 concepts specified by the standard and also on complex ontologies of the real world. Note that for the description with metadata of a particular video segment, more than one ontology that cover different domains may have to be used. The different ontologies may reflect the interests of different user communities, or simply describe aspects like affective aspects, privacy, etc.

Since MPEG-7 is not really known and used by people that are familiar with the construction and use of ontologies and rather difficult to use it for this purpose, we will create an upper layer of interaction with users, which will be based on OWL, the language that is considered today as a standard for the representation of ontologies. Retrieval languages and inference mechanisms that are being developed today by the OWL research community will also facilitate and influence our work on the user querying and profiling. We will generate a model and software for using inference mechanisms to infer higher-level concepts like events that occur starting from the primitive semantic concepts that are detected from the semantic extraction process. These higher-level concepts will be stored as metadata descriptions. In the OWL layer both an upper ontology based on MPEG-7 and domain ontologies compatible with MPEG-7 will be able to be described. These descriptions will be used to generate the MPEG-7 metadata that are stored in the servers of the information providers to be used by MPEG-7 applications including search engines (they are represented using XML).

The users may express directly queries that reflect their interest at a particular point in time for specific multimedia content. These queries will be submitted to the information providers to gather relevant data. The system however may expand the user queries with preference profile information from different user profiles in order to make them more effective and to personalize them to the registered user interests. We will analyze the expansion of queries with preference profiles in order to develop a mathematical framework and tools that will analyze the expanded query for semantic and syntactic correctness.

Risk management

One potential risk is the existence of readily available ontologies in all the areas that we would like to investigate/experiment. The FACS partners have already available a rather involved ontology (on soccer) which could be used for experimentation and demonstration across all the models of the architecture. Other ontologies may be obtained for testing individual modules depending on the application (e.g. educational). When such ontologies are not readily available they will be created in an ad-hoc manner fro demonstration/experimentation purposes. Based on our preliminary investigations MPEG-7 will be able to represent complex ontologies in different domains transparently to the applications. If we find any severe limitations we will make concrete proposals for MPEG-7 extensions to meet the needs of the applications.

B6.8.7. WP7: Management of content, context, and profile descriptions

This work-package aims at defining methods for the efficient management of contents, contexts, and profiles. This is necessary in order to identify media clips that are relevant to the interest of the users (searching paradigm), or vice versa to identify profiles of users that can be interested to a new media clip (filtering paradigm).

In the project we will represent all this descriptive information using XML. In fact, emerging standards for the descriptions of media content, knowledge representation, profiles, and contexts, make use of XML encoded data. As an example, consider MPEG-7 [4], MPEG-21 [4], RDF [8], OWL [6], CC/PP [47], P3P [7], which are all defined to be represented with XML.

Recently, several XML database management systems were proposed to deal natively with large amount of XML data. As an example consider Tamino [9], Xindice[11], Lore [3], Timber [10]. Accordingly, XML query languages, such as XQuery [13] and XPath [12] were also proposed to query such databases. However, current XML database technology is mainly focused on the issue of processing queries efficiently according to the *exact match paradigm* (in addition to some text retrieval functionalities). In this paradigm the result of a query is the set of data that exactly match what has been specified in the query definition. However, the capability needed to achieve the objectives of this project is beyond the simple exact match paradigm and the *approximate match paradigm* should be provided as well. In fact, the description of the content of a video, for instance, might contain ambiguity, imprecision, uncertainty, and subjectivity. In order to deal with the issue of approximate match search, XML query languages should be extended so that approximate match operators can be combined with exact match operators. Processing a query or matching a user profile with descriptions of content is a fuzzy process. The result of a query is a ranked set of documents that are more or less deemed to be relevant to the query. In this context, an active research field is the study of efficient algorithms and data structures (indexes) for efficient processing of combined approximate and exact match queries.

Another issue that will be considered in the project is the *highly distributed environment* where contexts, profiles, and content descriptions should be retrieved and filtered. We imagine that, according to what is happening nowadays with

RSS and Podcast feeds, a plethora of multimedia content feeds will be in the future available. In addition, the automated content description tools, developed in this project, will associate the shared multimedia content with the corresponding content descriptions. All this information is clearly highly distributed throughout the various sites that offer multimedia content feeds. The problem becomes even more difficult if we consider that profiles of users are also highly distributed. Users might store their profiles on repositories offered by their service providers, they can store their profiles directly on their PCs, or even on their mobile devices. In this context, the approximate matching problem, consisting of both finding relevant clips given a profile and finding profiles relevant to a media clip, is very difficult as discussed in the following.

A simple solution would be to have a centralised repository where all this information is stored. However, the high dynamicity and volatility of both available clips and user profiles makes this solution not efficient and not scalable. A more promising direction is offered by recent developments in the GRID field. Our distributed data environment can be managed by leveraging and enhancing current *GRID technology for data management and distributed query processing*. The GRID framework, in this context, will be also particularly useful to employ policies of privacy and trust and to manage the access rights to the various shared data (media clips, contexts, profiles). Various initiatives, as for instance gLite data management [1], OGSA-DAI [5], OGSA-DQP [50], GRACE[2], have considered the issue of data management in a GRID environment. However, some of them just consider a file-systems level data management (gLite), some deal with integrated access to different (relational and XML) databases (OGSA-DAI), some offer distributed query processing on multiple wrapped data sources (OGSA-DQP), others (GRACE) offer text retrieval and categorisation on GRID enhanced search engines.

In our case the problem is more difficult. Given the potentially large number of data sources, we have to consider our data environment as a single highly distributed database where complex combined exact and approximate match queries should be processed. Current distributed query processing techniques know where data of a certain type is stored and they have just to relate data in a site with data in other sites. For instance, the “customer” database entirely maintained at site A should be related with the “orders” database maintained at site B. However, in our case data of the same type (for instance content descriptions) are stored across a potentially very large number of sites that should all be searched or filtered. We classify existing data management approaches on the GRID as *vertical data management*, in contrast to the *horizontal data management* needed in our case. This poses new interesting problems to be solved in the area of data management in the GRID environment, especially if we consider that we need approximate match search and filtering capabilities to be available in addition to exact match functionalities.

To this aim, in the FACS project we will define new efficient XML horizontal distributed data storage paradigms, including new distributed access methods for efficient exact and approximate match search. Access methods and search algorithms that we will develop allow retrieval of relevant data, distributed in various sites, by accessing just a few promising sites depending on the specific approximate XML queries (generated from a user profile, when searching, or from a content description, when filtering) being processed.

We will also develop new query optimisation and query processing techniques by taking full advantage of the data management capabilities of the GRID environment. Different query execution strategies may imply different costs, in terms of query processing time and/or system load. We will exploit the GRID infrastructure to also make it possible to use query execution plans where sub-queries are possibly distributed to various nodes, to avoid bottleneck and provide load balance.

Risk management

Risks in the management of content, profile, and context descriptions rely in the failure of achieving the needed efficiency, in terms of response time, and scalability required by the functionality of the presentation generation process. Will the component, realized in this work package, offer good performance when the number of descriptions to be managed and the number of users to be served simultaneously increase? Will new content fed to the content environment be immediately indexed by the distributed indexes so that they can be immediately used?

We will use techniques appositively designed to deal will heavily distributed and dynamic information to cope with these issues, and we will carry out simulations to assess the efficiency and scalability properties of the developed techniques.

B6.8.8. WP8: Presentations delivery

While previous work packages mainly dealt with high level, semantic construction and manipulation of I-Objects, this work package deals with syntactic aspects of binding I-Object structure and content, and with the physical transformation of bound objects to various presentation platforms. It also studies methods of integrating interaction by the user in the client into a successive-refinement approach of successor object definition and presentation.

We plan to build upon an intermediate object description, which should adhere to the respective MPEG-21 standards as closely as possible. This object description will contain aspects of Digital Item Declaration (DID), Digital Item

Adaptation (DIA), and Digital Item Processing (DIP). It will make use of, and where necessary, extend various aspects of relevant W3C standards such as SMIL, RDF and related specification.

One task is to define and develop a series of binding strategies to associate abstract I-Objects with presentation environments. This abstract description, probably enriched with information currently not expressible in MPEG-21 (e.g. the selected presentation template), has to be matched against the capabilities of various end user devices, ranging from low end (e.g. handhelds, mobile phones) through mid-range (e.g. interactive TV) to high end (PC style).

Once a binding strategy is defined, we will design and implement a demonstrator in which a framework is illustrated that allows the easy implementation of multimedia distribution for existing and new platforms. To demonstrate the feasibility of the approach, some of the possible distribution paths will be implemented as prototypes. We envisage targeting at least PC-style internet devices (due to the broad range of available tools and resources), the Multimedia Home Platform (due to its attractiveness for a future mass market), and UTMS-level mobile devices (based on 3GPP/SMIL, DVB-H and IP Streaming technology) will be taken into account as well.

A key element of this work package is the development on an interaction architecture in which user interaction (in terms of the specification on incremental preferences, temporal hyperlink behaviour or direct I-Object interaction) is used to adapt the behaviour of the current I-Object, or the construction of successor I-Objects. An interaction model will be defined and a prototype interaction architecture will be integrated into the FACS demonstrators at the server and client levels.

Risk management

The objective of WP8 is to define a broad publishing architecture in which presentations based on I-Objects can be generated for a wide range of existing standards. The primary partners in the project have access to implementations of native players for these formats, so that no primary risk is seen in underlying platforms. There is a minor risk that all of the functionality required for client-side generation and interaction of content may not be available in low-power devices; in this case, we expect the scalable architecture to determine a methodology to coupe with heterogeneous devices.

B6.8.9. WP9: Dissemination and Exploitation

A structured dissemination and exploitation plan will be defined at the beginning of the project, in order to support an effective exploitation of the project results towards four main communities: researchers, contents providers, services providers, end-users. The exploitation plan will explain in detail how the output of the project would be introduced into common and individual strategies. Particularly, the consortium will have to identify the services and/or the products that could be derived from the project results, based on a scientific/technological survey and on a market survey.

Dissemination activities will be conducted through the following strategy:

- The results of the project, and mainly the reference architecture will result in articles and documentation to be submitted to conferences, workshops and selected high-level magazines (e.g. ACM Multimedia, VLDB Journal, ACM-SIGMOD, ACM-SIGIR, ACM TOIS, IEEE TPAMI, ACM Multimedia Systems Journal, IEEE Transactions on Neural Networks, IEEE Transactions on Signal Processing, etc.).
- The project will regularly organize (annually) a scientific workshop for the international research community, and a series of industrial days (according to the main releases of project results), inviting content and service providers, TV broadcasters, and potential investors. Both types of workshops will be organized with the support of the DELOS and MUSCLE NoEs. A letter of support of the Scientific Directors of the DELOS NoE and of the MUSCLE NoE is included in Appendix E.
- Since the FACS project is aimed at developing and experimenting the technologies needed to produce personalised and interactive digital content for entertainment, personal information, and edutainment, a profitable interaction with standardization bodies and user groups – e.g. mainly IETF, Mobile Entertainment Forum, FIAT/IFTA (International Federation of Television Archives), EBU (European Broadcast Union), W3C, MPEG, etc. – is expected.
- A project web site will be created at the beginning of the project and maintained during the entire project lifetime. Through this web site, the public documents produced within the FACS project will be made accessible, along with links to specific sites and documentation about the FACS technological issues.
- A two-monthly FACS Newsletter will be produced and delivered by e-mail and RSS feed to all communities interested to project results. All partners will contribute to the preparation of the newsletter.

The sharing of the expertise held by each partner will be a plus and will result rapidly in collaborative research and publications; in a longer timeframe, thanks to the employment of young researchers and Ph.D. students, it will result in opportunities for the new generation of researchers, both from Western and Eastern Europe.

Project exploitation will be done within the consortium and toward the EU research community.

The first one will be done towards the universities and research institutions that constitute the FACS consortium and that will strongly benefit from the scientific advances obtained from the cooperation with all other partners in a field that requires the expertise of many research groups. The FACS project will also address its results exploitation to the entire EU research community, with particular attention to technology and service providers. Technology providers will benefit from the research and feasibility assessment in bridging the gap between the current solution architecture and the flexible approach proposed in the FACS project. The FACS unified solution will provide useful indications for the engineering of new functionality into the real distribution environment, both in terms of network capabilities and of service provisioning schema. Service providers will interact with new and different mechanisms coming from the personalisation of services preserving their internal technological solution and with shorter time to market for service supplying. This will be made feasible by the FACS architecture and the specific components developed in the project.

The **project exploitation** is based on the following strategy:

- Dissemination activities in order to promote the adoption of the project "vision" and to demonstrate the feasibility of the project approach
- All software components, algorithm and techniques, and the resources (ontologies, I-Object templates, etc.) developed in the project will be made available as public domain. This approach is intended to support the diffusion and the adoption of project approach and results by the widest possible number of users.

B6.8.10. WP10: Project Management

The goal of project management is to ensure that the project is carried out respecting the objectives, the planned duration and the estimated costs. This will require intense communication between the partners, strict adherence to schedules and closely monitored progress.

A key activity to be carried out in this work package will be the continuous monitoring of project activities, control of the adherence to the work schedule and activation of immediate actions in case of significant deviations from the project's work plan.

This work package includes preparation of and participation in the regular project meetings, management of the communication among the partners, the coordination of reporting and administrative activities. The management structure and procedures are detailed in Chapter B5.

Project Management and Coordination will result in Activity and Financial Reports as required by the Commission. The coordinating partner is in charge of this work package. However, all the staff, management organisation, and the Project Management Board will directly support the Project Director, according to the scheme presented in Chapter B5. Administrative and financial aspects will be supported and coordinated within the work package.

B6.9. Graphical presentation of components (Pert diagram)

B6.10. Work Planning and Timetable (Gantt Chart)

Work package list (full duration of project)

Work-package No[4]	Work package title	Lead contractor No[5]	Start month[6]	End month[7]	Deliverable No[8]
WP 1	WP1: Research activity management and monitoring	1	1	36	D1.1.n, D1.2.n, D1.3.1, D1.4.n, D1.3.3
WP 2	WP2: Feature extraction	7	3	30	D2.1.1, D2.2.n, D2.3.1, D2.4.1, D2.3.2, D2.4.2
WP 3	WP3: Semantic and rhetoric annotation	6	3	36	D3.1.1, D3.2.1, D3.3.1, D3.2.2, D3.3.2
WP 4	WP4: Modelling and Supporting User Profiles and Contexts	5	3	33	D4.1.1, D4.2.1, D4.3.1, D4.4.1
WP 5	WP5: Intelligent multimedia objects modelling	3	3	30	D5.1.1, D5.2.1, D5.3.1, D5.4.1, D5.3.2, D5.4.2
WP 6	WP6: Development and support for Semantic Multimedia Content and Query Models	5	3	33	D6.1, D6.2, D6.3, D6.4
WP 7	WP7: Management of content, context, and profile descriptions	1	3	36	D7.1.1, D7.2.1, D7.2.2, D7.3.1
WP 8	WP8: Presentations delivery	4	10	36	D8.1.1, D8.2.1, D8.3.1, D8.4.1
WP 9	WP9: Dissemination and Exploitation	1	1	36	D9.1.1, D9.2.n, D9.3.n, D9.1.2, D9.4.n, D9.1.3
WP 10	WP10: Project Management	1	1	36	D10.1, D10.2.n, D10.3.n
	TOTAL				

[4] Work package number: WP 1 – WP n.

[5] Number of the contractor leading the work in this work package.

[6] Relative start date for the work in the specific work packages, month 0 marking the start of the project, and all other start dates being relative to this start date.

[7] Relative end date, month 0 marking the start of the project, and all ends dates being relative to this start date.

[8] Deliverable number: Number for the deliverable(s)/result(s) mentioned in the work package: D1 - Dn.

Deliverables list (full duration of project)

This is a complete list of FACS official deliverables. Different work packages will also provide additional *internal deliverables* (listed in the WP forms) in order to properly exchange information and intermediate results with other work packages.

Deliverable No	Deliverable title	Delivery date	Nature	Dissemination level
D10.1	Consortium Agreement	1	R	CO
D1.1.n	Report on the state of the art	3, 18, 36	R	PU
D2.1.1	WP2 Detailed research specification	6	R	CO
D2.2.n	Content selected for the experiments	6, 21	O	CO
D3.1.1	WP3 Detailed research specification	6	R	CO
D4.1.1	WP4 Detailed research specification for modelling user profile and context	6	R	CO
D5.1.1	WP5 detailed Research Specification	6	R	CO
D6.1	Detailed research specification for modelling metadata and query analysis	6	R	CO
D7.1.1	WP7 Detailed research specification	6	R	CO
D9.1.1	Initial dissemination and exploitation planning report	3	R	PU
D9.2.n	Production of promotion material	3, 18, 36	O	PU
D1.2.n	Global architecture specification report	4, 21, 36	R	PU
D9.3.n	Organization of workshop	6, 18, 36	O	PU
D10.2.n	Financial reports	6,12,18,30,36	R	CO
D10.3.n	Progress monitoring reports	6,12,18,30,36	R	CO
D8.1.1	WP8 Detailed research specification	14	R	CO
D1.3.1	First project result evaluation	18	R	PU
D1.4.n	Report on risk analysis and contingency plans	18, 27, 33	R	CO
D2.3.1	First Prototype video and image feature extraction implementation	18	P	PU
D2.4.1	First Prototype audio and music feature extraction implementation	18	P	PU
D3.2.1	First prototype multimedia semantic annotation implementation	18	P	PU
D3.3.1	First Prototype multimedia rhetorical annotation implementation	18	P	PU
D7.2.1	First prototype distributed management of description implementation	18	P	PU
D9.1.2	Revision of dissemination and exploitation planning	18	R	CO
D9.4.n	Report on scientific exploitation	18, 27, 36	R	CO

Deliverable No[9]	Deliverable title	Delivery date [10]	Nature [11]	Dissemination level[12]
D4.2.1	User profile models and tools	24	R	PU
D4.3.1	Context modelling report	24	R	PU
D5.2.1	Rhetorical Modelling/Matching Specification	24	R	CO
D5.3.1	Presentation Template Specification	24	R	CO
D5.4.1	I-Object Model Specification	24	R	CO
D6.2	Domain ontologies and metadata for content description report	24	R	PU
D6.3	Ontology integration mechanisms	24	R	PU
D7.2.2	Prototype distributed management of description implementation	24	P	PU
D2.3.2	Prototype video and image feature extraction implementation	27	P	PU
D2.4.2	Prototype audio and music feature extraction implementation	27	P	PU
D5.3.2	Collection of Presentation Templates	27	O	PU
D8.2.1	Prototype Media binding implementation	27	P	PU
D8.3.1	Prototype Server-side implementation	29	P	PU
D5.4.2	I-Object Creation/Access Environment	30	P	PU
D3.2.2	Prototype multimedia semantic annotation implementation	33	P	PU
D3.3.2	Prototype multimedia rhetorical annotation implementation	33	P	PU
D4.4.1	Usage history modelling and tools for the interactive generation of profiles and contexts	33	R	PU
D6.4	Retrieval and inference language and query processing	33	R	PU
D7.3.1	Prototype distributed approximate match implementation	33	P	PU
D8.4.1	Prototype Client Side implementation	33	P	PU

[9] Deliverable numbers in order of delivery dates: D1 – Dn

[10] Month in which the deliverables will be available. Month 0 marking the start of the project, and all delivery dates being relative to this start date.

[11] Please indicate the nature of the deliverable using one of the following codes:

R = Report
P = Prototype
D = Demonstrator
O = Other

[12] Please indicate the dissemination level using one of the following codes:

PU = Public

Work packages description (full duration of project)

Work package number	WP1	**Start date or starting event:**	1
Work package title WP1: Research activity management and monitoring			

Participant id	**ISTI-CNR**	CMU	CWI	FhG	TUC	TUWIEN	UVA
Person-months per participant:	10.0	1.0	1.5	2.0	4.0	6.0	6.0

Objectives: Controlling the correct execution of the project research technical activity, setting-up the initial common framework for coherent cross-disciplinary research activity, taking decision to solve encountered problems.

Description of work

T1.1 – State of the art monitoring.

Periodical monitoring of the state of the art related to the various project research activities.

Participants: all partners

T1.2 – Global Architecture and component functionality specification

Design of a common general architecture where all component level activities can be easily integrated. It provides a common framework to carry out cross-disciplinary research. The global architecture will be defined at the beginning of the project, revised after the first prototypes are evaluated, and during the final evaluation of the project results.

Participants: all partners

T1.3 – Project result assessment

Assessment of the intermediate and final research results obtained in correspondence of the main milestones of the project and production of reference documents on project results. Assessment is performed after acquiring information, by means of internal deliverables, on the specific result evaluation carried out in various work packages.

Participants: all partners

T1.4 – Risk and result analysis, and contingency plans

Results of the various activities are continuously monitored so that risks are immediately detected and corrective actions are taken

Participants: all partners

Deliverables

D1.1.n Report on the state of the art 3, 18, 36 (R)

D1.2.n Global architecture specification report 4, 21, 36 (R)

D1.3.1 First project result evaluation 18 (R)

D1.4.n Report on risk analysis and contingency plans 18, 27, 33 (R)

D1.3.3 Final Project result evaluation 36 (R)

PP = Restricted to other programme participants (including the Commission Services).

RE = Restricted to a group specified by the consortium (including the Commission Services).

CO = Confidential, only for members of the consortium (including the Commission Services).

Milestones[13] and expected result
M1: First assessment of the state-of-the art, Global functional architecture specification
M2: Evaluation of current results
M4: Final evaluation of project results

[13] Milestones are control points at which decisions are needed; for example concerning which of several technologies will be adopted as the basis for the next phase of the project.

Work packages description (full duration of project)

Work package number	WP2	**Start date or starting event:**				3	
Work package title WP2: Feature extraction							
Participant id	ISTI-CNR	CMU	CWI	FhG	TUC	TUWIEN	**UVA**
Person-months per participant:		8.0	1.0	1.0		35.0	42.0

Objectives Investigating and developing new feature extraction techniques for feature based searching and filtering of image, video, audio, and music documents; they will also be used by techniques developed in WP3 to generate semantic and rhetorical annotations of media clips.

Description of work

T2.1 Detailed research specification

Collection of functional requirements and functional specification for the research activity of this work package.

Participants: UVA, TUWIEN

T2.2 Multimedia content selection

Collection of image, video, audio, and music material for experimenting and validating the developed techniques. Collection will be performed at the beginning of the projects and will be updated after the first prototypes are delivered.

Participants: UVA, TUWIEN, CMU (They are in charge of selecting the needed material)

CWI, FhG (They should check that the content of the selected material is compatible with the types of presentation that will be automatically generated)

T2.3 Video and image feature extraction

Investigating and developing new feature extraction techniques for image and video document

Participants: UVA, CMU

T2.4 Audio and music feature extraction

Investigating and developing new feature extraction techniques for audio, and music documents

Participants: TUWIEN, CMU

T2.5 Results evaluation

Evaluation of the developed techniques

Participants: UVA, TUWIEN

Deliverables

D2.1.1 WP2 Detailed research specification 6 (R)

D2.2.n Content selected for the experiments 6, 21 (O)

D2.3.1 First Prototype video and image feature extraction implementation 18 (P)

D2.4.1 First Prototype audio and music feature extraction implementation 18 (P)

D2.3.2 Prototype video and image feature extraction implementation 27 (P)

D2.4.2 Prototype audio and music feature extraction implementation 27 (P)

Internal Deliverables

ID2.5.1 Feature extraction first evaluation 18 (R)

ID2.1.1 WP2 Detailed research specification revision 21 (R)

ID2.5.2 Final results evaluation 27 (R)

Milestones and expected result
M1: Detailed research specification, Selection of multimedia content for tests
M2: Preliminary feature extraction techniques
M4: Feature extraction techniques

Work packages description (full duration of project)

Work package number	WP3	**Start date or starting event:**				3	
Work package title WP3: Semantic and rhetoric annotation							
Participant id	ISTI-CNR	CMU	CWI	FhG	TUC	**TUWIEN**	UVA
Person-months per participant:		10.0				39.0	64.0

Objectives Developing techniques for the automated generation of semantic and rhetoric annotation of media clips.

Description of work

T3.1 Detailed research specification

Collection of functional requirements and functional specification for the research activity of this work package.

Participants: UVA, TUWIEN

T3.2 Multimedia Semantics annotation

Automatic generation of semantic annotations related to the content of images, video, audio, and music clips, by using machine-learning techniques.

Participants UVA, TUWIEN, CMU

T3.3 Multimedia Rhetorical annotation

Automatic generation of rhetorical annotations related to the content of images, video, audio, and music clips, by using machine learning techniques.

Participants UVA, TUWIEN, CMU

T3.4 Results evaluation

Evaluation of the developed techniques

Participants: UVA, TUWIEN

Deliverables

D3.1.1 WP3 Detailed research specification 6 (R)

D3.2.1 First prototype multimedia semantic annotation implementation 18 (P)

D3.3.1 First Prototype multimedia rhetorical annotation implementation 18 (P)

D3.2.2 Prototype multimedia semantic annotation implementation 33 (P)

D3.3.2 Prototype multimedia rhetorical annotation implementation 33 (P)

Internal deliverables

ID3.4.1 Multimedia annotation first evaluation 18 (R)

ID3.1.1 WP3 Detailed research specification revision 21 (R)

ID3.4.2 Final results evaluation 36 (R)

Milestones and expected result

M1: Detailed research specification

M2: Preliminary semantic and rhetoric annotation techniques

M4: Semantic and rhetoric annotation techniques

Work packages description (full duration of project)

Work package number	WP4	**Start date or starting event:**				3	
Work package title WP4: Modelling and Supporting User Profiles and Contexts							
Participant id	ISTI-CNR	CMU	CWI	FhG	**TUC**	TUWIEN	UVA
Person-months per participant:	35.0				40.0	6.0	9.0

Objectives

- The first objective of this work package is to develop a generalized and powerful model and software for describing semantic user preferences in an environment that supports MPEG-7 content descriptions augmented with domain ontology concepts. Current MPEG-7 profiles are very limited and inadequate for this purpose. They will be a special case of this model.
- The second objective is to provide a generalized and expandable model and software for describing user contexts also supporting personalized and contextualized ontologies and associated to user preferences.
- The third objective of the work package is to develop methodologies of dynamically adapting user preferences and their relationship to contexts through usage histories.

Description of work

T4.1 Detailed research specification

This task will collect the functional requirements and model the functional specification for the research activity of this work package focusing on the state of the art for personalization and contextualization that the project will enhance.

Participants: TUC, CNR

T4.2 Modelling and generating user preferences

This task will model multimedia user preferences in a generalized manner that can refer to the semantic MPEG-7 descriptions enhanced with ontology concepts. The existing support in the current MPEG-7 profiles will be only a special case of this generalized semantic approach. The task will also develop a visual tool for the interactive generation of such user preferences. These will be converted to a standard syntax for querying.

Participants: TUC, CNR (identify and define domain ontologies according to emerging standards)

UVA, TUWIEN (consultation on the type of semantic content that can be automatically extracted from documents, and can be used in the user profile)

UVA, TUWIEN (consultation on the type of features that can be automatically extracted from documents, to be also embedded in ontologies, for low-level description of user preferences)

T4.3 Modelling and generating user contexts

This task will provide a generalized model and structure for the user contexts. They will include general types like the user demographics, user preferences, personalized ontologies, organizational context as well as goal and task context. In each category existing ontologies and international standards will be taken into account and integrated dynamically. The task will also develop tools for generating personalized contextual ontologies and mapping them to common ontologies, as well as tools for interactive generation/update of contexts and context selection and association of contexts with preferences.

Participants: TUC, CNR

T4.4 Dynamic adaptation of user preferences

The task will develop new techniques for the automatic acquisition and/or dynamic update of user profiles and context selection based on the tracing and the analysis of the user's interaction with the end-user devices and on the information acquired from them. A generalized model of usage history will also be developed based on which preferences on semantic content will be inferred. The semantics of the content will be described again with the association of MPEG-7 with the domain ontologies, which result, to semantic content descriptors

Participants: CNR, TUC

Deliverables
D4.1.1 WP4 Detailed research specification for modelling user profile and context 6 (R)
D4.2.1 User profile models and tools 24 (R)
D4.3.1 Context modelling report 24 (R)
D4.4.1 Usage history modelling and tools for the interactive generation of profiles and contexts 33 (R)

Milestones and expected result
M1: Detailed research specification
M3: Models for descriptions of user contexts and profiles
M4: Techniques for interactive generation of contexts and profiles

Work packages description (full duration of project)

Work package number	WP5	**Start date or starting event:**	3
Work package title WP5: Intelligent multimedia objects modelling			

Participant id	ISTI-CNR	CMU	**CWI**	FhG	TUC	TUWIEN	UVA
Person-months per participant:			60.0	14.0	22.0	3.0	5.0

Objectives: Investigating and developing models for representing presentation templates and instances of personalised presentations based on intelligent multimedia objects (I-Objects). Modelling the intelligent object interactions with the user profiles and contexts. Production of a set of presentation templates for the experiments. Developing methodologies for choosing presentation templates according to the user preferences.

Description of work

T5.1 Detailed research specification

Collection of functional requirements and functional specification for the basis of the research activities of tasks T5.2, T5.3 and T5.4 of this work package.

Participants: CWI, UVA, TUWIEN, TUC, FhG

T5.2 Rhetorical modelling and matching

Developing models and structures for rhetorical information annotation within components of template presentations and media clips. Definition of rules for matching rhetorical annotation of presentation templates and media clips. A set of annotated clips will be defined for use in experimentation.

Participants: UVA, TUWIEN, CWI (development of rhetorical models and matching strategies)

CWI, FhG (Integration of rhetorical information in presentation templates)

T5.3 Presentation template modelling

Developing models and structures for presentation templates that are associated with descriptions of content class and can be match with specific user profiles. A set of presentation templates will be defined use in experimentation.

Participants: CWI, FhG

T5.4 I-Object structure, active behaviour, and interaction modelling

Modelling instances of personalised presentations (I-Objects) containing rhetoric that serve as instances of a presentation template by binding elements of the template with existing media clips. I-Objects have an active behaviour that allows them to adapt themselves to the user according to his/her interaction and the capability of the devices. The intelligent objects will be structured objects that contain for each of their components content metadata, state behaviour and they will be capable for interoperation with the environment, including with user contexts and profiles. We will study a number of application environments (such as education, learning, entertainment) to understand better, model and support such interactions.

Participants: CWI, FhG (development of the base model of the structure of I-Objects)

TUC (development of a model for the active behaviour and interaction with user contexts and profiles of I-Objects)

Deliverables

D5.1.1 WP5 detailed Research Specification 6 (R)

D5.2.1 Rhetorical Modelling/Matching Specification 24 (R)

D5.3.1 Presentation Template Specification 24 (R)

D5.4.1 I-Object Model Specification 24 (R)

D5.3.2 Collection of Presentation Templates 27 (O)

D5.4.2 I-Object Creation/Access Environment 30 (P)

Milestones and expected result
M1: Detailed research specification
M3: Models for presentation templates, Model for rhetorical descriptions, Model for I-Object representations, Collection of presentation templates
M4: Environment for I-Object creation and access

Work packages description (full duration of project)

Work package number	WP6	**Start date or starting event:**	3
Work package title WP6: Development and support for Semantic Multimedia Content and Query Models			

Participant id	ISTI-CNR	CMU	CWI	FhG	**TUC**	TUWIEN	UVA
Person-months per participant:	30.0		1.0	1.0	40.0	1.0	4,0

Objectives

- The first objective of the work package is to provide a generalized methodology for describing the content of multimedia data using MPEG-7/21 compatible descriptions of the multimedia material and the semantics of the real world that they represent. Multiple ontologies may be used for such a characterization.
- The second objective of the work package is to provide a methodology and software that facilitates as much as possible the integration of domain ontologies defined by user experts in ontologies (like OWL experts) in MPEG-7.
- The third objective of the work package is to define a language for retrieval and inference that will be used to express semantic user queries and user preferences. Since user queries may be augmented by user preferences the objective will be to develop methodologies and software to resolve syntactic and semantic conflicts (with the help of ontologies) before submitting queries for execution..

Description of work

T6.1 Detailed research specification

Collection of functional requirements and functional specification for the research activity of this work package.

Participants: TUC, CNR

T6.2 Domain ontologies and metadata for content description

Development of a methodology for producing semantic metadata descriptions that transparently supports and facilitates the integration of domain ontologies into MPEG-7. Study of different content and application domain classes and the primitives is needed for fast integration of domain ontologies. In addition description of important intermediate ontologies (connectors) is needed to allow faster integration.

Participants: TUC, CNR (identify and define domain ontologies according to emerging standards)

UVA, TUWIEN (consultation on the type of semantic content that can be automatically extracted from documents)

UVA, TUWIEN (consultation on the type of features that can be automatically extracted from documents, to be embedded in ontologies)

CWI, FhG, UVA, TUWIEN (consultation on the type of rhetorical information that can be automatically extracted from documents, and that is useful for presentation generation)

T6.3 Ontology integration middleware

Specification and development of middleware for supporting domain ontology integration in MPEG-7 applications, including ontology definition (like OWL) in an ontology language, and including the development of a visual environment that supports the interactive generation and/or integration of domain ontologies with MPEG-7 and their use for metadata generation.

Participants: TUC, CNR

T6.4 Retrieval and inference language and query processing

Specification and implementation of the query and inference language capabilities, the personalization and contextualization processes and the syntactic and semantic conflict analysis which takes place when the user queries are expanded for personalization and contextualization.

Participants: TUC, CNR

Deliverables

D6.1 Detailed research specification for modelling metadata and query analysis 6 (R)
D6.2 Domain ontologies and metadata for content description report 24 (R)
D6.3 Ontology integration mechanisms 24 (R)
D6.4 Retrieval and inference language and query processing 33 (R)

Milestones and expected result
M1: Detailed research specification
M3 Model for description of contents, Ontology integration middleware
M4 Retrieval and inference language and query processing

Work packages description (full duration of project)

Work package number	WP7	**Start date or starting event:**	3
Work package title WP7: Management of content, context, and profile descriptions			

Participant id	**ISTI-CNR**	CMU	CWI	FhG	TUC	TUWIEN	UVA
Person-months per participant:	76.0				15.0		

Objectives Investigating and developing techniques for efficient distributed management and approximate matching of descriptions of documents' content, user profiles, and contexts.

Description of work

T7.1 - Detailed research specification

Collection of functional requirements and functional specification of the component(s).

Participants: CNR, TUC

T7.2 - Distributed management of user profiles, contexts, and content descriptions

Investigating techniques (data structures and algorithms) for efficient distributed management of XML represented semi-structured data, on architectures based on GRID services.

Participants: CNR

T7.3 - Distributed approximate match search and filtering

Distributed access methods for efficient retrieval and filtering of data, based on approximate match on content and structure

Participants: CNR (development of access methods for approximate filtering)

TUC (development of specific approximate matching functions for retrieval and filtering of content descriptions)

T7.4 - Results evaluation

Evaluation of the developed techniques

Participants: CNR, TUC

Deliverables

D7.1.1 WP7 Detailed research specification 6 (R)

D7.2.1 First prototype distributed management of description implementation 18 (P)

D7.2.2 Prototype distributed management of description implementation 24 (P)

D7.3.1 Prototype distributed approximate match implementation 33 (P)

Internal deliverables

ID7.4.1 Distributed management of descriptions first evaluation 18 (R)

ID7.1.1 WP7 Detailed research specification first revision 21 (R)

ID7.4.2 Final results evaluation 36 (R)

Milestones and expected result

M1: Detailed research specification

M2: Preliminary distributed repository for content, contexts and profile descriptions

M4: Distributed repository for content, contexts and profile descriptions, Approximate matching and filtering techniques

Work packages description (full duration of project)

Work package number	WP8	**Start date or starting event:**				10	
Work package title WP8: Presentations delivery							
Participant id	ISTI-CNR	CMU	CWI	**FhG**	TUC	TUWIEN	UVA
Person-months per participant:	5.0		72.0	32.0			

Objectives In this work package we will design a framework for delivering I-Objects to the user. We will strive to use existing standards for multimedia delivery wherever possible and appropriate. Presentation of I-Objects should be possible on a wide range of devices: from low end (e.g. handhelds, mobile phones) through mid-range (e.g. interactive TV) to high end (PC style). To demonstrate the feasibility of our approach, some of the possible distribution paths will be implemented as prototypes.

Description of work

T8.1 - Detailed research specification

In this task we will evaluate the relevant standards for packaging and delivering multimedia presentations (e.g. SMIL, MPEG-4, MPEG-4/XMT, DVB/MHP) with respect to their suitability for presenting I-Objects. Where necessary, additional requirements and functional specifications will be produced. The outcome of this task will strongly determine the work in the remaining tasks of this WP, especially the choice of prototypes to implement.

Participants: CWI, FhG

T8.2 Binding selected media clips with selected presentation layouts

Previous work packages will produce a set of media clips and several descriptions, including user preferences, selected presentation layouts and device profiles. In this task we will combine these elements into one coherent, abstract I-Object. It is abstract in the sense that it is independent of packaging, transport and presentation device.

Participants: CWI, FhG (instantiation of personalised presentations)

CNR (access to rhetorical descriptions)

T8.3 Server-side transformation of presentation

Due to the large differences in available resources (bandwidth, memory, cpu, input devices etc.), different physical representations of the same I-Object have to be generated before delivery. In this task we will evaluate necessary server-side transformations of the abstract I-Object developed in T8.2. Prototypes for some devices will be developed to demonstrate the feasibility of our approach.

Participants: CWI, FhG

T8.4 Client-side rendering of presentations

Depending on the selected output device and the outcome of task 8.1, I-Object might be presented using standard viewers. Otherwise we will use or enhance tools accessible to project partners.

Participants: CWI, FhG

T8.5 - Results evaluation

Evaluation of the developed techniques

Participants: CWI, FhG

Deliverables

D8.1.1 WP8 Detailed research specification 14 (R)

D8.2.1 Prototype Media binding implementation 27 (P)

D8.3.1 Prototype Server-side implementation 29 (P)

D8.4.1 Prototype Client Side implementation 33 (P)

Internal deliverables

ID8.5.1 Final results evaluation 36 (R) (**Internal**)

Milestones and expected result

M2: Detailed research specification

M4: Techniques for binding media with presentation templates, Techniques for server-side transformation of I-Objects, Techniques for client side rendering of I-Objects

Work packages description (full duration of project)

Work package number	WP9	**Start date or starting event:**	1
Work package title WP9: Dissemination and Exploitation			

Participant id	**ISTI-CNR**	CMU	CWI	FhG	TUC	TUWIEN	UVA
Person-months per participant:	6.0		1.5	1.0	4.0	1.0	1.0

Objectives Development of the Consortium's strategy for effective dissemination and exploitation. Diffusion of the project results and techniques in the business environment and in the academic world. Development and establishment of exploitation plans for each Partner, coherent with the overall Consortium strategy.

Description of work

T9.1 Dissemination and exploitation planning

A strategy for exploitation of the project's results will be identified for the Consortium and for single Partners. A comprehensive plan will be produced, which will be progressively updated, to take new information arising from the project into account.

Participants: all partners

T9.2 Production of promotional material

Production of promotional material such as a Web site, project leaflet, brochures, will be done at the beginning of the project and will be updated during the project lifetime.

Participants: all partners

T9.3 Organisation of project workshops

A series of workshops will be organised to promote projects results in collaboration with related network of excellences such as DELOS and MUSCLE.

Participants: all partners

T9.4 Scientific exploitation

Scientific exploitation will be carried-out by publishing research result of the project in international journals and conferences, and through a FACS Newsletter produced every two months and delivered via e-mail and RSS.

Participants: all partners

Deliverables

D9.1.1 Initial dissemination and exploitation planning report 3 (R)

D9.2.n Production of promotion material 3, 18, 36 (O)

D9.3.n Organization of workshop 6, 18, 36 (O)

D9.1.2 Revision of dissemination and exploitation planning 18 (R)

D9.4.n Report on scientific exploitation 18, 27, 36 (R)

D9.1.3 Revision of dissemination and exploitation planning 36 (R)

Milestones and expected result

M1: Dissemination and exploitation planning, Production o promotion material

M4: Production of promotion material, Organisation of workshops

Work packages description (full duration of project)

Work package number	WP10	**Start date or starting event:**	1
Work package title WP10: Project Management			

Participant id	**ISTI-CNR**	CMU	CWI	FhG	TUC	TUWIEN	UVA
Person-months per participant:	36	1	1	3	2	2	2

Objectives This work package will ensure the overall management of the FACS project. This requires a control of the technical quality of contributions, and appropriate timing of activities. Rules for the quality control of all project activities will be defined and adopted. Progress of the work will be monitored against the milestones and the objectives defined in the project programme.

All the activities carried out in this work-package will be conducted according to the management structure of the Consortium, as described in section B5. ISTI-CNR is responsible for this work package as overall coordinator of the Project.

Description of work

Task 10.1 – Project Administration and Resource Monitoring

This task will provide support for Financial and Project Administration as well as the monitoring of appropriate and adequate resource allocation of all partners. The Project Director will be in charge of the day-to-day running of the project and the Management Board is responsible for the main strategic, and managerial decisions. The Executive Scientific Committee, coordinated by the Project Director, will handle the technical management of the project. Details of the project management structure are given in Section B5.

This task will consist of the following activities:

- *Administrative coordination*
  - Project coordination and related steering activities (Project Management Board activities)
  - Relations with the Commission, project reporting and organization of project reviews
  - Definition and signature of Consortium Agreement
  - Conflict resolution
- *Technical coordination*
  - Scientific Executive Committee activities
  - Work harmonization and management of technical decision-making process.
- *Liaison activities*
  - Participation in partner meetings
  - Participation in project reviews
  - Participation in activities with other related IST projects

Task 10.2 – Work plan monitoring and updating

This task is dedicated to two main activities:

- The monitoring of all project activities to ensure that they are carried out according to planned scheduling
- Continuous checking that the work to be done is in accordance with the original work plan. In case of delays, new activities that are considered necessary due to scientific and technology evolution, a new Management Plan will be produced.

Task 10.3 – Reporting to EU

This task will be the responsibility of the Project Director. It includes all the reporting aspects as defined by the VI Framework Programme Rules.

A **Progress Monitoring Report** will be produced every six moths. It will contain a management-level overview of the activities carried out, a description of progress toward the scientific and technological objectives, a description of progress toward the milestones and deliverables foreseen, and identification of problems encountered during the project and actions taken to correct them.

A **Financial Report** will be produced every six months containing a *summary cost statement* prepared by each participant and a *cost certificate* per participant, a *management-level justification* prepared by each participant of its overall costs incurred, linking these costs to the resources deployed and activities carried out by the participant. The financial report will also contain a *summary financial report* prepared by the coordinator and the administrative co-ordinator, bringing together the incurred costs of the consortium and the requested Community contribution.

Deliverables	
D10.1 – Consortium Agreement	1 (R)
D10.2.n – Financial reports	6, 12, 18, 24, 30, 36 (R)
D10.3.n – Progress monitoring reports	6, 12, 18, 24, 30, 36 (R)

B7. Other issues

B7.1. Ethical issues

Participants in the FACS project will conform to current legislation and regulations in their respective countries, as well as EC regulations and international conventions and declarations. No specific problem is foreseen at this stage, however, particular care will be taken into account in correspondence of the following issues:

Intellectual property right:

The issue of *Intellectual Property Rights (IPR)* is central to topics studied in FACS. Indeed, one of the goal of FACS is to develop a framework where produced material (multimedia presentations, document annotations), can be managed, distributed, and possibly *traded*, in accordance to needs of providers and consumers. However, the project intends to provide the scientific and technological advances to enable the automatic synthesis of multimedia presentations, while it does not have the objective to deal explicitly with IPR issues, which may require extensive investigations possibly considered in other research projects and in commercial implementations.

Protection of individuals:

The FACS project will be realised in accordance with the **rules of protection and processing of personal and sensible data**: Directive 95/46/CE (protection of personal date treatment) and the Directive 97/66/CE (personal data processing and tutelage of privacy in the sector's telecommunication).

The Directive on data's protection is applied to "processing" data: collected, achievement and diffusion of personal data.

The FACS project makes use of data related to the individual, like, user preference, user behaviour, user contextual information, user information needs, user interests, in order to provide users with personalised and context-aware services.

In FACS the personal data processing can be carried out only if the person gives, unambiguously, consent. The person decides, unconditionally and specifically, his/her consent after adequate information of possible risks (in accordance with European Convention of the **humans' rights** safeguard and fundamental freedom, as interpreted by European Court of human right).

Data regarding the user behaviour is automatically collected, in order to build profiles used by personalisation services. However, such user profiles are in no way publicly available and indeed are only available to software systems and not human agents.

The sensible data processing is applied with more restrictive rules, and generally is not allowed. The user can contact the responsible of treatment to know if exist a treatment of own data (**consumer's defence**). If the data are not exact, or treatment in unlawful mode, the consumer can asks the cross out.

Ethical issues form

A. Proposers are requested to fill in the following table

Does your proposed research raise sensitive ethical questions related to:	YES	NO
Human beings		X
Human biological samples		X
Personal data (whether identified by name or not)	X	
Genetic information		X
Animals		X

B. Proposers are requested to confirm that the proposed research does not involve:

Research activity aimed at human cloning for reproductive purposes,

Research activity intended to modify the genetic heritage of human beings which could make such changes heritable[14]

Research activity intended to create human embryos solely for the purpose of research or for the purpose of stem cell procurement, including by means of somatic cell nuclear transfer;

Research involving the use of human embryos or embryonic stem cells with the exception of banked or isolated human embryonic stem cells in culture.[15]

Confirmation : the proposed research involves none of the issues listed in section B	**YES**	**NO**
	X	

B7.2. Gender issues

The gender issues would be integrated into the project according to **adoption of a multi-disciplinary approach and the proposals of mainstreaming policies on gender issues.**

B7.2.1. Gender and law recommendations

FACS would cope gender issues starting from this awareness: GENDER must be considered not only as biological difference but also "*as the socially and culturally ascribed characteristics of and relations between, women and men*" and as an issues concerning "*the deep rooted values and concepts that underline the thinking, behaviour and actions of women and men in all areas of socio economic life*[16]". It would develop its programme and action considering the specific laws or Commission Communications:

- The Treaty of Amsterdam, article 2,3 (1997).
- Commission Communication (1996) "Incorporating equal opportunities for women and men into all Community policies and activities".
- Commission Communication (1999) "women and Science: mobilising women to enrich European research".
- The Community Framework strategy on Gender Equality (2001/2005) .

FACS would consider the gender dimension according with the two-pronged strategy combining mainstreaming with specific actions, that means to consider and promote:

- The participation of women and men in the project activities (**research by women**):
- The mainstreaming of gender in the management of each project cycles (**research for women**)
- The impact of the project in terms of the treatment of gender issues (**research about women)**

Through the promotion of the following actions:

[14] Research relating to cancer treatment of the gonads can be financed

[15] Applicants should note that the Council and the Commission have agreed that detailed implementing provisions concerning research activities involving the use of human embryos and human embryonic stem cells, which may be funded under the 6th Framework Programme, shall be established by 31 December 2003. The Commission has stated that, during that period and pending establishment of the detailed implementing provisions, it will not propose to fund such research, with the exception of the study of banked or isolated human embryonic stem cells in culture.

[16] "Gender in Research" a study for the European Commission by Mary Braithwaite, Tacitus sprl, April 2001, Executive Summary, pg.5.

- It would be equal represented men and women into the consortium, not only in researchers, but also in validation phases, not to waste female scientific potentialities that are often lost in scientific careers. (suggested women percentage is 40%)
- It would be also avoided an unequal representation of women in decision making processes and leadership
- It would be promoted a proactive approach to stimulate the gender sensitivity of interactions between management bodies and project coordinators at the negotiation and implementation stage and specially to avoid the adoption of neutral policies or solution.

Appendix A

FACS partners are involved in a number of European and National projects that are related to FACS activities. This is a list of these projects.

- **Eureka/ITEA project Passepartout** (2005-2006), aims at studying the selective capture of broadcast content and its distribution to dissimilar devices within a home network environment. Passepartout provides a solid base of experience with existing content storage, subsetting and distribution based on heuristic user needs and profiling. The Passepartout project does not address the direct semantic and rhetoric construction of content, however, nor does it consider the on-demand generation of content for the scaled distribution to a range of user devices. CWI is active in the project.
- **National project Ambulant (Stichting NLnet),** has as its goal the development of open source reference implementations of standards-based media players (specifically those based on the W3C SMIL standard). We expect to be able to reuse technology developed by Ambulant as part of the FACS demonstrator. CWI is active in the project.
- **ECHO**, was IST KA III project coordinated by ISTI-CNR. It follows an open architecture approach to distributed digital film archive services providing semi-automatic metadata extraction and acquisition from digital film information, non-English speech recognisers (Italian, French, Dutch) for the purpose of indexing, searching and retrieval, cross-language retrieval capabilities, intelligent access to digital films, automatic film summary creation, collection mechanisms, privacy and billing mechanisms. ISTI-CNR was the coordinator of ECHO.
- **DELOS NoE**, coordinated by ISTI-CNR, will define unifying and comprehensive theories and frameworks over the life cycle of Digital Libraries information and it will build interoperable multimodal/multilingual services and integrated content management. A strong relationship will be established with DELOS in order to disseminate and promote project results and to have a continuous view of the evolution of multimedia content management in Digital Libraries. ISTI-CNR is coordinator of this project.
- The **MUSCLE Network of Excellence** aims at supporting a research network to encourage close collaboration between research groups in multimedia datamining and machine learning. UVA and TUWIEN are partners in the Network.
- The **CoreGRID Network of Excellence** aims at strengthening and advancing scientific and technological excellence in the area of Grid and Peer-to-Peer technologies. ISTI-CNR is involved in this Network.
- **Eureka/ITEA project PELOPS.** The Project has the objective to automate the content creation, metadata extraction and production of live sport events. It will invest in the automated generation, collection, and re-use of metadata. ISTI-CNR is partner in the project.
- The **DILIGENT** IP project is coordinated by ISTI-CNR. It will create an advanced test-bed that will allow virtual e-Science communities to share knowledge and collaborate in a secure, coordinated, dynamic and cost-effective way. The DILIGENT test-bed will be built by integrating Grid and Digital Library (DL) technologies.
- **Informedia,** The Informedia project at CMU was initiated in 1994 as one of six national Digital Library Initiative (DLI) projects; it has pioneered numerous advances in automated video understanding. CMU is involved in this project.
- **VACE – Video Analysis and Content Extraction** (US – ARDA) This basic research program pursues breakthrough advances in semantic object detection, activity understanding, and event recognition from a variety of video sources from broadcast news to aerial imagery. CMU is involved in this project.
- **AQUAINT – Advanced Question Answering for Intelligence** (US – ARDA) This research program develops technology for automated question answering addressed as a cooperative, information-gathering process between a user and the system in which each has its own information context. CMU is involved in this project.
- **ECVision - European research network for cognitive computer vision systems** is a research network which was formed to promote research, education, and application systems engineering in cognitive AI-enabled computer vision through focussed networking, multi- disciplinary peer-interaction, targeted identification of priority issues, and wide-spread promotion of the area's challenges and successes within both the academic and industrial communities. CMU is involved in this project.
- **CHIRON,** that develops reference material presenting and analysing research outcomes, experiments and best practice solutions for new forms of e-learning, based on integration of broadband web-, digital TV- and mobile

technologies for ubiquitous applications in the sector of non-formal and informal life-long learning. TUC is partner in the project.

- **CAMPIELLO**, that was experimenting with the use of innovative ICT to develop new links between the local communities and the visitors of the Historical Cities of Art and Culture. The project has developed a system to support the information flow and the processes between the different actors involved within a defined physical environment and appropriate interaction paradigms adapted to the specific users, content and physical environment of the Historical Cities. TUC is partner in the project.
- **ECVision** - European research network for cognitive computer vision systems is a research network which was formed to promote research, education, and application systems engineering in cognitive AI-enabled computer vision through focussed networking, multi- disciplinary peer-interaction, targeted identification of priority issues, and wide-spread promotion of the area's challenges and successes within both the academic and industrial communities. UVA is partner in the project.
- **National project MultimediaN (Multimedia Netherlands)** involves the knowledge creation and transfer on handling of video, pictures, audio, and language in ICT. Knowledge creation is focused on four themes: processing Multimedia data streams: combined video, audio and speech gaining, extracting, and analysis; Connecting and interacting to different modularities of computers around us: ambient databases and emotional interfaces; Providing semantic access to multimedia databases: internet next generation search technology and individual adapted interfaces; Enriching content and discover knowledge by intelligent and adaptive pattern recognition, adding structure and annotations to multimedia data. UVA is partner in the project.
- **Image and Video Analysis for Artistic Rendering**: This project, funded by the "Hochschuljubilaeumsstiftung der Stadt Wien" focuses on the development of algorithms for non-photorealistic rendering of real images and video scenes. It is devoted to simulation of human sketching ability to present various thoughts, ideas and to describe the world around just with few rough strokes. Our focus is to present real scenes as stylized sketches by employing depth information to select only significant contours that play the most important role in object description. TUWIEN is involved in this project.
- **3D Video Analysis for Interactive Multimedia Applications**: The Video3D project focuses on the 3D analysis of moving scenes that were captured by two (or more) synchronized video cameras. The work entails research in the fields of stereo analysis and optical tracking, human motion capture, and MPEG encoding. The project is funded by the Austrian Science Fund (FWF). TUWIEN is involved in this project.
- **VizIR**: VizIR project aims to provide a set of methods for visual information retrieval. These include methods for feature extraction, similarity definition, query processing and user interfaces for querying and refinement. VizIR will implement most of the visual descriptors in the MPEG-7 standard. TUWIEN is involved in this project.
- **SOMeJB: The SOM-enhanced JukeBox**: The SOMeJB system builds upon the principles of the SOMLib system to create a digital library for music by combining a variety of technologies from the fields of audio processing, neural networks, and information visualization. It is based on the self-organizing map (SOM), a popular unsupervised neural network, and its extension, the Growing Hierarchical Self-Organizing Map (GHSOM), used to organize pieces of music available as, e.g., mp3 files, according to their musical sound characteristics, i.e. creating a kind of genre-based organization. Islands of Music and Weather Charts provide an intuitive interface to the system. TUWIEN is involved in this project.

Appendix B

FACS activities are related to a number of European and National projects. This is a list of these projects.

- **HyNoDe**, a service platform, supporting Multimedia News Stories Authoring, efficient storage and Retrieval of News Stories, and Personalized delivery to the end-users.
- **COSMOS**, a mobile middleware platform supporting IP based workflow applications, running on top of wireless and satellite infrastructures.
- **iMEDIA**, an end-to-end solution for providing targeted interactive advertising in a DTV environment. It is a mediator, a marketplace between the Content Providers (TV Channels), the Advertising Companies/Advertisers and the provider of the DTV infrastructure (Network Provider).
- **UPTV**, that creates advanced and expandable architectures and systems for TV-Anytime applications. UP-TV aims to provide personalized access to broadband information in an interactive way, allowing the user to browse through the content, search for selected content items and organize the content in various ways independently from time and place
- **UWA**, that defines a set of methodologies, notations, and tools to support the design and fast prototyping of complex, multi-device, ubiquitous web applications.
- **MyTV**, a platform that will enable consumers to have access to content and services at their convenience, independent of the moment of broadcasting supporting non-linear browsing of television content, interactive and targeted advertising, and easy navigation through the massive amount of content offered.
- **ShareIT**!, an end-to-end system that enables easy access to, and transfer of, personal content between local storage devices using home-to-home networks and that enables innovative services through a seamless combination of on-line, broadcast, and stored content.
- **SAMBITS**, brings MPEG-4 and MPEG-7 technology to the broadcast industry and the related Internet services. It provides multimedia services to a terminal that can display any type of general interest integrated broadcast/internet services with local and remote interactivity.
- **PISTE**, an advanced system for interactive coverage of sports events, offering a high level of graphical information for statistical purposes, choice of viewing conditions and other.
- **MELISA**, successor of PISTE, additionally involving enabling of interactive in-play betting services as well as embedded advertising.
- The **COLLATE** Collaboratory project comes close to what is ultimately needed in cultural-heritage knowledge disclosure: it "aims at the development and practical usage of a content-centric, user-driven information system for the management of surrogates of fragile historic multimedia objects. As a distributed Web-based multimedia repository, it will function as a 'collaboratory' supporting distributed user groups by dedicated knowledge management facilities such as content-based access, comparison and in-depth indexing/annotation of digitised sources."
- **CHIMER** (Children's Heritage Interactive Models for Evolving Repositories; http://dbs.cordis.lu/fep-cgi/srchidadb). CHIMER aims to establish an open international network of children, teachers and museologists for developing an Open Evolving Multimedia Multilingual Digital Heritage Archive as a long-term storage medium for European cultural repositories.
- **COINE** (Cultural Objects in Networked Environments) (http://dbs.cordis.lu/fep-cgi/srchidadb). Empowering European citizens to tell their own stories lies at the heart of the COINE (Cultural Objects in Networked Environments) Project. It will provide the tools needed to create structured, World Wide Web-based environments which are hospitable to local cultural activity but which allow content to be shared locally, regionally, nationally and internationally.
- **ECHO** (European Cultural Heritage Online) (http://echo.mpiwg-berlin.mpg.de, http://www.mpi.nl/echo) is a new project that has as task to provide a rich interdisciplinary access to objects of cultural heritage. Aspects of interoperability at the metadata level between the 4 included disciplines are one of the core aspects.
- **INTERA**: (Integrated European language Resource Area) is an attempt to solve interoperability problems on a vertical line by creating not only a large metadata domain of language resources, but also by integrating the domain of resource descriptions with those of tool descriptions. The goal is that dependent on the type of selected resources appropriate tools will be selected automatically.
- **AXMEDIS** (Automating Production of Cross Media Content for Multichannel Distribution) is an IP Project aiming at providing a solution to automate, accelerate, and restructure the production process of cross media content.

- **PENG** (PErsonalised News content programming) is a project aiming at defining a flexible, personalized and context-aware system for the gathering, filtering, retrieval and presentation of multimedia news for news professionals
- **REVEAL THIS** (Retrieval of Video and Language for The Home user in an Information Society) project aims at developing content programming technology able to capture, semantically index, categorize and cross-link multi-platform, multimedia and multilingual digital content, as well as provide the system user with semantic search, retrieval, summarization and translation functionalities.

Appendix C – References

[1] "EGEE gLite User's Guide, Overview Of gLite Data Management", Document identifier: EGEE-TECH-570643-v1.0, March 4, 2005, https://edms.cern.ch/document/570643
[2] "GRACE, Grid search & Categorisation Engine", project "IST-2001-38100" in the IST 2002 Framework 5, under CPA9: "Grid technologies and their applications", http://www.grace-ist.org/
[3] "Lore" http://www-db.stanford.edu/lore/
[4] "MPEG - Moving Picture Expert Group", http://www.chiariglione.org/mpeg/
[5] "OGSA-DAI Open Grid Services Architecture Data Access and Integration OGSA-DAI", http://www.ogsadai.org.uk/
[6] "OWL Web Ontology Language Reference" W3C Recommendation 10 February 2004, http://www.w3.org/TR/2004/REC-owl-ref-20040210/
[7] "P3P - Platform for Privacy Preferences", http://www.w3.org/P3P/
[8] "RDF – Resource Description Framework" http://www.w3.org/RDF/
[9] "Tamino", http://www.tamino.com
[10] "Timber" http://www.eecs.umich.edu/db/timber/
[11] "Xindice" http://xml.apache.org/xindice/
[12] "XPath, XML Path Language" http://www.w3.org/TR/xpath
[13] "XQuery, an XML Query Language" http://www.w3.org/TR/xquery/
[14] A. Rauber, E. Pampalk, D. Merkl: The SOM-enhanced JukeBox: Organization and Visualization of Music Collections based on Perceptual Models. In: Journal of New Music Research (JNMR), 32(2):193-210, Swets and Zeitlinger, June 2003.
[15] A. Vetro, "Mpeg-21 digital item adaptation: enabling universal multimedia access," IEEE Multimedia, vol. 11, no. 1, pp. 84–87, Jan.-Mar. 2004.
[16] A.W.M. Smeulders, F. de Jong: Multimedia information technology for indexing EU-conf video archives, invitation .
[17] A.W.M. Smeulders, L. Hardman, G. Schreiber, JM Geusebroek: An integrated multimedia approach to cultural heritage, ACM- MIR 2002.
[18] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, R. Jain: Content-based image retrieval at the end of the early years, IEEE PAMI, 1349–1380, 2000.
[19] An, Y., Borgida, A., Mylopoulos, J.: Refining Semantic Mappings from Relational Tables to Ontologies. In the Proceedings of SWDB 2004
[20] Arnheim, R. Art and Visual Perception: A Psychology of the Creative Eye. University of California Press, Berkeley, 1974.
[21] B. Logan. Mel Frequency Cepstral coefficients for Music Modelling. In Proceedings International Symposium on Music Information Retrieval (ISMIR 2000), Plymouth, MA. 2000.
[22] Bernstein, P., Giunchiglia, F., Kementsietsidis, A., Mylopoulos, J., Serafini, L., Zaihrayeu, I.: Data Management for Peer-to-Peer Computing: A Vision. In the Proceedings of the Int'l Workshop on the Web and Databases, 2002
[23] Bordwell, D. and Thompson, K. Film Art: An Introduction. McGraw Hill, New York, 2003.
[24] C. Snoek, M. Worring, J. Geusebroek, D. Koelma, F. Seinstra: The MediaMill TRECVID 2004 Semantic Video Search.
[25] C. Tsinaraki, E. Fatourou, S. Christodoulakis, "An Ontology-Driven Framework for the Management of Semantic Metadata describing Audiovisual Information", in Proc. of CAiSE, Velden, Austria, 2003, pp 340-356
[26] C. Tsinaraki, P. Polydoros, F. Kazasis, S. Christodoulakis, "Ontology-based Semantic Indexing for MPEG-7 and TV-Anytime Audiovisual Content", Special issue of Multimedia Tools and Applications Journal on Video Segmentation for Semantic Annotation and Transcoding, 2004
[27] C. Tsinaraki, P. Polydoros, S. Christodoulakis, "Integration of OWL ontologies in MPEG-7 and TVAnytime compliant Semantic Indexing", in Proceeding of CAiSE 2004
[28] C. Tsinaraki, P. Polydoros, S. Christodoulakis, "Interoperability support for Ontology-based Video Retrieval Applications", in Proceedings of CIVR 2004
[29] C. Tsinaraki, P. Polydoros, S. Christodoulakis, "Interoperability support for Ontology-based Video Retrieval Applications", in the CIVR 2004 special issue of the IEE VIS Journal
[30] Creative commons, Website, www.creativecommons.org
[31] Downie, J.S.: Music information retrieval . In: Annual Review of Information Science and Technology 37:295-340, Information Today, Medford, NJ, 2003.
[32] E. Wold, T. Blum, D. Kreislar, and J. Wheaton. Content-based Classification, Search, and Retrieval of Audio. IEEE Multimedia 3(3):27-36, 1996.

[33] Elisabeth Andre. WIP and PPP: A Comparison of two Multimedia Presentation Systems in Terms of the Standard Reference Model. Computer Standards & Interfaces 18 (1997), pp. 555-563.
[34] G Schreiber, B Dubbeldam, J Wielemaker, B Wielinga. Ontology-based photo annotation IEEE Intell Systems 2001
[35] G. Sten, J. Rittscher, A. Hoogs. Enabling video annotation using a semantic database extended with visual knowledge, ICME 2003.
[36] Giunchiglia, F., Zaihrayeu, I.: Implementing Database Coordination in p2p Networks. In the Proceedings of the 2nd Workshop in Peer-toPeer and Grid Computing (SemPGrid'04), 2004
[37] Hieu Tat Nguyen, A. W. M. Smeulders: Everything gets better all the time apart from the amount of data. CIVR 2004, Dublin.
[38] Internet Archive: audio and video holdings, Website, www.archive.org
[39] ISMIR, International Conference on Music Information Retrieval, Website, www.ismir.net
[40] J.M. Geusebroek, R. van den Boomgaard, A.W.M. Smeulders, H. Gerards: Kleuren invarianten, IEEE trans PAMI, 2002. [Geusebroek 2004] J.M. Geusebroek, G.J Burghouts, A.W.M. Smeulders: The ALOI data set, Int. Jounr. Comp. Vis, 2005
[41] J.S. Downie. Annual Review of Information Science and Technology, volume 37, Chapter on „Music Information Retrieval", pp. 295-340, Information Today, Medford, NJ, 2003. http://music-ir.org/downie_mir_arist37.pdf
[42] J.T. Foote. Content-based Retrieval of Music and Audio. In Proceedings of SPIE Multimedia Storage and Archiving Systems II, 3229:138-147. 1997.
[43] John Bateman, Jörg Kleinz, Thomas Kamps, and Klaus Reichenberger. Towards Constructive Text, Diagram, and Layout Generation for Information Presentation, In: Computational Linguistics 27(3), pp. 409-449, September 2001.
[44] K. Mikolajczyk, C. Schmid: Scale and affine invariant interest point detectors. Nt. Journ. Comp. Vis 63 – 86, 2004.
[45] Kantere, V., Kiringa, I., Mylopoulos, J., Kementsietsidis, A., Arenas, M.: Coordinating Peer Databases Using ECA Rules. In the Proceedings of VLDB 2003
[46] Kementsietsidis, A., Arenas, M., Miller, R.J.: Data mapping in peer-to-peer systems: Semantics and algorithmic issues. In ACM SIGMOD 2003
[47] Klyne G., Reynolds F., Woodrow C., Ohto H., Hjelm J., Butler M., Tran L., "Composite Capability/Preference Profiles (CC/PP): Structure and Vocaburaries 1.0", W3C recommendation, January 2004
[48] L. Hollink, G. Schreiber, B. Wielinga, M. Worring. Classification of User Image Descriptions. Int. J. Human-Com Studies, 2004.
[49] LegalTorrents, Website, www.legaltorrents.com
[50] M. Nedim Alpdemir, Arijit Mukherjee, Anastasios Gounaris, Norman W. Paton, Paul Watson, Alvaro A.A. Fernandes: "OGSA-DQP: A Grid Service for Distributed Querying on the Grid", Proc. 9th International Conference on Extending Database Technology (EDBT), 858-861, 2004.
[51] N. Cristianini and J. Shawe-Taylor . An introduction to support vector machines and other kernel-based learning methods, Cambridge University Press, 2000
[52] N. Maillot, M. Thonnat, and A. Boucher. Towards ontology-based cognitive vision. MVA Journal, 16(1):33–40, 2004.
[53] Niels Ole Bernsen. Defining a Taxonomy of Output Modalities from an HCI Perspective, In: Computer Standards and Interfaces 18, pp. 537-553, 1997.
[54] Noah Wardrip-Fruin (Editor), Nick Montfort (Editor), *The New Media Reader*, The MIT Press
[55] P. Bouquet and F. Giunchiglia and F. van Harmelen and L. Serafini and H. Stuckenschmidt, "C-OWL: Contextualizing Ontologies", Proc. of the 2nd International Semantic Web Conference
[56] Pampalk, E., Rauber, A., Merkl, D.: Content-based Organization and Visualization of Music Archives. In Proceedings of ACM Multimedia 2002, 570-579, Juan-les-Pins, France, December 2001. ACM.
[57] J. Ossenbruggen, J. Geurts, F. Cornelissen, L. Rutledge, and L. Hardman: "Towards Second and Third Generation Web-Based Multimedia" - Tenth International World Wide Web Conference, pp. 479-488, May 1-5, 2001, Hong Kong.
[58] J. Ossenbruggen, J. Geurts, L. Hardman, and L. Rutledge, "Towards a Formatting Vocabulary for Time-based Hypermedia" - Twelfth International World Wide Web Conference, pp. 384-393), ACM Press, Budapest, Hungary, May 20-24, 2003.
[59] J. Geurts, J. Ossenbruggen, and L. Hardman, "Application-Specific Constraints for Multimedia Presentation Generation" - Proceedings of the International Conference on Multimedia Modelling 2001 (MMM01), pp 247-266), November 5-7, 2001, CWI, Amsterdam.
[60] Podcast.net, Website www.podcats.net
[61] Prud'hommeaux E., Seaborne A. (Eds.), "SPARQL Query Language for RDF", W3C Working Draft, October 2004, http://www.w3.org/TR/rdf-sparql-query/
[62] R. Fergus, P. Perona, A. Zissermann: Object class recognition by unsupervised scale invariant learning, 2003.

[63] S.Christodoulakis, C.Tsinaraki (editors), "State of the Art in Audiovisual Content – Based Retrieval, Information Universal Access & Interaction including Datamodels & Languages", book submitted for publication
[64] T. Blum, D.F. Keislar, J.A. Wheaton, E.H. wold. Method and Article Manufacture for Content-based Analysis, Storage, Retrieval, and Segemntation of audio Information. U.S. Patent 5.918.223, 1999.
[65] T. Gevers, A Smeulders: Color based object recognition, PattRecognition, 453-464, 1999.
[66] T. Kohonen. Self-Organizing Maps. Springer, 2001.
[67] TREC Video retrieval evaluation, 2001-2004, http://www-nlpir.nist.gov/projects/trecvid/
[68] V. Jijkoun, M. de Rijke, and J. Mur. Information Extraction for Question Answering: COLING 2004)
[69] William C. Mann, Christian M. I. M. Matthiesen, and Sandara A. Thompson. Rhetorical Structure Theory and Text Analysis, technical report ISI/RR-89-242, November 1989.
[70] McCarthy, J., "Generality in Artificial Intelligence", Communication of the ASM. Vol. 30, No. 12, pp,1030-1035, 1987
[71] Ghidini C., Giunchiglia F., "Local Models, Semantics, or Contextual Reasoning = Locality+Compatibility", Artificial Intelligence 127(2), pp. 221-259, April 2001
[72] Lee M., Wilks Y., "An ascription-based approach to speech acts", in Proceedings of the 16th International Conference in Computational Linguistics COLING-96, 1996
[73] Giunchiglia F., Serafini L., "Multilanguage hierarchical logics, or how can we do without modal logics", Artificial Intelligence 65 (1), pp. 29-70, 1994
[74] Parsons S., Sierra C., Jennings N.R., "Agents that reason and negotiate by arguing", Journal of Logic and Computation, 8, 3, pp. 261-292, 1998
[75] Crystal, D. "A Dictionary of Linguistics and Phonetics", 3rd ed. Blackwell, Oxford, UK, 1991
[76] Poesio, M., Traum, D. R., "Conversational actions and discourse situations", Computational Intelligence, Volume 13 Issue 3 Page 301, 1997
[77] Kamp, H., "A theory of truth and semantic representation", In Formal Methods in the study of Language. Edited by J. Groenendijk, T. Janssen, and M Stokhof. Mathematical Center, Amsterdam 1981
[78] Asher N., Lascarides A. "Logics of Conversation", Cambridge, UK, Cambridge University Press, 2003.
[79] Zadrozny, S., Kacprzyk, J. "FQUERY for Access: towards human consistent querying user interface", Proceedings of the 1996 ACM symposium on Applied Computing
[80] Belloti, R., Decurtins, C., Grossniklaus, M., Norrie, M.C., Palinginis, A. "Modelling Context for Information Environments", In Proceedings of Ubiquitous Mobile Information and Collaboration Systems, Riga, Latvia 2004
[81] Klemke, R., "Context Framework – an Open Approach to Enhance Organizational Memory Systems with Context Modelling Techniques". In Proceedings of the Third International Conference on Practical Aspects of Knowledge Management, 2000
[82] Schilit, B., Adams, N., Want, R., "Context-aware computing applications." In Proceedings of IEEE Workshop on Mobile Computing Systems and Applications, pages 85-90, Santa-Cruz, California, December 1994
[83] Bulterman, D.C.A, Hardman, L., Jansen, J. Mullender, K.S. and Rutledge, L. "GRiNS: A GRaphical Interface for Creating and Playing SMIL Documents", Computer Networks and ISDN systems, 10, (1998), pp. 519-529.
[84] Bulterman, D.C.A and Rutledge, SMIL 2.0: Interactive Multimedia for Web and Mobile Devices, Springer-Verlag, Heidelberg, 2004.
[85] Oratrix Development BV., GRiNS, http://www.oratrix.com/GRiNS/.
[86] Weitzman, L., and Wittenberg, K. "Automatic Presentation of Multimedia Documents UsingRelational Grammars", Proceedings of ACM Multimedia 94, San Francisco, USA, September 1994, pp. 443--451.
[87] Zhou, M.X. "Visual Planning: A Practical Approach to Automated Presentation Design" Proceedings of the International Joint *Conference on Artificial Intelligence 99 (IJCAI 99)*, pp. 634-641.
[88] A. Hanjalic, L-Q. Xu, Affective vides Content Representation and Modelling, IEEE Trans. on Multimedia, vol. 7, no. 1, pp. 143-154, February 2005.
[89] Li, T. and Mitsunori O. Detecting Emotion in Music. Proceedings of the International Conference on Music information Retrieval (ISMIR 2003), pp. 239-240, Baltimore, MD, Oct. 26-30,2003.
[90] Liu, D., Lu L., Zhang H.J. Automatic Mood Detection from Acoustic Music Data. Proceedings of the International Conference on Music information Retrieval (ISMIR 2003), pp 81-87, Baltimore, MD, Oct. 26-30,2003.
[91] StreamSage, Web Site http://www.streamsage.com
[92] Infomedia Project, Web Site http://www.informedia.cs.cmu.edu/
[93] Pustejovsky, J., Ingria, B., Sauri, R., Castano, J., Littman, J., Gaizauskas, R., Setzer, A., Katz, G., and Mani, I, "The Specification Language TimeML".. 2004. in The Language of Time: A Reader, Mani, I., Pustejovsky, J., and Gaizauskas, R. (eds.). 2004.. Oxford University Press, to appear.
[94] Pustejovsky, J., "TimeML Specification (Version 1.2)" [HTML Document], in TimeML Documents. http://www.timeml.org/timemldocs.html

[95] Wiebe, J., et.al., "Recognizing and Organizing Opinions Expressed in the World Press" in New Directions in Question Answering, Maybury, M (ed.),2003 AAAI Spring Symposium, AAAI Press, Palo Alto, CA, March 2003

[96] Stoyanov, V., Cardie, C., Litman, D., Wiebe, J., "Evaluating an Opinion Annotation Using a New Multi-Perspective Question and Answer Corpus". To appear in AAAI-2004 Spring Symposium on Attitude and Affect in Text(EAAT-2004)

[97] Wactlar, H.D., Kanade, T., Smith, M.A., and Stevens, S.M. Intelligent Access to Digital Video: Informedia Project, *IEEE Computer*, 29(5), 46-52.

[98] Wactlar, H.D., Christel, M.G., Gong, Y., and Hauptmann, A.G. Lessons Learned from the Creation and Deployment of a Terabyte Digital Video Library, *IEEE Computer* 32(2): 66-73.

[99] Christel, M., Winkler, D., and Taylor, R. Multimedia Abstractions for a Digital Video Library, ACM Digital Libraries 97 (Philadelphia, PA, July 1997), 21-29.

Appendix D – CV of key project personnel

This appendix contains the CV of key persons involved in the project. The list is subdivided per partner.

Consiglio Nazionale delle Ricerche – Istituto di Scienze e Tecnologie dell'Informazione "Alessandro Faedo" (ISTI-CNR)

Pasquale Savino graduated in Physics at the University of Pisa, Italy, in March 1980. In April 1983 he received a post graduate specialization in Computer Science (Corso di Specializzazione in Calcolo Automatico) from the University of Pisa. From 1983 to 1995 he has worked at the Olivetti Research Labs in Pisa where he coordinated projects in the area of multimedia information systems and multimedia communication; since 1996 he has been a member of the research staff at CNR-IEI in Pisa, working in the area of multimedia information systems. Actually he is senior researcher at ISTI-CNR. He has participated and coordinated several EU-funded research projects in the area of multimedia information systems. In particular, he was

- The project leader and project manager of the ESPRIT Project MULTOS (Multimedia Office Systems). The project, started in June 1985 and completed in June 1990 required an effort of approximately 100 man/years. Main topics addressed were the management and content-based retrieval of multimedia documents.
- The project leader of the work-part on multimedia databases and information retrieval in the ESPRIT Project TROPICS (1989-1990). Main objective of TROPICS was the design and development of a parallel computer to be used in several application environments.
- The project manager of the ESPRIT Project OSMOSE (Open Standard for Multimedia Optical Storage Environments) (1990-1994). The project had the purpose to define and experiment on multiple platforms a standard representation of multimedia interactive applications.
- The project manager of the EPSRIT Project HYTEA (HYperText Authoring), (1990-1993). The project had the objective to define and implement a set of tools for the design and development of multimedia interactive applications.
- The project manager of the DELTA Project MALIBU (Multimedia And distance Learning In Banking and bUsiness environments) (1992-1995). The project had the purpose to develop, experiment and evaluate computer based training applications using telecommunication services.
- The project manager of the EUREKA Project ACROPOL (1992-1994). The project had the objective to design, develop and experiment multimedia information kiosks to be used by citizens in several European cities.
- The project manager of the ESPRIT Project M-CUBE (Multiple Media Multiple Communication Workstation) (1992-1995). This project has been dedicated to the definition, development and experimentation of multimedia interactive systems able to support the remote communication among users geographically distributed. The communication was based on the use of public networks (ISND) and satellite TV. The project had a cost of approximately 72 man/years.
- The project manager of the ESPRIT Project OASIS. During the project were developed and experimented multimedia applications and kiosks to be used as a support for remote product sales.
- He was involved in the CEC-funded ESPRIT LTR Project HERMES (Foundations of High Performance Multimedia Information Management Systems), where he was the lead of the Work Part on Multimedia Content Addressability. This Project, completed in September 1998, had the purpose to study and experiment new techniques for the archiving and the content-based retrieval of multimedia data.
- He was the CNR Project Manager of the EC funded Project "Eurosearch" Telematics Application Program – Language Engineering.
- He was the CNR Project Manager of the EC funded Project "EUROgatherer - Personalised Information Gathering System" Telematics Application Program – Information Engineering.
- He was the Project Manager of the EC funded Project ECHO (European CHronicles On-line) KA III – Multimedia Content and Tools.

His research activity is focused on the definition and experimental evaluation of new access structures supporting an efficient and effective retrieval of text documents and images; he has also worked on the definition of an appropriate architecture for a system supporting the archival and content-based retrieval of multimedia documents. Recently he worked on the definition of a model tailored for multimedia data representation and on a Multimedia Query Language that exploits the features of the proposed model. He has also contributed to the definition of new access structures supporting similarity retrieval.

Moreover, he is working in the area of Multimedia Digital Libraries, in particular video digital libraries.

He has published more than 70 scientific papers in international journals and conferences in the areas of multimedia document retrieval and information retrieval.

Giuseppe Amato (born 1968) graduated in Computer Science at the University of Pisa, Italy, in 1992 and achieved his PhD in Computer Science at the University of Dortmund, Germany, in 2002. In 1994 he was a member of the research staff at CNR-CNUCE in Pisa, working in the area of object-oriented databases and persistent programming languages. Since 1995 he has been a member of the research staff at CNR-IEI in Pisa, working in the area of Multimedia Information Systems, and Digital Libraries. He is currently a permanent member of the research staff at the ISTI-CNR in Pisa. His main research interests are content-based retrieval of multimedia documents, access methods for similarity search of multimedia documents, metadata for multimedia documents, multimedia digital libraries, XML native databases. He has published several papers in international journals and conferences in the areas of information systems and multimedia information retrieval. He has participated in several EC funded research actions in the areas of object-oriented databases and content-based retrieval of multimedia data. He was involved in

- the EC funded ESPRIT project FIDE-2, where he investigated concurrency control strategies and for object oriented persistent programming languages,
- the EC funded ESPRIT project HERMES, where he worked at a model and query language for content based retrieval of videos and access structures for similarity search,
- the EC funded "EUROGgatherer –Personalised Information Gathering System" Telematics Application Program – Information Engeneering, where he worked at the development of a personalised filtering systems for on-line textual information,
- the EC funded ECHO (Europeean Chronicles On-line) Key Action III, Multimedia Content and Tools, where he investigated metadata for multimedia documents and access methods for approximate similarity search.
- the EC funded DELOS Network of Excellence, where he is mainly involved in investigating native XML databases specialized for multimedia metadata management.

Fosca Giannotti graduated in Computer Science (Laurea in Scienze dell'Informazione, summa cum laude) in 1982, from the University of Pisa. She is currently a senior researcher at Institute of Information Science and Technologies (ISTI-CNR) of the Italian National Research Council in Pisa where she leads the Knowledge Discovery and Delivery Laboratory. She has been a researcher at the R&D labs of several Italian software companies, and a visiting scientist at The Microelectronics and Computer Technology Corporation (MCC), Austin, Texas, USA, involved in the LDL (Logic Database Language) project. Her current research interests include data mining query languages, knowledge discovery support environment, web-mining, spatio-temporal reasoning, and spatio-temporal data mining. She has been involved in several research projects both at national and international level, holding both management and research positions. She has been the coordinator of DeduGIS, an Esprit working group funded by the European Commission, with ten academic and industrial partners from four UE countries. She is currently responsible for the Action on Web Mining of the national project ECD funded by the Italian Government. She has taught classes on databases and data mining at universities in Italy and abroad, and has coordinated a master course on data mining in collaboration with the faculty of Economics at the University of Pisa. She served in the organization and in the scientific committee of various conferences in the area of logic programming, databases and data mining. She served as program chair of ECML/PKDD 2004 conference on Machine Learning and Knowledge Discovery in Data Bases.

Publications relevant to the project:

- G. Amato, G. Mainetto, P. Savino, "An Approach to Content-Based Retrieval of Multimedia Data", *Multimedia Tools and Applications*, Kluwer Academic Publishers, Vol. 7, N. 1/2, pp. 9-36, July 1998.
- P. Zezula, P. Savino, G. Amato, F. Rabitti, "Approximate similarity retrieval with M-tree", *The VLDB Journal,* Vol. 7, N. 4, pp. 275-293, 1998.
- P. Savino, "Building an Audio-visual Digital Library of Historical Documentaries: the ECHO Project", D-Lib Magazine, Vol. 6, No. 11, November 2000
- G. Amato, F. Rabitti, P. Savino, P. Zezula, , "Region Proximity in Metric Spaces and its use for Approximate Similarity Search", ACM Transactions on Information Systems, Vol. 21, No. 2, April 2003, pp. 192-227
- C. Gennaro, F. Rabitti, P. Savino, "The use of XML in a Video Digital Library", in Intelligent Search on XML Data Applications, Languages, Models, Implementations, and Benchmarks, Lecture Notes in Computer Science, Vol. 2818 Blanken, H.; Grabs, T.; Schek, H.-J.; Schenkel, R.; Weikum, G. (Eds.) 2003, ISBN: 3-540-40768-5
- V. Dohnal, C. Gennaro, P. Savino, P. Zezula, "Similarity Join in Metric Spaces", ECIR'03, pp. 452-467
- G. Amato, F. Falchi, C. Gennaro, F. Rabitti, P. Savino, P. Stanchev, "Selection of MPEG-7 Image Features for Improving Image Similarity Search on Specific Data Sets", 7th IASTED International Conference on COMPUTER GRAPHICS AND IMAGING (CGIM 2004) August 17-19, 2004 Kauai, Hawaii, USA
- G. Amato, C. Gennaro, F. Rabitti, P. Savino, "Milos: a Multimedia Content Management System for Digital Library Applications", 8th European Conference on Digital Libraries, 12-17 September 2004, University of Bath (UK)

- G. Amato, F. Falchi, C. Gennaro, F. Rabitti, P. Savino, P. Stanchev, "Improving Image Similarity Search Effectiveness in a Multimedia Content Management System", International Workshop on Multimedia Information Systems August 25 - 27, 2004 Washington DC, USA
- G. Amato, F. Debole, F. Rabitti, P. Savino, P. Zezula, "A Signature-Based Approach for Efficient Relationship Search on XML Data Collections", *Second International XML Database Symposium(XSym 2004)* In conjunction with VLDB 2004 Toronto, Canada, 29-30 August 2004
- F. Bonchi, F. Giannotti, C. Gozzi, G. Manco, M. Nanni, D. Pedreschi, C. Renso, and S. Ruggieri, Web log data warehousing and mining for intelligent web caching. Data and Knowledge Engineering (DKE), 32(2):165-189, October 2001.
- F. Giannotti, C. Gozzi, G. Manco, Clustering Transactional Data. PKDD 2002: 175-187 2001
- Fosca Giannotti, Cristian Gozzi, Giuseppe Manco: Characterizing Web User Accesses: A Transactional Approach to Web Log Clustering. ITCC 2002: 312-324
- V. Bacarella, Fosca Giannotti, Mirco Nanni, Dino Pedreschi: Discovery of ads web hosts through traffic data analysis In Proceedings of the 9th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery June 13, 2004, Maison dela Chimie, Paris, France. Pp. 76-81.
- Fosca Giannotti, Mirco Nanni, Dino Pedreschi, F. Samaritani: WebCat: Automatic Categorization of Web Search Results. SEBD 2003: 507-518

Carnegie Mellon University (CMU)

Howard Wactlar is Vice Provost for Research Computing, Associate Dean, and Alumni Research Professor in the School of Computer Science, Carnegie Mellon University, Pittsburgh, USA. He was co-founder of the DoD funded Software Engineering Institute FFRDC and presently directs several basic research initiatives for defense and intelligence agencies. He was primary architect and remains project director of the Informedia Digital Video Library, one of the initial US Digital Library Initiative projects, and is the principal investigator for the NSF-funded Digital Human Memory Machine and CareMedia projects. The latter applies machine understanding of observed human behaviour patterns to healthcare. He has been instrumental in establishing the US, European and Asian digital library research programs. He holds basic patents for processes related to automated video understanding, is widely published and a contributing editor to the scientific literature and a member of numerous US and foreign government advisory and review panels. He has co-chaired the JCDL conference and served on numerous conference program committees. His current research centers on multimedia information systems, machine learning and intelligent systems, and their application to improving health care.

Tobun Dorbin Ng is a Systems Scientist in School of Computer Science at Carnegie Mellon University. He earned the MS and PhD degrees in management information systems from the University of Arizona in 1993 and 2000, respectively and received his BS in business administration from the University of Arizona in 1990. He has worked on analyzing semantic components – concepts – in large-scale textual and multimedia databases since the beginning of the 1990s. His research consists of extracting concepts using heuristics and natural language processing methods, associating concepts for generating thesaurus-like concept network, exploring concept spaces for contextual query expansion, and clustering concepts for summarization. He is a member of ACM and IEEE. His research interests include digital libraries, intelligent and multimedia information retrieval, semantic extraction and context analysis, semantic interoperability for human computer interaction and information analysis, large-scale knowledge discovery using high performance supercomputers, knowledge management, web-based search engine and user interface development, neural-network computing and collaborative computing.

Selected publications relevant to the project:

- "Towards Automatic Analysis of Social Interaction Patterns in a Nursing Home Environment from Video", Chen, D., Wactlar, H., Yang, J., 6th ACM SIGMM International Workshop on Multimedia Information Retrieval (MIR'04), New York, New York, October, 2004
- "Articulated Motion Modelling for Activity Analysis", Gao, J., Wactlar, H., Hauptmann, A., Collins, R.T., IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'04), Workshop on Articulated and Nonrigid Motion (ANM'04), Washington, DC, vol. 1, pp. 20, July, 2004
- "Automated Analysis of Nursing Home Observations". Wactlar, H., Yang, J., Qi, Y., Yan, R., Gao, J., Hauptmann, A., IEEE Pervasive Computing, vol. 3, no. 2, pp. 15-21, April-June, 2004.
- "Data Analysis for a Multimedia Library". Hauptmann, A., Wactlar, H., Jin, R. Text- and Speech-Triggered Information Access, Renals, S. and Grefenstette, G. (eds), Springer, Berlin, pp. 6-37, 2003.

- "A System of Video Information Capture, Indexing and Retrieval for Interpreting Human Activity" . Wactlar, H., Bharucha, A., Stevens, S., Hauptmann, A., Christel, M., Third International Symposium on Image and Signal Processing and Analysis (ISPA'03), September 18-20, 2003.
- "Informedia at TRECVID 2003: Analyzing and Searching Broadcast News" Video, Hauptmann, A., Ng, D., Baron, R., Chen, M-Y., Christel, M., Duygulu, Huang, C., Lin, W-H., Wactlar, H., Moraveji, N., Papernick, N., Snoek, C.G.M., Tzanetakis, G., Yang, J., Yan, R., Jin, R., Proceedings of (VIDEO) TREC 2003 (Twelfth Text Retrieval Conference), Gaithersburg, MD, November 17-21, 2003
- "Linking Visual and Textual Data on Video", Duygulu, P., Ng, D., Papernick, N., Wactlar, H., Workshop on Multimedia Contents in Digital Libraries, Crete, Greece, June 2-3, 2003
- "Collages as Dynamic Summaries of Mined Video Content for Intelligent Multimedia Knowledge Management" , Ng, T.D., Wactlar, H., Hauptmann, A., Christel, M., AAAI Spring Symposium Series on Intelligent Multimedia Knowledge Management, Palo Alto, CA, March 24-26, 2003
- "Video Retrieval Using Speech and Image Information", Hauptmann, A., Ng, T.D., Jin, R., Electronic Imaging Conf. (EI'03), Storage and Retrieval for Multimedia Databases, Santa Clara, CA, Jan, 2003
- "Collages as Dynamic Summaries for News Video". Christel, M., Ng, D., Wactlar, H., Hauptmann, A. Proceedings of the ACM Multimedia 2002, Juan-les-Pins, France, December 1-6, 2002.
- "Enriching Perspectives in Exploring Cultural Heritage Documentaries Using Informedia Technologies". Ng, T.D., Wactlar, H., 4th Intl Workkshop. on Multimedia Information Retrieval, Juan-les-Pins, France, Dec, 2002.
- "Video Classification and Retrieval with the Informedia Digital Video Library System", Hauptmann, A., Ng, T.D., Baron, R., Lin, W., Chen, M., Derthick, M., Christel, M., Jin, R., Yan, R., Text Retrieval Conference (TREC02), Gaithersburg, MD, November, 2002
- "Multi-Modal Information Retrieval from Broadcast Video Using OCR and Speech Recognition", Hauptmann, A., Ng, T.D., Jin, R., Joint Conference on Digital Libraries (JCDL'02), Portland, Oregon, pp.160-161, July 13-17, 2002
- "Enhanced Perspectives for Historical and Cultural Documentaries Using Informedia Technologies". Wactlar, H., Chen, C. Joint Conference on Digital Libraries (JCDL'02), Portland, Oregon, July, 2002.
- "Digital Video Archives: Managing through Metadata". Wactlar, H., Christel, M., In Building a National Strategy for Digital Preservation: Issues in Digital Media Archiving, Library of Congress, April, 2002.
- "Improving Access to Digital Video Archives through Informedia Technology." Christel, M., Hauptmann, A., Wactlar, H., Journal of the Audio Engineering Society, 49(7/8), July/August 2001.
- "Complementary Video and Audio Analysis for Broadcast News Archives". H. Wactlar, A. Hauptmann, M. Christel, R. Houghton, A. Olligschlaeger, Communications of the ACM, 32(2), pp.42-47, Feb 2000.
- "Lessons Learned from the Creation and Deployment of a Terabyte Digital Video Library" . Wactlar, H., Christel, M., Gong, Y., Hauptmann, A., IEEE Computer 32(2), pp. 66-73, 1999.
- "Informedia Experience-On-Demand: Capturing, Integrating and Communicating Experiences Across People, Time and Space ". Wactlar, H., Christel, M., Hauptmann, A., Gong, Y. ACM Computing Surveys, 31(9), June, 1999.
- "Intelligent Access to Digital Video: The Informedia Project". Wactlar, H., Kanade, T., Smith, M., and Stevens, S., IEEE Computer, 29(5), May 1996.

Stichting Centrum voor Wiskunde en Informatica – CWI

Dick C.A. Bulterman received a Ph.D. in computer science from Brown University (USA) in 1981. His dissertation studied animation languages to model parallel architectures. He received an Sc. M. from Brown in computer science in 1977 and a bachelor's degree in economics from Hope College in 1973. In 1980, he joined the Laboratory for Engineering Man/Machine Systems (LEMS) within Brown's Division of Engineering. In 1984 he was award an IBM young faculty member award and in 1985, he was a visiting professor of engineering at the Technical University of Delft. In 1988 he joined CWI as head of the department of computer systems and telematics. In 1991, he founded CWI's interdisciplinary Multimedia System group which transformed itself into the Multimedia and Human-Computer Interaction theme. In 1997, he won the McKinsey New Ventures '97 prize, the proceeds from which went to form Oratrix Development BV, a CWI-startup company specializing in multimedia authoring and presentation systems. In 2002, he rejoined CWI and started the Convergent Media Infrastructures theme.

Bulterman's research interests revolve around distributed models for supporting the capture, transfer and rendering of time-sensitive multimedia documents over a heterogeneous infrastructure. At present, his group looks at applications in the domains of medical multimedia, home-based media capture and enrichment and interfaces for temporal documents in the transportation sector. The group also develops the Ambulant Open Source SMIL player and incremental authoring

environments. The core problems include the development of document activity specification that can be used as the basis for transportable applications, the development of protocol rules for adaptive communication of multimedia data, and distributed resource allocation algorithms.

Bulterman has held teaching appointments at Brown, the University of Utrecht, Leiden University and the University of Amsterdam. He is a founding member of the W3C Synchronized Multimedia working group and contributed heavily to the development of the SMIL-1, SMIL-2 and SMIL-2.1 specifications. He participates in the Dutch public/private BRICKS consortium as head of the PDC-3 project on models for end-to-end multimedia infrastructures, he is a member of the ITEA *Passepartout* project and is director of the *AMBULANT*-I/-II projects. He was the technical director of the ESPRIT CHAMELEON project and has participated in over a dozen other European and national projects. His group (and company) produced the *CMIF* and *GRiNS* multimedia authoring and presentation formats. He has published 75 articles on networking, computer graphics and multimedia topics. He has contributed chapters to two books and multimedia systems and is (together with Lloyd Rutledge) the author of *SMIL 2.0: Interactive Multimedia for Web and Mobile Devices* (published by Springer Verlag in 2004.)

He is associate editor of ACM *Trans. on Multimedia Communications, Computing and Applications* (TOMCCAP) and associate editor of ACM/Springer *Multimedia Systems Journal.* Bulterman also serves on the editorial board *Multimedia Tools and Applications*. He was program chair of the ACM Multimedia '99, Multimedia Modelling 2001 and WWW-9 conferences. He is an active member of the ACM special interest group on Multimedia and is a member of Sigma Xi, the ACM and the IEEE Computer Society.

Lynda Hardman (http://www.cwi.nl/~lynda/) heads the Multimedia and Human-Computer Interaction group at CWI and is part-time full professor at the Technical University of Eindhoven. She obtained her PhD from the University of Amsterdam in 1998, having graduated in Mathematics and Physics from Glasgow University in 1982. During several years of working in the software industry she was the development manager for Guide - the first hypertext authoring system for personal computers (1986).

She was a member of the W3C working group that developed the first SMIL recommendation. The research projects she currently leads focus on different aspects of the automated generation of hypermedia presentations, with emphasis on aspects of discourse and design and on underlying (Semantic) Web technologies.

She is a member of the editorial board for the New Review of Hypermedia and Multimedia, was co-programme chair for ACM Hypertext in 2003 and has reviewed for the ACM Multimedia, ACM Hypertext and WWW conference series.

Publications relevant to the project:

- Dick C.A. Bulterman and Daniel F. Zucker, *SMIL 2.1 Extended Mobile Profile*, W3C Candidate Recommendation, W3C March 2005.
- Dick C.A. Bulterman, *SMIL 2.1 Layout Module Functional Specification*, W3C Candidate Recommendation, W3C March 2005.
- Dick C.A. Bulterman and Lynda Hardman, *Structured Multimedia Authoring*, ACM Trans. on Multimedia Computing, Communications and Applications, 1(1) 2005.
- Dick C.A. Bulterman, *Is It Time for a Moratorium on Metadata?,* IEEE Multimedia, 11(4) , Pp. 10-17, October - December 2004.
- Jacco van Ossenbruggen, Frank Nack, and Lynda Hardman. *That Obscure Object of Desire: Multimedia Metadata on the Web (Part I)*. In: IEEE Multimedia 11(4) Pp. 38-48, October - December 2004.
- Dick C.A. Bulterman, *Animating Peer-Level Annotations Within Web-Based Multimedia*, Eurographics Multimedia 2004, Nanjing, China, 27-28 October 2004.
- D.C.A. Bulterman, A.J. Jansen, K. Kleanthous, K. Blom, and D. Benden, *Ambulant: A Fast, Multi-Platform Open Source SMIL Player*, ACM Multimedia 2004, New York, Oct 2004.
- Dick C.A. Bulterman and Lloyd Rutledge, *SMIL 2.0: Interactive Multimedia for Web and Mobile Devices*, Springer-Verlag, Heidelberg, May 2004.
- Frank Nack, Amit Manniesing, and Lynda Hardman. *Colour Picking - the Pecking Order of Form and Function*. In: Proceedings of the eleventh ACM International Conference on Multimedia, ACM Press, Berkeley, USA, November 2 - November 8, 2003
- Dick C.A. Bulterman, *Using SMIL to Encode Interactive, Peer-Level Multimedia Annotations*, Proc. of ACM DocumentEngineering 2003, Grenoble, France, November 2003, pp. 32-41.

- Joost Geurts, Stefano Bocconi, Jacco van Ossenbruggen, and Lynda Hardman. *Towards Ontology-driven Discourse: From Semantic Graphs to Multimedia Presentations*. In: Second International Semantic Web Conference (ISWC2003) (pages 597-612), October 20-23, 2003, Sanibel Island, Florida, USA.
- Dick C.A. Bulterman, *SMIL Authoring Systems: The State of the Art*, Proc of SMIL Europe 2003, Paris, February 2003, 47-53.
- Dick C.A. Bulterman, *SMIL 2.0: Overview, Concepts, and Structure*, IEEE Multimedia, 8(4), 2001, pp. 82-88.
- Dick C.A. Bulterman, *Authoring Systems*, in *Readings in Multimedia Computing and Networking* (K. Jeffay and H.J. Zhang, Eds.), Morgan Kaufmann, San Francisco, 2001..
- Dick C.A. Bulterman, Repurposing Broadcast Content for the Web, European Broadcast Union (EBU) Technical Review, 287, June 2001, pp. 1-10.

Fraunhofer – Gesellschaft zur Foerderung der angewandten Forschung e.V. - Institute for Media Communication (FhG)

Dr. Sepideh Chakaveh received her BSc. in Electronics Engineering from Nottigham Trent Polytechnic in 1993 UK, and since completion of her Ph.D. in Space Science & Astrophysics in 1996 from Kent University at Canterbury UK has been employed at the following research establishments:

Postdoctoral Research Assistance: Astrophysics Group, Imperial College, London, UK (9/86-6/89),
Rutherford Appleton Laboratories, OXON, UK (6/89-12/89),
Max Planck Institut Für Kernphysik, Heidelberg, Germany (1/90-1/93).
Software Scientist, University of Göttingen Observatory, Göttingen,Germany (1/93-1/95).
Project Manager, GMD German Institute for Information Technology, St.Augustin, Germany (1/95 to 6/97). Head of the Telematic & Telecommunication Application Group DELTA, German Institute for Information Technology (7/97 to Date).
Membership in UNESCO's Experts commission on Communication & Information Technology.
Head and the initiator of the European Research Consortium for Informatics and Mathematics (ERCIM) E-Learning Working group.

Joachim Kaeber received his diploma in computer science from the Technical University Berlin, Germany, in 1982. He joined GMD in 1984. Since 1995, he is a member of the research staff at Fraunhofer FhG. He has been working on several projects in the area of interactive digital TV in cooperation with industrial partners, the main focus being on design and implementation of playout systems.

Publications relevant to the project

- Chakaveh, Sepideh; Geuer, Olaf; Werning, Stefan; Borggrefe, Sorina; Haeger, Ralf, "Application of web-based visualisations to interactive television: a practical approach", Visual Communications and Image Processing Conference, Lugano, Schwitzerland, 2003
- S. Chakaveh, O. Geuer, S. Werning, S. Burggrefe, R. Haeger, "Application of Web-based Visualisations to Interactive Television: deinewahl02 – A Demonstrator for Interactive Television's Political Programs", Proceedings of SPIE, Vol. 5150, ISSN 0277-786X, S. 1062–1067
- Maad Soha, "The Potential and Pitfall of Interactive TV Technology: An Empirical Study, Television in Transition", MIT –MiT3, 2003
- Geuer, Olaf, "Applications for interactive television", 3rd ERCIM E-Learning Working Group Workshop, Nice, France, 2002
- Maad, Soha, "Universal Access to Multimodal ITV Content: Challenges and Prospects", 7th ERCIM Workshop on "User Interfaces for All", Special Theme: "Universal Access", Paris, France, 2002
- Chakaveh Sepideh, "Pedagogical Applications of the new Technology", UNESCO World Cultural Heritage, Leipzig, Germany 2000

Technical University of Crete (TUC)

Prof. Stavros Christodoulakis is Director of MUSIC/TUC and Professor of the Department of Electronic and Computer Engineering of the Technical University of Crete, Greece. Professor Christodoulakis holds a PhD in Computer Science from the Department of Computer Science of the University of Toronto. Professor Christodoulakis has been professor in the Department of Computer Science of the Universities of Toronto and Waterloo, Associate Chairman and Chairman of Graduate Studies in the University of Waterloo, Member of the Board Directors of the Institute of Computer Research in the University of Waterloo, as well as the Institute of Speech of Greek Ministry Research and Technology, and the Lambrakis Research Foundation. He has also been Chairman of the Department of Electronic and Computer Engineering of the Technical University of Crete, Member of the Senate, as well as Director of

the Laboratory of Telecommunications and the Laboratory of Software Engineering in the Technical University of Crete. Professor Stavros Christodoulakis is member of the Scientific Board of the DELOS II Network of Excellence on Digital Libraries, coordinates the DELOS II Cluster "Audio-visual and non-traditional objects" that focuses on metadata capture for audio-visual content, universal access and interaction with audio-visual libraries, together with the management of audio-visual content in digital libraries.

Prof. Christodoulakis has been a member of the Editorial Board of several International Scientific Journals, Program Committee Chairman, Area Program Committee Chairman, European Program Committee Chairman, and Program Committee Member in many of the most important International Conferences in the areas of Multimedia Systems, Databases, Information Retrieval, and Office Information Systems. He has also been reviewer for the top quality Journals in the areas of databases and information systems, reviewer for the NSF of USA, of NSERC of Canada, and of the European Community ESPRIT Projects. He is the author of many scientific articles in journals, conferences and books in the areas of databases, information systems, multimedia systems, document management systems, and system performance. He has also been invited and presented Keynote Speeches and Tutorials in many international conferences.

Relevant publications

- C. Tsinaraki, E. Fatourou, S. Christodoulakis, "An Ontology-Driven Framework for the Management of Semantic Metadata describing Audiovisual Information", in Proceedings of CAiSE 2003
- C. Tsinaraki, P. Polydoros, F. Kazasis, S. Christodoulakis, "Ontology-based Semantic Indexing for MPEG-7 and TV-Anytime Audiovisual Content", Special issue of Multimedia Tools and Applications Journal on Video Segmentation for Semantic Annotation and Transcoding, 2004
- C. Tsinaraki, P. Polydoros, S. Christodoulakis, "Integration of OWL ontologies in MPEG-7 and TVAnytime compliant Semantic Indexing", in Proceedings of CAiSE 2004
- C. Tsinaraki, P. Polydoros, S. Christodoulakis, "Interoperability support for Ontology-based Video Retrieval Applications", in Proceedings of CIVR 2004, selected for publication in the CIVR 2004 special issue of the IEE VIS Journal

Technische Universität Wien (TUWIEN)

Andreas Rauber is Associate Professor at the Department of Software Technology and Interactive Systems (ifs) at the Vienna University of Technology (TU-Wien). He furthermore is head of the iSpaces research group at the eCommerce Competence Center (ec3).

He received his MSc and PhD in Computer Science from the Vienna University of Technology in 1997 and 2000, respectively. In 2001 he joined the National Research Council of Italy (CNR) in Pisa as an ERCIM Research Fellow, followed by an ERCIM Research position at the French National Institute for Research in Computer Science and Control (INRIA), at Rocquencourt, France, in 2002.

In 1998 he received the OeGAI Award of the Austrian Society for Artificial Intelligence (OeGAI), and the Cor-Baayen Award of the European Research Consortium for Informatics and Mathematics (ERCIM) in 2002. His team won the Rhythm Classification Contest within the Audio Description Contest of the International Conference on Music Information Retrieval (ISMIR). He has published numerous papers in refereed journals and international conferences and served as PC member and reviewer for several major journals, conferences and workshops. He is a member of the Association for Computing Machinery (ACM), The Institute of Electrical and Electronics Engineers (IEEE), the Austrian Society for Artificial Intelligence (OeGAI). He serves on the board of the IEEE Technical Committee on Digital Libraries (TCDL), and is a member of the EU FP6 DELOS Network of Excellence on Digital Libraries as well as the EU FP6 MUSCLE Network of Excellence on Multimedia Understanding through Semantics, Computation and Learning.

He is actively involved in several research projects in the field of Information Retrieval, focusing on the organization and exploration of large information spaces.

His research interests cover the broad scope of digital libraries, including specifically text and music information retrieval and organization, information visualization, as well as data analysis and neural computation.

- S. Leitich, A. Rauber. Information Retrieval in Digital Libraries of Music. Proceedings of the 6th Russian Conference on Digital Libraries (RCDL 2004), pp.207 215, September 29 - October 1 2004, Pushchino, Russia.
- Rauber. Creation and Exploration of Musical Information Spaces. Proceedings of the International Conference on Digital Libraries (ICDL04), pp. 741-748, February 24-27, New Delhi, India.
- Rauber, E. Pampalk, D. Merkl: The SOM-enhanced JukeBox: Organization and Visualization of Music Collections based on Perceptual Models. In: Journal of New Music Research (JNMR), 32(2):193-210, Swets and Zeitlinger, June 2003.
- Rauber, D. Merkl: Text Mining in the SOMLib Digital Library System: The Representation of Topics and Genres. In: Applied Intelligence, Vol. 18, No. 3, pp. 271-293, Kluwer, May/June 2003.

- E. Pampalk, A. Rauber, D. Merkl: Content-based Organization and Visualization of Music Archives. In: Proceedings of ACM Multimedia 2002, pp. 570-579, December 1-6, 2002, Juan-les-Pins, France.
- Rauber, D. Merkl, and M. Dittenbach: The Growing Hierarchical Self-Organizing Map: Exploratory Analysis of High-Dimensional Data. In: IEEE Transactions on Neural Networks, Vol. 13, No 6, pp. 1331-1341, November 2002. IEEE.
- Rauber, E. Pampalk, D. Merkl: Using Psycho-Acoustic Models and Self-Organizing Maps to Create a Hierarchical Structuring of Music by Musical Styles. In: Proceedings of the 3rd International Conference on Music Information Retrieval (ISMIR 2002), pp.71-80, October 13-17, 2002, Paris, France.
- Rauber, E. Pampalk, D. Merkl: Content-based Music Indexing and Organization. In: Proceedings of the 25. Annual International ACM SIGIR

University of Amsterdam (UVA)

Arnold W.M. Smeulders graduated from Technical University of Delft in physics in 1977 (M.Sc.) and in 1982 from Leyden University in medicine (Ph.D.) on the topic of visual pattern analysis. In 1994, he became full professor in multimedia information analysis at the University of Amsterdam. In 1996, he was treasurer of the Faculty and from 1997 until 2001 he served as the director of the Informatics Institute at the University of Amsterdam. Currently, he is scientific director of the Intelligent Systems Lab, of 60 staff members, the MultimediaN public-private partnership of 30 institutions and companies with a budget of 32 MEuro and 100 participants, and of the ASCI national research school.

He has an interest in cognitive vision, content-based image retrieval, the picture-language question as well as in systems for the analysis of video. He has graduated 28 PhD-students. Currently, he heads the ISIS research group of 12 Ph.D students and 10 staff and support members concentrating on theory, practice and implementation of multimedia information analysis including image databases and computer vision. The group has an extensive record in co-operations with Dutch institutions and industry in the area of multimedia and video analysis. He has written 250 papers in refereed journals and conferences.

He received a Fulbright grant at Yale University in 1987, and a visiting professorship at the City University Hong Kong in 1996, and again at Tsukuba Japan in 1998. In 2000, he was elected fellow of International Association of Pattern Recognition. He was associated editor of IEEE transactions PAMI. Currently he is associated editor of the Int. Journal for Computer Vision as well as the IEEE trans Multimedia.

He is a member of the steering committee of the Visual Information Systems and IEEE's International Conference on Multimedia and Expo series. He participates in de EUVision, DELOS and MUSCLE networks of excellence. He was keynote speaker and chairman of the program committee of conferences including the IEEE Multimedia conference in Florence in 1999, ICIP 2000, CVPR in 2001 and CIVR in 2004 in Dublin. Currently he is general chair of ICME2005 in Amsterdam.

Nicu Sebe is an assistant professor in the Faculty of Science, University of Amsterdam, The Netherlands, where he is doing research in the areas of multimedia information retrieval and human-computer interaction in computer vision applications. He is the author of the book Robust Computer Vision—Theory and Applications (Kluwer, April 2003) and of the upcoming book Machine Learning in Computer Vision. He was a guest editor of a CVIU special issue on video retrieval and summarization (December 2003) and was the co-chair of ACM Multimedia Information Retrieval Workshops, MIR'03 & MIR'04 (in conjunction with ACM Multimedia conferences). He also was the co-chair of the first Human Computer Interaction Workshop, HCI '04 (in conjunction with ECCV 2004) and is the co-chair of the upcoming IEEE Workshop on Human computer Interaction Workshop (in conjunction with ICCV 2005). He is the guest editor of two special issues on multimedia information retrieval and human computer interaction in ACM Multimedia Systems journal and Image and Vision Computing Journal. He was the technical program chair for the International Conference on Image and Video Retrieval, CIVR 2003. He was a visiting researcher in the Beckman Institute, University of Illinois at Urbana-Champaign (2002) and was a research fellow of the British Telecomm in Ipswich (2003). He has published more than 50 technical papers in the areas of computer vision, content-based retrieval, pattern recognition, and human-computer interaction and has served on the program committee of several conferences in these areas. He is a member of the IEEE and the ACM.

Publications relevant to the project:

- N. Sebe, I. Cohen, T.S. Huang, Multimodal Emotion Recognition, in Handbook of Pattern Recognition and Computer Vision (chapter 4.1), C.H. Chen and P.S.P. Wang, Eds., World Scientific, ISBN 981-256-105-6, January 2005
- N. Sebe, I. Cohen, A. Garg, T.S. Huang, Machine Learning in Computer Vision, Springer, to appear in March 2005.

- Cohen, N. Sebe, F.G. Cozman, M.C. Cirelo, T.S. Huang, Semi-supervised Learning of Classifiers: Theory and Algorithms and Their Applications to Human-Computer Interaction, IEEE Trans. on PAMI, Vol. 26, No. 12, pp. 1553-1567, December 2004.
- D.P. Huijsmans, N. Sebe, How to Complete Performance Graphs in Content-Based Image Retrieval: Add Generality and Normalize Scope, IEEE Trans. on PAMI, Vol. 27, No. 2, pp. 245-251, February 2005.
- C.G.M. Snoek, M. Worring, Multimodal Video Indexing: A Review of the State-of-the-art, Multimedia Tools and Applications, 25(1):5-35, January 2005.
- C.G.M. Snoek and M. Worring, Multimedia Event based Video Indexing using Time Intervals, IEEE Trans. on Multimedia, 2005 (in press).
- J. van de Weijer, Th. Gevers, J.-M. Geusebroek, Edge and Corner Detection by Photometric Quasi-Invariants, IEEE Trans. on PAMI, Vol. 27, No. 4, pp. 625-630, April 2005.
- Th. Gevers, Robust Segmentation and Tracking of Colored Objects in Video, IEEE Trans. on Circuits and Systems for Video Technology, pp. 776- 781, June, 2004.
- Th. Gevers and H.M.G. Stokman, Robust Histogram Construction from Color Invariants for Object Recognition, IEEE Trans. on PAMI, 26(1), pp. 113-118, January 2004.
- Th. Gevers and H. M. G. Stokman, Classification of Color Edges in Video into Shadow-Geometry, Highlight, or Material Transitions, IEEE Trans. on Multimedia, 5(2), 2003.
- A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content based image retrieval at the end of the early years. IEEE Trans. on PAMI, 22(12):1349-1380, 2000.

Appendix E – Letters of support

DELOS
NETWORK OF EXCELLENCE ON DIGITAL LIBRARIES

Pisa, 17 March, 2005

Dr. Costantino Thanos
ISTI-CNR

TO WHOM IT MAY CONCERN

LETTER OF SUPPORT

With this letter I should like to express the firm interest and commitment of the DELOS Network of Excellence in activating a cooperation with the FACS (Fully Automated multimedia Content management and personalised Synthesis) project if the project proposal is successful. This is a well conceived proposal which addresses issues that are currently of great significance and of high interest to the DELOS community.

In DELOS we have activated clusters that address several of the research topics explored in FACS, such as Information Access and Personalization, Audio-visual and non traditional Objects in Digital Libraries, and Knowledge Extraction and Semantic Interoperability.

We would thus be happy to activate a fruitful cooperation with the FACS project and to support FACS in the dissemination of project results through our Network.

Yours Sincerely,

Costantino Thanos
Scientific Coordinator
DELOS Network of Excellence

DELOS
Network of Excellence on
Digital Libraries
www.delos.info

ISTI-CNR
Via G. Moruzzi 1
56124 Pisa, Italy

Phone: +39 050 315 2910
Fax: +39 050 315 3464
E-Mail: costantino.thanos@isti.cnr.it

Centrum voor Wiskunde en Informatica

Our reference EP/PNA4
Date March 16th, 2005
Telephone +31-20-592 4225
E—mail E.Pauwels@cwi.nl

To whom it may concern.

LETTER OF SUPPORT

Dear Madam, Sir,

With this letter I should like to express the firm interest and commitment of the MUSCLE Network of Excellence in activating cooperation with the FACS (Fully Automated multimedia Content management and personalized Synthesis) project if the project proposal is successful. Several of the research topics addressed in the FACS project are of key interest to our Network, and it would be our intention to establish a fruitful cooperation through joint meetings and workshops, and possibly exchange or joint use of software and test-data.

Of particular relevance to MUSCLE are the planned activities on automated semantic annotation of sound and video and machine learning on multimedia data. We would also be happy to support FACS in the dissemination of project results through the MUSCLE network.

Do not hesitate to contact me if you require further information.

Sincerely,

Dr. Eric J. Pauwels
Scientific Coordinator
MUSCLE Network of Excellence

CENTRUM VOOR WISKUNDE EN INFORMATICA
KRUISLAAN 413 - 1098 SJ AMSTERDAM
POSTBUS 94079 - 1090 GB AMSTERDAM

Addresses
Kruislaan 413 1098 SJ Amsterdam
P.O. Box 94079 1090 GB Amsterdam
info@cwi.nl www.cwi.nl
Tel (020) 592 9333 Fax (020) 592 4199

Bank accounts
ABN-AMRO 43.60.53.705
Postbank 462890
Centrum voor Wiskunde en Informatica

Registration Numbers
41198731 KvK Amsterdam
Alg. Voorwaarden nr. 3851 KvK Amsterdam
BTW nr. NL 002953390B01