



# 6-monthly Activity Report 04/2007–09/2007

Deliverable T18 / Public Part

MediaCampaign Identifier: MC-T18-JRS-02-6monthlyManRepPubP02.doc

Author(s) and company: Herwig Rehatschek (JRS), Helen Hasenauer (JRS),  
Herwig Zeiner (JRS), Mark Guinney (NMR-UK), Julien  
Nioche (USFD), Borislav Popov (ONTO), Walter  
Plaschzug (HSA), Marco Masetti (SOF); Roeland  
Ordelman (UT), David van Leeuwen (TNO)

Work package/task: WP 1 / T1.1

Document status: Final

Confidentiality: Public

Keywords: project highlights, project progress, summary of WP  
efforts, resource consumption

Abstract: This document summarises the overall project results  
and progress for the period 04/2007 (PM 013) – 09/2007  
(PM 18).

## DOCUMENT HISTORY

Version	Date	Reason of change
1	2007-10-25	document created
2	2007-11-06	Input from WP leaders incorporated



## Table of Contents

---

<b>1 Executive Summary .....</b>	<b>1</b>
<b>2 Introduction .....</b>	<b>3</b>
2.1 Purpose of this document.....	3
2.2 Scope of this document.....	3
2.3 Related Documents .....	3
<b>3 WP1: Project Management .....</b>	<b>4</b>
3.1 Management Issues .....	4
3.2 Main Management Activities.....	5
3.3 Project Meetings.....	5
3.4 Quality assurance plan and procedures.....	7
3.4 Corrective Actions and conflict resolution .....	7
<b>4 Project Progress and Status .....</b>	<b>8</b>
4.1 WP2: Requirements, Data Collection and Evaluation.....	8
4.2 WP3: Campaign Modelling .....	10
4.3 WP4: Media Presence and Campaign Analysis.....	11
4.3.1 <i>Cross Media Advertisement Detection</i> .....	11
4.3.2 <i>Audio analysis</i> .....	11
4.3.3 <i>Visual analysis</i> .....	12
4.3.4 <i>Textual analysis</i> .....	12
4.4 WP5: Knowledge Fusion and Campaign Discovery.....	13
4.4.1 <i>Creative detection</i> .....	13
4.4.2 <i>Knowledge Fusion</i> .....	13
4.4.3 <i>Campaign Discovery</i> .....	13
4.4.4 <i>Evaluation</i> .....	13
4.5 WP6: Component Frameworks .....	14
4.6 WP7: Exploitation and Dissemination .....	15
4.6.1 <i>Swedish potential customer in Stockholm</i> .....	15
4.6.2 <i>Demo system</i> .....	15
4.6.3 <i>Dissemination Activities Report (D7.2.3)</i> .....	16
4.6.4 <i>Exploitation &amp; Use Plan (D7.1.2)</i> .....	16
4.6.5 <i>Scientific publications</i> .....	16
4.6.6 <i>Public web-site</i> .....	16
<b>5 Glossary .....</b>	<b>17</b>

# 1 Executive Summary

---

MediaCampaign targets Action Line IST-2004-2.4.7 "Semantic-based knowledge and content systems" of the IST work programme 2004 (IST FP6-027413). MediaCampaign's scope is on discovering, inter-relating and navigating cross-media campaign knowledge. The project's main goal is to automate to a large degree the detection and tracking of media campaigns on television, the Internet and in the press.

The main event during the last reporting period was the 1<sup>st</sup> EC review of the project. The main outcome of the review was that MediaCampaign has made significant progresses in accordance with the initial plan and had good technical results so far. Dissemination is proceeding relatively well with a completed project brochure, a first version of the website and six scientific publications. The exploitation planning has begun with the production of a market analysis (D7.1.1). The project appears to be well-coordinated and minor changes in budgets and staff were justified<sup>1</sup>. The reviewers gave five recommendations, which will be closely followed and taken into consideration by the consortium within the next year.

Before the review meeting an integration meeting took place, where the first MediaCampaign prototype was set-up. For this prototype all analysis modules were available in a first prototype version and with all necessary communication interfaces implemented. This prototype was fully integrated and supported the full workflow for press and TV. All modules with basic functionality were available. The software was successfully presented during the 1<sup>st</sup> EC review in two steps: first the components and the scientific advances made were presented by the individual partners, second a workflow demonstration was given showing the workflow from the acquisition until the delivery system. This workflow was demonstrated for the media TV and press. The first MediaCampaign prototype was documented within D6.4.1 which was submitted to the EC.

The second main activity within the last reporting period was the first evaluation of the MediaCampaign system. All results have been manifested within D2.3 which was submitted in-time to the EC. Generally positive results have come back from all of the component tests done. Most of the system integration tests performed have been successful within the scope of the requirements at this stage. Tests that have not, and require further attention have been identified and will be dealt with in the next year. Some tests, however, will need to be further measured as they will only be deemed a success if they provide a good enough speed of delivery as well as a high quality of creative and campaign detection. All in all the results of the tests have provided a good foundation for the next stages of testing.

In terms of exploitation an important contact to a Swedish company could be established which is highly interested in solutions for media monitoring. There was already a first meeting with the company where a demo system was demonstrated. This company is also highly interested in other fields of media-analysis (brand detection) and therefore a potential customer for several MC components. Following the presentation of the demo system in summer 2007, there was some more detailed discussion about Internet workflow and logo recognition. Requirements for Internet workflow have been discussed and a test-installation for logo-recognition at the customer's site has been prepared.

Additionally a demo system for the press-workflow that is set-up at HSA has been used to do a careful evaluation and find out problematic cases. Based on that isolated samples the failure rate should be reduced in the next version. Some preparations with NMR-UK have been done, in order to setup the press workflow at a real user-site. In connection with dissemination JRS presented a MediaCampaign poster at the K-Space summer school (SSMS'07) in Glasgow, UK. There are two scientific publications in this period, namely Marijn Huijbregts, Roeland Ordelman and Franciska de Jong. Annotation of Heterogeneous Multimedia Content Using Automatic Speech Recognition. SAMT, December 5-7 2007, Genova, Italy and Alps Adria Acoustics Association Conference, Graz, Austria: Presentation of the publication "Speech Music Discrimination in Mixed Audio Content" by B. Rettenbacher, M.Fellner from JRS. Furthermore there was made a major update of the public website

---

<sup>1</sup> some quotations from official EC review report.

with a detailed description of the MediaCampaign prototype - as recommended by the reviewers. The prototype description is accessible by the menu on the left side, following the link "MediaCampaign Prototype".

## 2 Introduction

---

### 2.1 Purpose of this document

---

This document gives an overview over the project achievements within the reporting period, which last from 04/2007 (PM 13) until 09/2007 (PM 18). The Management part is not included in the public version of the document.

### 2.2 Scope of this document

---

This document contains the activities and milestones reached within the third six months of the MediaCampaign project.

### 2.3 Related Documents

---

In order to get a more detailed insight in the project progress this document should be read together with the following already submitted major deliverables:

- MC-D641-SOF-08-FirstPrototype
- MC-D23-NMRUK-10-MidtermEval

Furthermore in order to get a general idea of the MediaCampaign project the public website <http://www.media-campaign.eu> shall be visited, where latest dissemination activities, public deliverables, the project objectives and the technical achievements can be found.

### 3 WP1: Project Management

WP leader	JRS
Objectives	<ul style="list-style-type: none"> <li>Lead the MediaCampaign project to technical, organisational and financial success.</li> <li>To set up communication and project management structures.</li> <li>Clear definition of tasks along with responsibilities and a work distribution that, on the one hand, avoids excessive interactions between the participants but, on the other hand, ensures that all the necessary communications between the participants will be established.</li> <li>Ensure the overall project quality. A Quality Assurance (QA) Plan will be applied to all MediaCampaign activities and deliverables. QA will ensure mediation of errors as early in the life cycle as possible. The QA plan consists of planned and systematic activities to determine and ensure achievement of MediaCampaign quality objectives.</li> </ul>
Milestones and expected results	<ul style="list-style-type: none"> <li>Quality Assurance Plan</li> <li>Internal communication structure</li> </ul> <p>Organization of kick-off meeting, project policy board meetings (at least every 6 months), work package meetings (on demand) and EC reviews (every 12 months)</p>
Deliverables	<p><b>T6, T18, T30</b>      Half year management report and project progress summary</p> <p><b>T12, T24</b>         Annual reports and financial statements</p> <p><b>D1.2</b>                Quality Assurance Plan</p>

Main objective of WP1 "project management" is to lead MediaCampaign to technical, organisational and financial success. This goal is ensured by using of available communication and project management structures and a clear definition of tasks along with responsibilities. Furthermore by a work distribution which, on the one hand, avoids excessive interactions between the participants but, on the other hand, ensures that all the necessary communications between the participants will be established.

To ensure the overall project quality the Quality Assurance (QA) Plan (D1.2) is used. QA will ensure detection of errors as early in the life cycle as possible. The QA plan consists of planned and systematic activities to determine and ensure achievement of MediaCampaign quality objectives.

#### 3.1 Management Issues

After a successful and efficient first year the project is well on track and all deliverables have been submitted in-time to the EC. Several policy board and integration meetings have been organized during the first 18 months which consequently contributed to the successful completion of the first milestone "Overall specification" (PM 7 to pave the road to reach the second milestone

"MediaCampaign prototype" (PM 14). The MediaCampaign prototype was shown at the 1<sup>st</sup> EC review held on 6 June 2007 in Graz.

### 3.2 Main Management Activities

---

Main activities of consortium management in the last reporting period included the preparation of the

- The main activity within the last reporting period was the preparation of the annual report consisting of the activity report, the management report, the public version of the activity report and all the financial statements. JRS also offered a pre-check service to the partners in order to help them with preparation and in order to ensure the highest possible quality in the delivery of the FS. Furthermore JRS provided numerous excel sheets in order to assist partners with the preparation of the FS. The annual report 2007 was submitted completely and in time mid of May to the EC.
- Consortium Organisation (Management Constellation, Project Co-ordination and Management, Project Co-ordination and Management Board, Work Package Managers, Conflict resolution, Quality Assurance & Corrective Actions)
- Project Monitoring/Deliverables (Monthly Activity Reports, etc.). This activity includes the preparation of the Bi-monthly, Six-monthly Activity Report for period 1/4/2007-30/9/2007, monitoring project progress, setting up the agenda, and preparation for the integration meetings (25/04-26/04/2007) hosted by USFD, and 04/06/2007 hosted by JRS in Graz (parte of the 1<sup>st</sup> EC review)
- From January 2008 until June 2008 Judith Kessens will we the main contact person at TNO. David van Leeuwen will be on a sabbatical in Berkeley.
- Contractual Framework, Payments/Cost Statements

### 3.3 Project Meetings

---

Several project meetings have been held within the reporting period Apr. 2007 – Sept. 2007. The Project Policy Board meeting (PBM) is an official plenary meeting where administrative issues and topics of the project policy nature are discussed and where project decisions take place. More details concerning the organisation and the decision structure within the project MediaCampaign can be found in the Description of Work (Annex I of EC-Contract) or in Deliverable D1.2 QA Plan and Objective Tree. Within the reported period two policy board meetings were held.

Additionally, WP meetings and workshops took place. Minutes for PBM were circulated latest 2 weeks after each project meeting and were confirmed by all MediaCampaign partners. The final versions of the minutes are available via the MediaCampaign internal document store and can be downloaded by all partners.

Meeting	Date	Location	Main Topics (Participants)	Results
Integration-PBM	25 – 26 April 2007 integration meeting (technical partners) 27 April: PBM (all partners)	Sheffield (UK)	Goal: Review preparation and Integration  Participants: (all partners)	To set-up the first MediaCampaign prototype → last integration & tests before the EC review and to agree on a test data set from the query data (May 2006) of the common test data set which demonstrates the functionality of the MC system in the best way (to be shown at the EC review).
EC-Review	4 – 6 June 2007 4 June Integration 5 June Rehearsal (same agenda as for review!) 6 June EC Review	Graz (AT)	EC-Review  Participants: (all partners, reviewers: Mr. Nuno Correia and Mr. Dave Snowden)	Review Result: Despite some setbacks the consortium has a functional prototype system that implements first versions of all the system components for television and print media. There have been some difficulties in obtaining a suitable speech recognition system for German but English and Dutch are supported.  The reviewers have no hesitation in recommending that the project continue.  The project should proceed according to its current plan.
Policy Board Meeting	20. – 22. September 2007	Genoa (IT)	Goal: Internet workflow, test material, evaluation results, future developments  Participants: (all partners)	Set up of Internet workflow, discussion and decision on new test material, discussion on set-up of system at NMR-UK, discussion of evaluation results, set-up of development plan for MC Integrated Prototype & Integration.

**Planned Meetings:**

Meeting	Date	Location	Subjects (Participants)
Integration Meeting	29 – 30 November 2007	Soesterberg (NL)	Goal: set-up of prototype, in specific Internet workflow Participants: all technical partners
Integration PBM	6 – 8 February 2008 3 days (2 days integration, 1 day PBM on 8 Feb 2008)	Bracknell (UK)	Set-up of MediaCampaign prototype at NMR-UK Preparation of annual report

### 2.3 Quality assurance plan and procedures

---

This section gives a short overview on how the project performed in terms of reaching its objectives and quality of management. This section does not assess technical performance. This is addressed in the corresponding Deliverables D2.2 test scenario plan and especially in D2.3 Mid-term evaluation report.

Within D1.2 the procedures which will be undertaken during the project in order to guarantee delivery of high quality project results and Deliverables that are inline with the evaluation criteria of the IST programme were defined. For this purpose an internal review procedure was manifested within the project and agreed with all partners. Next to this the Deliverable gives an overview on internal communication and reporting tools used and on financial issues.

The document provides all project partners with a summary of the most important project procedures (project monitoring, reporting tools, financial management, internal communication structures, etc.). Additionally this document comprises partner and contact lists as well as documentation and communication standards in order to enable quick and efficient communication within the project consortium.

Resources within the project were effectively monitored by the co-ordinator (by detailed bi-monthly reports) and appropriately used according to the reached objectives.

### 3.4 Corrective Actions and conflict resolution

---

No corrective actions had to be undertaken in the last reporting period.

## 4 Project Progress and Status

### 4.1 WP2: Requirements, Data Collection and Evaluation

WP leader	NMR-UK
Objectives	<ul style="list-style-type: none"> <li>• Production of test scenario document and check list</li> <li>• Define the guidelines for the data collection and for the training of the system</li> <li>• Provision of relevant content with which the system can be tested off-line in as close-to-reality conditions as possible</li> <li>• Ongoing feedback from testing throughout the development stage</li> <li>• Define and characterise operational environment and conditions</li> <li>• Analyze the Advertising Campaign Domain, provide hints for heuristics for campaign detection</li> <li>• Define a methodology for the metrics and the evaluation</li> <li>• Produce URD containing the above information</li> </ul>
Milestones and expected results	<ul style="list-style-type: none"> <li>• User requirements, Advertising Campaign Domain Analysis &amp; Draft of Training and Test Data (URD)</li> <li>• Definition of user groups and use cases</li> <li>• Selection of test material for testing MediaCampaign system</li> <li>• Test scenario plan – to be defined for testing MediaCampaign Prototype</li> <li>• Mid-term evaluation report MediaCampaign Prototype</li> <li>• Final Evaluation Report</li> </ul>
Deliverables	<b>D2.3</b> Mid-Term Evaluation Report

The major focus of WP2 in the last reporting period laid on the evaluation of the first MediaCampaign prototype, which was successfully demonstrated during the 1<sup>st</sup> EC project review in Graz. Hence working concentrated on Deliverable D2.3 "Mid-term evaluation report". It was successfully completed by using evaluation results from analysis modules and software components. The document provided the reader with details of the various component and system integration tests for the MediaCampaign project. The population of content for D2.3 was acquired from each of the partners responsible for their respective components and system tests. It also provided the results from the tests done which gave an idea of how successful the each of the tests were. If any tests were found to be unsuccessful, or run manually with the view to be later run automatically, then follow up details were provided with the partners intentions for the future.

Providing test data: Various data was supplied to various partners for testing in the different project modules. NMR had to provide the ground truth for the query and historic data which the partners

would be testing with. They could then use this detail to draw a comparison with the automated system.

NMRUK have also had to be more definitive about what a campaign is. A description of its definition, along with examples, was provided to the partners and is manifested within D2.3.

Providing feedback of NMR to user requirements, software modules, and strategic design decisions:  
As recommended during the first EC project review the definition of campaigns – as given within D2.1 - was revised according to results from NMR-UK internal discussions and new technical possibilities which evolved by availability of the first MediaCampaign prototype. Next to the campaign definition also the creative definition was revised and is manifested now in this document.

Component Tests were applied to each of the following components: audio segmentation, jingle recognition, video and image OCR, logo recognition, image/ video fingerprinting, text analysis and creative detector. For each test information was provided on the definition of the component and what the test is trying to achieve, the scope of the test performed, whether the test has any dependencies, which of the three media types the component will be applied to, whether the quality of the material has a bearing on the test performed, a performance test based on the hardware used, whether licences are required to use the component, consistency test results, stability test results, stress test results, and can optimisation be applied to the component. In addition NMR has done some individual testing and reporting of the AdClipper and AdComparer applications and provides feedback the respective partner – in this case feedback to HSA.

The ground truth was manually acquired from which the results can be compared. Results are provided using precision and recall, so we have a common measurement and a possibility to compare between the components. Follow up actions are attributed to components where optimisation can be explored or implemented, and where some tests may have failed and therefore need to be addressed and re-tested. The evaluation summary concludes each component. It holds a description, the results, and follow up actions.

Within System integration tests individual software modules were combined and tested as a group, and then all together. Each test provides on a defined input an expected output. A description of what is input and what can be expected from the output is shown. What the output reveals determines whether the test is successful or not. The medium, by which the information is moved, such as web services, is also mentioned. Stress tests were performed where applicable.

Two presentations were done at the Policy Board Meeting in Sheffield. The first was a UK Market Analysis which included media capture competitors of NMRUK and the media monitoring market world wide. It provided information about how the work done by each of the partners and MediaCampaign as a whole will benefit NMRUK. The second presentation was regarding work done on the AdClipper and AdComparer applications. It included the types of testing done by the operations team, and feedback to HSA about how we would like the functionality to work differently for optimising the operators working environment.

At the review meeting in Graz two presentations to the partners and EC board were done. The first was 'The User Requirements, and Selection of Test Content. This presentation included the scope of the project, what information NMR currently has with respect to its current work flow, and the speed with which it is delivered. What we would like from the MediaCampaign project when it is complete. It also provided scope to the data that we are currently using to test the system. The second presentation was 'Test Data & Self Assessment Plan'. This presentation showed the various test that have to be performed within each component, and the tests between each component. It also established a uniform measurement from which these tests are to show the results so that effective comparisons can be drawn. Tests were also established for user acceptability and cost efficiency. Details were provided about current rates experienced within NMR as guidance to what we expect from the finished project.

Four presentations were done by NMRUK at the next policy board meeting in Genoa. These included: a.) The main results of the conclusions from D2.3. b.) The definition of a campaign with examples. c.) Statistics for the addition of new sectors Finance and Telecom to IT and Automotive. The current Internet workflow within NMR, how this differs from the press and TV workflow and what NMR would like to get from the MediaCampaign project regarding internet creatives.

## 4.2 WP3: Campaign Modelling

WP leader	ONTO
Objectives	<ul style="list-style-type: none"> <li>• Design of a media presence and campaign ontology (MEPCO) allowing for representation, interlinking, and consolidation of data across different media.</li> <li>• Develop specific media presence models for press, TV, and Internet.</li> <li>• Enable semantic representation and mapping of brand and product related data (catalogues, feature-related terminology, etc.). MEPCO will allow for (i) extension with domain ontologies specific to particular markets and (ii) cross-linking artefacts (e.g. products) between those.</li> <li>• Formally model the domains and markets described in WP2.</li> </ul>
Milestones and expected results	<ul style="list-style-type: none"> <li>• (I) overall specification</li> <li>• design of MEPCO V1 as major input for (II) MediaCampaign prototype</li> <li>• design of the final MEPCO version as major input for the MediaCampaign integrated prototype</li> <li>• public available MEPCO ontology</li> </ul>
Deliverables	<ul style="list-style-type: none"> <li>• No deliverables in the reporting period!</li> </ul>

A central focus in WP3 of this working period has been the incremental update of MEdia Presence and Campaign Ontology – MEPCO. It has been designed according to the analysis of the domain and existing terminology and models. MEPCO has been developed as an extension of the basic upper level ontology PROTON<sup>2</sup> reducing the overall effort needed for a domain ontology by reusing the existing entity classes and properties defined there.

Major activities in WP 3 within the last reporting period were:

- Adjustments of the modelling of instances of creatives, as a result of the prototype-related testing and the input from KFCD module. This included revising the model provided as a template to be used through the web service interface for campaign discovery.
- Revision of MEPCO as a result of testing the first integrated prototype and the interaction between KFCD, OMS and Delivery System.
- Ongoing work on improving the modelling of campaigns and creatives in MEPCO provoked by observations in the prototype and the demands of the campaign discovery.
- Changes to the modelling provided as template for passing creatives to the storage.
- Revision of other modelling issues appearing in the prototype, like: why media company is not subclass of company
  - xsd:string stays as a class
  - hasAdvertiser vs advertiserOfCreative - both properties appear for creatives

<sup>2</sup> <http://proton.semanticweb.org>

- Changes related to incorrect citations of classes and properties in instance data and provided templates
- USFD experiments with modelling approaches in MEPCO related to accommodating more metadata in the ontology

### 4.3 WP4: Media Presence and Campaign Analysis

WP leader	JRS
Objectives	<ul style="list-style-type: none"> <li>• Implementation of a cross-media (Press, Internet, TV) advertisement detector</li> <li>• Development of an audio analysis module featuring speech transcripts (3 languages), jingle recognition and word spotting</li> <li>• Implementation of visual analysis module with new "subimage &amp; subtext" approach and brand detection</li> <li>• Development of text analysis modules for the analysis of text from ASR, Internet and Press in order to spot new campaigns</li> </ul>
Milestones and expected results	<ul style="list-style-type: none"> <li>• (II) MediaCampaign prototype software - Subsystem</li> <li>• (III) MediaCampaign integrated prototype - Subsystem</li> </ul>
Deliverables	<ul style="list-style-type: none"> <li>• D4.2 Draft of Semantic Analysis Components</li> </ul>

Within this work package one main task is the development of several analysis modules (part of the MediaCampaign analysis framework) having as a result the detection of new ads within a single media and a single language. Dependent on the media type (e.g. press, TV, Internet (in planning state)) the analysis results will be forwarded either to the AdComparer (Press), or to the TVComparer (TV), or to the InternetComparer (Internet). These components detect new and existing creatives within a single media and a single language. The analysis workflow is controlled by the Collection Processing Manager (CPM) which is developed and described in WP6.

#### 4.3.1 Cross Media Advertisement Detection

Within this task four applications are developed, namely AdClipper, AdComparer (Press Workflow), TVComparer (TV Workflow) and InternetComparer (Internet Workflow). HSA improved the AdClipper and AdComparer software components and finalized the TVComparer implementation. These three applications are in form of a stable version available. The InternetComparer is currently under design and development.

#### 4.3.2 Audio analysis

The developed Audio Segmentation Module temporally segments the TV audio stream into segments of the classes "music", "speech", "other" sounds and "silence" and segments containing class mixtures. JRS improved the implementation of the Audio Segmentation in this period.

The speech recognition system was further improved in this working period. The ASR module is fully integrated in the MC communication infrastructure for the first prototype.

The Jingle Recognition Module matches fingerprints extracted from reference jingles with fingerprints extracted from TV spots by similarity matching. The Jingle Recognition Module has been improved e.g. adding task and result logger, MPEG-7 to AES31-1999 converter for visualisation of analysis results. In addition an implementation and testing of an alternative jingle recognition method based on

beat-synchronous chroma features was implemented. In addition JRS implemented a fully automated evaluation system for audio segmentation and jingle recognition. The main issues were: recognition criteria, computational performance, and consistency.

TNO finalized a first version for the Wordspotter module for the languages (English, Dutch and German). This module is fully integrated in the MC communication infrastructure for the first prototype and has been successfully demonstrated at the first review.

### **4.3.3 Visual analysis**

The different visual analysis modules have been implemented and continuously improved during the last 6 months of the project. One special focus was a detailed evaluation of all visual analysis modules.

The visual analysis is currently applied to the two main workflows for press (image fingerprinting module, press OCR module) and TV advertisements (TV fingerprinting module, logo recognition, TV OCR module) and will be applied in the near future also on the Internet Workflow.

- **Image Fingerprinting:** Advertisements in press usually consist of a single image which can include text, such as the company and product brand, a slogan or general product information. In cases where no text exists the visual similarity is the only modality that can be used to extract information from an investigated creative.
- **The image fingerprint module** was improved by the additional use of Histogram of Gradient (HoG) features for similarity search and by fine-tuning of the exact matcher. This can be seen as one major achievement in this working period.
- **Video Fingerprinting:** Video fingerprinting is the pendant to image fingerprinting for TV advertisements. Further improvements have been integrated in this module.
- **Logo recognition:** The modules are further developed in this working period.
- **OCR modules:** The modules have been further developed in this working period. For the video text block detection system was developed that uses temporal information to generate text blocks and which provides the possibility to recognize text (not only in still images) over a time period.

Furthermore the xml structure of all visual analysis modules was updated to the need of our project partners. All analysis modules were realized as WindowsServices and all basic libraries were checked for memory leaks.

### **4.3.4 Textual analysis**

Text Analysis modules have been designed and implemented during the last 18 months of the project. These modules are fully integrated in the MediaCampaign architecture and communicate with the different repositories via web services. The Text Analysis module finds instances of MEPCO concepts such as advertiser and product in textual documents obtained from OCR or ASR. The focus of the last 6 month was a detailed evaluation of the text analysis module. Further improvements will be carried out in the next working period.

#### 4.4 WP5: Knowledge Fusion and Campaign Discovery

WP leader	USFD
Objectives	<ul style="list-style-type: none"> <li>Use the composite data model to decide, for any particular semantic block, document or shot sequence, whether it belongs to an existing campaign or forms evidence for a new campaign. Develop hybrid SVM and heuristic model for fusion of evidence from: <ul style="list-style-type: none"> <li>OBIE from multiple sources from multiple media</li> <li>Features from audiovisual analytics</li> </ul> </li> </ul>
Milestones and expected results	<ul style="list-style-type: none"> <li>(II) MediaCampaign prototype software – cross relation of media campaigns</li> <li>(III) MediaCampaign integrated prototype - cross relation of media campaigns including knowledge base</li> </ul>
Deliverables	<ul style="list-style-type: none"> <li>No deliverables during the reporting period!</li> </ul>

The work package Knowledge Fusion and Campaign Discovery combines the information generated by the analysis modules described in the WP4, converts it into concepts of the MEPCO ontology of WP3 and finds relations between instances of these concepts. The relations between Creatives are kept within Campaign entities, and span over different media and languages. The instances of MEPCO are stored in the Ontology Management Service (OMS). The work package consists of the Cross Media Interlinking and Campaign Detection and Tracking tasks. No specific deliverables for WP 5 were due during the reporting period.

The main activities during the period have focused on the issues described in the following sections.

##### **4.4.1 Creative detection**

This module detects whether a given spot is an instance of an existing creative or the first instance of a new creative. This module combines the output of the analysis modules (audio, video and text). A new version of this module has been released for the MediaCampaign prototype and further improvements were continuously integrated in this module.

##### **4.4.2 Knowledge Fusion**

This module is used as a bridge between the creative detection and the campaign discovery. It is in charge of the communication but also of the mapping to the semantic concepts defined in the MEPCO ontology and used by the Campaign Discovery. A number of adjustments and modifications have been made on this component during the last 6 months.

##### **4.4.3 Campaign Discovery**

A new version of the Campaign Discovery module has been included in the MediaCampaign prototype and is fully functional. In particular, the output of this module is being used by the MC Delivery Portal and allows the visualisation of the data from individual spots to campaigns. Ongoing work has been done on this module to improve its functionalities and performance.

##### **4.4.4 Evaluation**

The mid-term evaluation covered the results of the creative detection module and the campaign discovery. This has been done in collaboration with the exploitation partner NMR. This initial evaluation will be used as a baseline for the following versions. A number of possible features has

been described and will be used in the next releases of the Creative Detector and Campaign Discovery.

#### 4.5 WP6: Component Frameworks

WP leader	SOF
Objectives	<ul style="list-style-type: none"> <li>• Definition of the overall MC system architecture</li> <li>• Design and implementation of MC workflow management component</li> <li>• Design and implementation of the MC cross-media knowledge repositories (essence, meta-essence, ontologies)</li> <li>• Design and implementation of applications for end-user delivery/reporting of MC analysis results</li> <li>• Integration of the components and set-up of a working demonstrator</li> </ul>
Milestones and expected results	<ul style="list-style-type: none"> <li>• (I) Overall system specification</li> <li>• (II) MediaCampaign Prototype (M 14)</li> <li>• (III) Media Campaign Integrated Prototype (M 27)</li> <li>• (IV) Final Demonstrator and MC Case Study tests at NMR-UK completed (M30)</li> </ul>
Deliverables	<ul style="list-style-type: none"> <li>• D6.2.1 Draft of Workflow Manager, Cross-Media Knowledge Stores &amp; Delivery, Navigation and Reporting</li> <li>• D6.4.1 Media Campaign first Prototype</li> </ul>

WP6 deals with the design, implementation and testing of the MediaCampaign system as a framework of components. The major achievement of this WP in the last reporting period was the completion and successful demonstration of the first integrated MediaCampaign prototype during the first EC review in June 2007. This corresponds to the fulfilment of the second milestone of the MC project.

All framework components have been implemented and framework integration has reached a good stage. Press and TV workflow have been tested during different integration meetings. This was quite an important period for this project work package. In June 2007 there has been the first project review where the system was presented to the Commission. The first period months (April and May) were spent mostly consolidating each system component and gluing them all together. The prototype is extensively described with D6.4.1 "MediaCampaign Prototype".

Regarding Task 6.1 (Architecture and Workflow Management) the main achievements have been the release of an improved Collection Processing Manager (CPM). Furthermore the definition of an Internet workflow has started. Part of the Internet workflow is the Internet grabber; this piece of software takes also care of the checking of internet material acquired by special sides (e.g. for finance, cars, etc.).

Regarding Task 6.2 (Cross-Media Knowledge Stores) the main results in the project period have been the submission of D6.2.1, with a detailed description of the different storages.

Regarding Task 6.3 (Delivery, Navigation and Reporting) a first release of the Delivery system was deployed for the first review, design and implementation of the web platform and of the provided services is still ongoing. The first prototype of the Delivery component was described in D6.2.1 too.

In the last part of the period, partners started focusing on the system set-up at NMR (UK), activities which go under Task 6.4 (Case Study) and Task 6.5 (Case Study, Demonstration). In this connection the next integration meeting hosted by TNO will be specifically focused on this issue.

#### 4.6 WP7: Exploitation and Dissemination

WP leader	HSA
Objectives	<ul style="list-style-type: none"> <li>• Transfer goals and achievements of the project to the public</li> <li>• Prepare the basis for commercialisation of project results after successful completion</li> <li>• To ensure maximum visibility of MediaCampaign results, to promote take-up of the technology in relevant industrial sectors, and to minimise the overheads in exploiting the project's R&amp;D outputs, both in commercial applications of the technology and in further research on related areas.</li> </ul>
Milestones and expected results	<ul style="list-style-type: none"> <li>• PM1: First version of the website</li> <li>• PM12: market analysis and first dissemination activity report</li> <li>• PM24: Update of Dissemination Activity Report</li> <li>• PM30: Exploitation Plan</li> </ul>
Deliverables	<ul style="list-style-type: none"> <li>• D 7.2.3 – D 7.2.5 Dissemination Activities Report</li> </ul>

The workpackage consists of two tasks, the first deals with exploitation planning and is headed by HSA, the second handles dissemination activities and is headed by USFD.

All deliverables have been submitted without significant delay. All important deliverables have been internally reviewed and revised according to the comments of the internal reviewer.

Overall it is to state that WP7 runs so far as planned, without major modifications and/or deviations from the schedule. The communication and climate is perfect and all partners did so far contribute with a satisfying level.

The main activities for the WP within the last reporting period are the following:

##### **4.6.1 Swedish potential customer in Stockholm**

The demo-system in Graz has been shown to a potential Swedish company. This company is also highly interested in other fields of media-analysis (brand detection) and therefore a potential customer for several MC components.

Following the presentation of the demo system in summer 2007, there was some more detailed discussion about Internet workflow and logo recognition. Requirements for Internet workflow have been discussed and a test-installation for logo-recognition at the customer's site has been prepared.

##### **4.6.2 Demo system**

A demo system for the press-workflow has been configured and set-up at HS-ART, in order to permit potential customer demonstration. It has been verified by a web-demo the staff of NMR-UK.

Some preparations with NMR-UK have been done, in order to prepare a test-setup at a real user-site. The system in Graz has been used with NMR-UK test data in order to verify the quality of the tools. Feedback has been given to the development-team for a next round of improvements.

#### **4.6.3 Dissemination Activities Report (D7.2.3)**

This deliverable has been compiled with contributions from various partners and has been submitted and accepted as a consequence of the first project review.

#### **4.6.4 Exploitation & Use Plan (D7.1.2)**

A first version of an exploitation plan has been completed and updated. Following the feedback from the first project review, a draft use plan has been added and submitted together with the annual report 2007. The document is a working document to be finally turned into the deliverable D7.1.2.

The product roadmap has been updated and improved by the findings made in the discussion with potential users (e.g.: Swedish customer).

#### **4.6.5 Scientific publications**

The following paper from UT has been accepted for the SAMT 2007 conference:

- Marijn Huijbregts, Roeland Ordelman and Franciska de Jong. Annotation of Heterogeneous Multimedia Content Using Automatic Speech Recognition. SAMT, December 5-7 2007, Genova, Italy.

Other relevant scientific presentations:

- SSMS'07 summer school in Glasgow, UK: MediaCampaign poster presented by JRS.
- Alps Adria Acoustics Association Conference, Graz, Austria: Presentation of the publication "Speech Music Discrimination in Mixed Audio Content" by B. Rettenbacher, M.Fellner from JRS.

#### **4.6.6 Public web-site**

There was a major update of the public website performed in the last reporting period in connection with the feedback from the first project review. A detailed description of the MediaCampaign prototype was added. Furthermore the web server was placed now outside from USFD solving the domain issues also mentioned during the first EC review (pages below the start page had a "USFD" domain and not the media-campaign domain). This issue is now solved.

## 5 Glossary

---

Terms used within MediaCampaign project sorted alphabetically.

- API** Application Programming Interface
- Campaign** A campaign represents a number of creatives semantically belonging together. A campaign has a certain duration and can be cross country and cross media. See also "spot" and "creative"
- Creative** A Creative represents all the occurrences of a similar spot. A Creative is not cross media (e. g. TV and press) and not cross language. See also "spot"
- CPM** Collection processing manager. Central component which controls the analysis workflow via web services.
- DB** Database
- DBMS** Database Management System
- EMS** Essence management store – holds all essence of MC, i.e. TV spots, press images and meta essence such as extracted audio files.
- JAPE RULES** Extraction patterns used to identify new entities
- HTML** Hypertext Markup Language
- HTTP** Hypertext Transfer Protocol
- IDE** Integrated Development Environment
- IE** Information Engineering
- iFS** Internet File System
- ISO/ANSI** International Organization for Standardization / American National Standards Institute
- META-ESSENCE** Under this term we understand all essence and additional data which is generated by subsystems (e.g. lo-res videos, keyframes, OCR text, ...).
- MMS** Metadata Management Store – holds all data extracted by the analysis systems, i.e. low-level and mid-level features such as tagging of spots in connection with logos.
- OMG** Object Management Group
- OMS** Ontology management store – holds all semantic knowledge of MC, i.e. the MEPCO ontology and the campaign knowledge.
- OS** Operating System
- PL/SQL** Procedural Language/SQL
- Re-detection** The same creative is re-detected in a single media. E.g. the ad of Renault Megane appeared on 5 May 2005 on "TV ORF1" 5.00 pm and on 7 June 2006 on "TV ORF1" at 4.00 pm.
- Scene** A consecutive series of shots connected through transitions (hard cut, fade, panning, ...) that constitutes a logical unit of action in a video. It is defined by "bridging features" such as same visual/audio content, music/speech/noise segments or text overlays.
- SGML** Standard Generalized Markup Language
- Shot** A consecutive series of pictures representing coherent visual content, e.g. when having an interview with two persons, and the camera is changing

between the two faces (depending on who is talking), each face would be a shot.

**SOAP** Simple Object Access Protocol

**SQL** Structured Query Language

**Spot** An occurrence of a Creative, e.g. TV spot seen on a given channel at a specific time. The incoming material obtained from the Media Acquisition step is initially available as a spot and later attached to a new or existing Creative. See also "creative".

**S&R** Search and Retrieval

**Tracking** Tracking of creatives by means of history over media and countries. When a creative is re-detected (see also "re-detection"), it is inserted into a data structure which represents the history of this creative. This enables users to query the database for appearances of this creative in the past (e.g. it had appeared on 1 April 2006 in press "The Times" in US p.5 and on 3 April 2006 on "TV Sky 1" at 19.00 in UK.

**WP** Work Package

**UML** Unified Modelling Language

**URI** Uniform Resource Identifier

**XML** Extensible Markup Language

**XPath** XML Path Language

**XSLT** XSL Transformation

#### Partner Acronyms:

**HSA** HS-Art Digital Service GmbH

**JRS** JOANNEUM RESEARCH Forschungsgesellschaft mbH

**NMR-UK** Nielsen Media Research UK

**ONTO** Sirma AI EAD

**SOF** Softeco Sismat SpA

**TNO** Netherlands Organisation for Applied Scientific Research

**USFD** University of Sheffield

**UT** University of Twente