

# 1 Publishable Summary

## 1.1 Project Objectives

The main goal of *inEvent* is to develop new means to structure, retrieve, and share large archives of networked, and dynamically changing, multimedia recordings, mainly consisting here of meetings, video-conferences, and lectures. Several partners of the *inEvent* consortium have indeed access to (and continuously generate) such large multimedia repositories, which keep being enriched everyday by new recordings, as well as social network data. The resulting resources often share common or related information, or are highly complementary, but also come from different sources, in different formats, and different types of metadata information (if any). Hence, it is still impossible to properly search across those very rich multimedia resources simply based on metadata.

Exploiting, and going beyond, the current state-of-the-art in audio, video, and multimedia processing and indexing, *inEvent* aims at developing and testing a system that addresses the above problem by breaking our multimedia recordings into interconnected “hyper-events” (as opposed to hypertext) consisting of a particular structure of simpler “facets” of rich audio-video recordings, which are then easier to search, retrieve and share. Building and adaptively linking such “hyper-events”, as a means to search and link networked multimedia archives, should result in more efficient search systems, in which information can be retrieved based on “insights” and “experiences” (in addition to the usual metadata).

Reaching the aforementioned goal requires challenging RTD efforts going much beyond current state-of-the-art in the fields of knowledge representation, audio processing, video analysis, semantics of information, and exploitation of social network information.

## 1.2 Project progress overview and most significant achievements

Progress of the project so far has been very satisfactory. As discussed in Section 2, the project has currently reached all its milestones and has developed all the necessary data capture and analysis tools, specification of standard data (hyper-event) formats to analysis services, and specification and first prototype of the associated archiving system (possibly demonstrated at the review meeting). This first prototype is actually coming 12 months earlier than expected but was considered necessary to maximize collaboration between the partners and integration potential of the different technologies.

More specifically, the main achievements of the project after the first 12 months, can be summarized as follows:

- Definition and exploitation of the fundamental architecture of capture, analysis and storage blocks, and matching the requirements of Radvision and Klewel’s capture device for each of the defined use cases (Lecture and Meeting).
- Analysis of user requirements, user interaction and user evaluation. Work began by generating a set of appropriate user scenarios that can guide the development of the system. Based on these use-cases, we defined the system requirements from the user’s perspective. Part of the system requirement definition was defining the interaction scenarios of the user with the system.
- Definition of the characteristics of the data that will be used throughout the project, and collecting the initial data set so that it is available for processing and indexing to all the consortium member. The resulting data has been made available and started being processing for audi-video feature extraction, semantic indexing, and hyper-event building.

The initial dataset for *inEvent* is ready to use. The dataset consists of video and metadata files from three different sources: two project partners, Radvision (meeting recordings) and Klewel (lecture recordings), and the external TED repository (<http://www.ted.com>) to perform additional experiments on out-of-project data. The metadata files contain information that will be used by indexing engine in order to index the related media files.

- Developing automatic video and speech processing algorithms that can extract relevant clues from the extremely challenging data on which *inEvent* focuses. This included the development of new approaches for speaker diarization and linking, development of multispeaker conversational speech-to-text transcription systems for the lecture and videoconference domains, and development of video processing algorithms robust to the high variability of the *inEvent* data, including detection and tracking of human body parts, face detection, person counting in videoconference scenarios, and analysis of VGA videos containing presentations.
- Specification and development of a first prototype of the *inEvent* archiving system. The environment consists of a database, file repository, application server, and the first version of the archiving system deployed on it. The system is able to handle different REST API calls, defined through the backend API protocol of the archiving system.
- Hyper-event based archiving system can already be demonstrated; a very preliminary prototype system accommodating search by example (hyper-events related to a given one) and personalized recommendations (hyper-events of interest to a given user) were implemented and tested on the large TED archive (ca. 1100 events).

### 1.3 Expected final results

Ultimately, the main goal of *inEvent* could thus be summarized as developing new ways to replace the usual “hypertext” links (linking text and “information” bits) by multi-faceted “hyper-events”, linking different “experiences/insights” related to dynamic multimedia recordings. The resulting system should be able to automatically assimilate rich original audio-video media (mainly project generated data, in our case), index and structure hyper-events using federated approaches (audio-video feature extraction, semantic linking and structuring), and supporting user interaction and queries along the hyper-events.