1 Publishable Summary

One-hundred hours of video uploaded to YouTube per minute, the success of online conferences like TED, and the propagation of video conferencing in the business world have led to an avalanche of recorded multimedia content. Yet this content, by its very nature, often lacks important metadata (such as audio transcripts, A/V semantic interpretation, links between events, which can be personalized) that is needed to structure databases and create meaningful links between content, hence maximizing its exploitation opportunities.

The *inEvent* European FP7 Project, by building a collaboration between industry and academia from 2011 to 2014, has achieved a new phase in recording, processing, automatically recommending and navigating this deluge of multimedia data through the creation of **interconnected hyper-events**, where hyper-events (by analogy to hyper-texts and hyper-links) are defined as complex audio-video recordings that are indexed across several dimensions and linked/searched along different facets.

The overarching goal of the *inEvent* project was to create an end-to-end solution for archiving and accessing these vast multimedia databases. Going beyond current state-of-the-art multimedia processing and indexing, the project generated intelligent, standardized and interoperable database content, and shared it on a web-based platform where search and innovative visualization tools (e.g., navigation graphs) help the user discover and connect related content.

1.1 Project objectives

The main goal of *inEvent* was to develop new means to structure, retrieve, and share large archives of networked, and dynamically changing, multimedia recordings, mainly consisting 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 have 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 representation, *inEvent* aimed at developing and evaluating a system that addresses the above problem by processing and structuring raw multimedia recordings as interconnected *hyper-events* (by analogy to hyper-text) composed of simpler "facets" of rich audio-video recordings, which are then easier to search, retrieve and share. Building and adaptively linking such hyper-events according to their relatedness or similarity, *inEvent* offers innovative and efficient means to search, connect, and navigate through network multimedia archives in an end-user web portal.

Reaching the aforementioned goal required challenging RTD efforts going much beyond current state-of-the-art in the fields of knowledge representation, audio processing, video analysis, semantics of multimedia information, and exploitation of social network information. More specifically, and as further discussed in Section 2.1, *inEvent* had identified in its Description of Work (DoW) **four major objectives** which had to be achieved over the three years of the project:

1. **Objective 1**: Extraction of hyper-events, including semantic interpretation, from large scale, heterogeneous A/V data (presentations). *inEvent* will automatically extract semantic and social metadata from multimodal resources to enable the construction of dynamic, complex hyper-events.

2. **Objective 2**: Building and exploiting links between hyper-events. *inEvent* shows that the concept of hyper-events can serve as the fundamental organizational principle for dynamic multi-

media systems, in order to help users to access relevant media information in an efficient and intuitive way.

3. **Objective 3**: Demonstrating on real problems, and testing with users from different communities, the usefulness of the novel structure on large amounts of multimedia archives.

4. **Objective 4**: Standardization of the multimedia event archives. Since multimedia recordings will come from heterogeneous sources, environments, and applications, it is important to make sure that the processed multimedia resources accommodate multiple types of metadata and enable common processing towards the automatic extraction of semantic and social metadata.

The present Deliverable presents the main work done during *inEvent* Y3 towards these objectives, while Deliverable D6.5 (Final Report) presents an overview of the main achievements resulting from the *inEvent* three-year efforts:

- In **Section 1.2**, we give an overview of what we see as the main *inEvent* Y3 achievements.
- In **Section 2.1**, we summarize the main Y3 achievements in the light of the four objectives.
- In **Section 2.2**, we discuss the actions taken to comply with the comments and recommendations of the last review report.

1.2 Project progress overview and most significant achievements

As discussed in Section 2, *inEvent* has currently reached all its milestones. Comments from the last review report were also particularly helpful, and greatly motivated us to focus most of our last year resources on:

1. Full end-to-end system integration: As discussed in detail in Deliverable D3.6, and summarized in Section 3.4, page 32, of the present report.

2. Iterative user evaluation and system improvement: As discussed in detail in Deliverable D4.6, and summarized in Section 3.5, page 36, of the present report.

3. Compilation of all *inEvent* technology components and their IPR status: As also suggested in the last review report, we paid special attention to all technology components, properly listed in a single table also including integration status (integrated in portal or proof of concept), how it is delivered (web service, executable, open source code, library, command line library, etc), how each technology component has been evaluated (yes/no, scientific publication comparing with state-of-the-art, on different datasets to show inter-operability, user feedbacks, QA, etc), IP originator (institution) and IPR agreement (open source, ownership, binary only, etc), and main interested industrial partner(s). This is discussed in detail in **Deliverable D5.4**, and summarized in Section 3.6.3, page 41.

4. System integration: Besides further development and improvement of the technology components, the *inEvent* system provides a fully integrated end-to-end solution to multimedia data archiving, processing, indexing, searching (browsing), and recommendation. As one of the *inEvent* central components, the archiving system described in Section 3.4.3, page 33, was further developed and finalized according to the block diagram presented in Figure 3, page 13, to comply with all the integration needs of the technology components, the interface (portal), and the user requirements. It also had to be adapted (based on clear and publicly available

APIs) to be fully automatic, being able to upload input data (coming from Klewel, Radvision, or TED), call relevant web services (based on the defined API specifications), automatically upload extracted metadata, and synchronize/link them with the raw data, as well as social data; and answer user's portal requests while browsing or searching for specific information.

5. User requirements and user-based evaluation: The portal's design includes a graphical interface to show the network of relationships among recordings in an archive and a "hopping" interface that shows a timeline with details of e.g. slide changes, applause, and speech in a meeting or lecture to speed up browsing. As these features are highly innovative, users required seeing them within working interfaces in order to comment upon them. A first round of interviews demonstrated that users value the visual attractiveness of the graphical front-end to the archive, as well as the information it provides beyond simple ranked lists (state of the art). However, they did lead us to recast our interface concept so that users would understand how to use it more easily. A second round of interviews with the recast interface demonstrated acceptance of the recasting, but among a small set of potential users with higher technology skills than our expected user pool. Smaller exercises were used to check the interpretability of our icon choices, particularly in the hopper, and display options for showing transcription and summarizing the content of a recording at a glance. We then engaged in two exercises to evaluate the interface among larger and more realistic user communities: via a workshop with contacts from business and NGO contacts from the areas around Geneva and Lausanne, and via a web-based usability survey.

6. Long-term sustainability: As discussed in Section 3 of Deliverable D5.4, several measures have been taken to guarantee the sustainability of the *inEvent* system, including:

   - API specifications: The *inEvent* API specifications were reviewed, homogenized, extended, and published online. The API was extended to describe specifications for sending and receiving requests from processing web servers responsible for audio and video analytics (developed in WP2). The latest API version is clearly described and available at: http://inevent.klewel.com/api/.
     Users can quickly jump to a defined construct to check its required and optional attributes, and access the specifications for issuing a web service request for archiving, retrieval, and processing.
   - The source code of the *inEvent* portal is now published as open source on Github:
     https://github.com/idiap/inevent
     under the MIT license (yielding a free access to the software without restriction, and without limitation to use, copy, publish, and distribute it with an acknowledgement). A quick installation guide is also provided.
   - To ensure the sustainability of the *inEvent* portal for at least two years beyond the end of the project, Klewel dedicated one of its own servers for the *inEvent* end-user portal and the *inEvent* archiving and repository system. As mentioned previously, the *inEvent* portal exchanges data with the archiving system using REST web services. Up until the third year of the project, the IBM servers used to host the archiving system. In order to support sustainability, IBM replaced some proprietary components with open-source ones, and provided the binary code for the archiving system to Klewel. The database and the media repository were also transferred to the Klewel server dedicated for the *inEvent* portal. Keeping the portal up and running beyond the project, is rewarding for inEvent communities (such as FRESH) who would benefit from a prolonged access to their recordings.

Moreover, after having reached a stable and user-validated version, keeping the portal alive would encourage other end-user communities, researchers and developers to use and build upon it, and promote the adoption of the underlying *inEvent* technologies, which usage is demonstrated in the portal.

1.3 Project website and statistics

1.3.1 *inEvent* website

To properly wind down the project, we fully revisited the *inEvent* web site:
https://www.inevent-project.eu/
to provide clear and concise information about the project.

- We include on the homepage only the most relevant Y3 information, such as the training event, the *inEvent* business workshop, our collaboration with ICC'2014, and participation to the FRESH conference. Various webcasts are also available in the right column, together with a copy of the public leaflet. The menu on the left side gives access to all demonstration systems, intranet (wiki, with list and minutes of meetings, workshops, etc), as well as publications.
- Under https://www.inevent-project.eu/project/gallery we made available a collection of schemas and photographs that can be used to document the project.
- The Intranet at https://www.inevent-project.eu/intranet contains a Wiki for internal communication and links to the SVN server, as two ways to share notes and software.
- As made explicit in more detail just below, we have consolidated all possible *inEvent* resources into one single space of the *inEvent* web site (named "Codes, API, and data resources"):
  http://www.inevent-project.eu/codes-api-data-resources
  including pointers to software (with source code), APIs, and data resources.

1.3.2 Codes, API and data resources

To guarantee the full and public accessibility and potential exploitation of the *inEvent* system, we have consolidated all possible *inEvent* resources into one single space of the *inEvent* web site (named "Code, API & Data Resources") including pointers to the following software, APIs, and data resources:
http://www.inevent-project.eu/codes-api-data-resources

- Open source code:
  - Front-end code of the *inEvent* portal:
    https://github.com/idiap/inevent
    The open source code of the inEvent online portal is available on *github*, under the MIT license (yielding a free access to the software without restriction, and without limitation to use, copy, publish, and distribute it with an acknowledgement). A quick installation guide is also provided.
  - Speaker diarization source code:
    http://www.idiap.ch/software/speaker-diarization-toolkit/spkr_diar_ib_rel.tar.gzt

  - Hyperevent recommendation code for two dimensions:
    Content-based recommendation: https://github.com/idiap/cbrec
    Emotion-based recommendation: https://github.com/idiap/emorec
  - Transcription producer and meeting summarization code:
    https://github.com/laic/summarization

- API specifications:

  - Main links:
    http://inevent.klewel.com/api
    http://inevent.klewel.com/api/-DTOdefinitions
    The API specifications are available online. The Data Transfer Objects and REST services defined for archiving, retrieving, and processing multimedia are described at the URL above.
  - Public dataset retrieval:
    http://inevent.klewel.com/api/#Retrieval
    The retrieval section of the API documentation describes how to download all the dataset, or a subset of it based on the filter and search query specified. Every hyper-event retrieved is returned along with:
    * Its metadata
    * Related events (through emotion-based and content-based similarity)
    * Tracks (video files, slides, audio and video processing outputs when available)
  - Access to Audio and Video Processing Services via the REST API:
    http://inevent.klewel.com/api/#Processing
    The processing section of the API documentation describes the specifications defined to issue a web service processing request for:
    * Speaker diarization.
    * Automatic speech recognition.
    * Video analytics.

- Public dataset description:

  - The public dataset covers 1250 hyperevents representing approximately 1500 hours from Klewel, Radvision, and TED talks along with their metadata, and the output of audio and visual analytics (for a subset of them), in the form of tracks.
    Radvision's dataset contains a total of 14 recordings of *inEvent* meetings, *inEvent* webinar 2013, and Fresh 2014. A detailed description of the TED dataset used is available here: https://www.idiap.ch/dataset/ted/.
    Klewel's dataset contains university lecture and events such as international congresses, conferences, and symposiums recorded using Klewel's capture station. Klewel's dataset consists of 123 hyper-events, corresponding to approximately 73 recorded hours of content including EPFL lectures and presentations from the Djangocon 2012, Fresh2012, and Lift 2011 conferences.

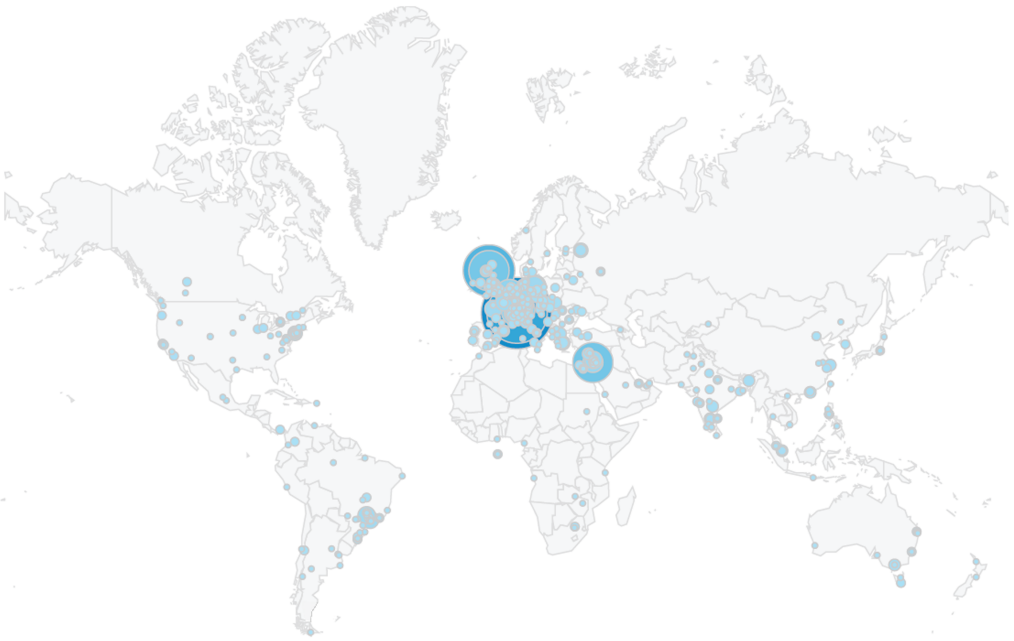

Figure 1: *Cities from where website sessions were initiated.*

1.3.3 Project website statistics

As requested by the review panel, we provide below some statistics regarding the *inEvent* website.

- The *inEvent* website has received 6'871 session visits, cumulating in 31'603 page views.
- 50% of these visits were new sessions, while 50% of the visitors were returning ones.
- The average time of visits is 3.5 minutes.
- Without any surprise, the home page is the most visited one, however it is worth mentioning that the Publications page is the third most visited page and the Demo page the fifth one very close behind.
- Visitors are from all around the world. Figure 1 shows the cities from where the sessions were initiated. The majority of the visitors are from Europe (70%), and the remaining 30% are mainly visitors from Asia and the Americas.

1.4 Project Deliverables

Deliverables are available to the reviewers at:
https://www.inevent-project.eu/intranet/wiki/deliverables.
They are protected by a password which is made available through a secure channel.

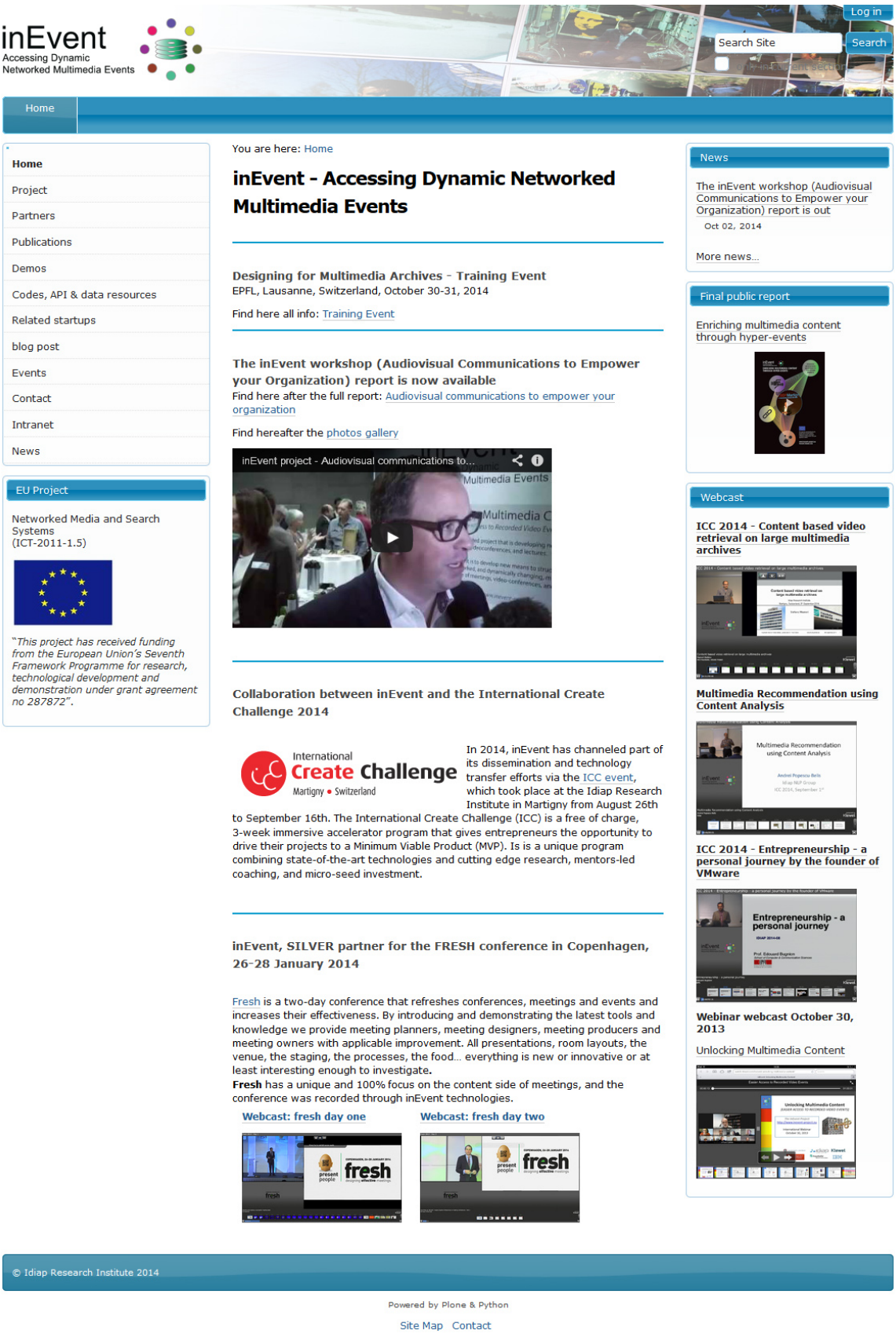

Figure 2: *Snapshot of the home page of the* inEvent *website as of November 21, 2014. Under "Publications", 27* inEvent *papers are available. Under "Demos", several demonstrations of* inEvent *components are available, including Automatic speech recognition of lectures, Audio-video processing,* inEvent *Portal,* inEvent *navigation graph, and Automatic recommendation of lectures. The webcast of the* inEvent *webinar (Easier Access to Recorded Video Events) is also fully available under "Events".*