

### 3.1 Publishable summary

#### 3.1.1 Project synopsis

During the past decade, the rapid development of digital technologies has led to a great increase in the availability of multilingual and multimedia content worldwide. This content can be repetitive or complementary across political, cultural, or linguistic borders but can be also contradictory and in some cases unreliable. The consumption of such large amounts of content regardless of its reliability and cross-validation can have important consequences on the society and especially on journalism, media monitoring and international investments.

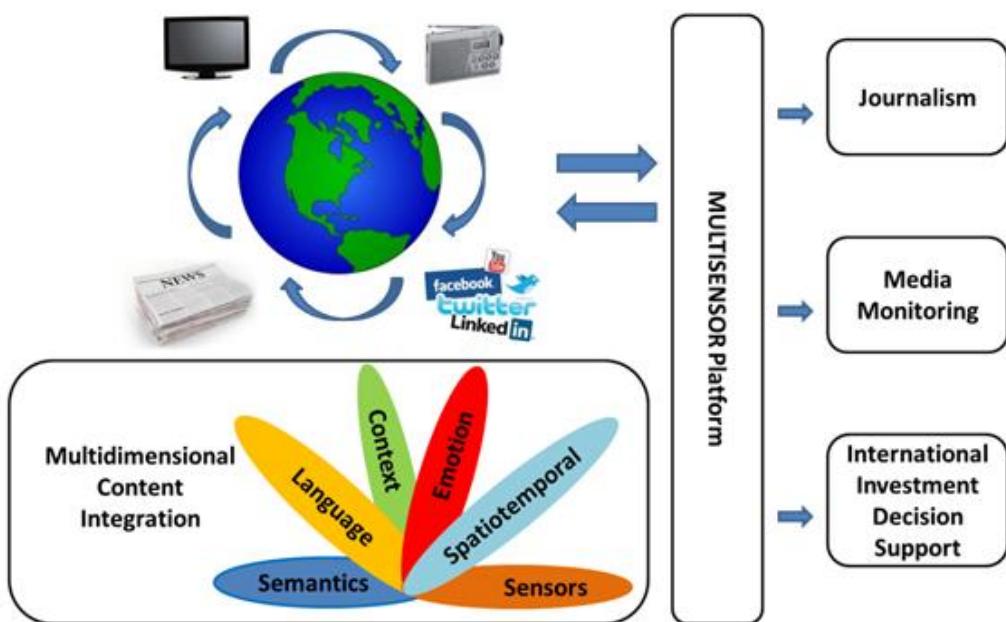


Figure 1. The MULTISENSOR concept

In this context, MULTISENSOR aims at providing unified access to **multilingual** and **multicultural** economic, news story material across borders, that ensure its **context-aware**, **spatiotemporal**, **sentiment-oriented** and **semantic interpretation**, and that correlate and summarise the content into a coherent whole. MULTISENSOR envisages an integrated view of heterogeneous resources sensing the world (i.e. sensors) such as web articles, newspapers, international TV and social media. The approach of MULTISENSOR will build upon the concept of **multidimensional content integration** (Figure 1) by considering the following dimensions for mining, correlating, linking, understanding and summarising heterogeneous material: language, multimedia, semantics, context, emotion, as well as time and location. The **overall goal of MULTISENSOR** is to research and develop a **unified platform**, which will allow for the **multidimensional content integration** from heterogeneous sensors, with a view to providing **end-user services** such as journalism, commercial media monitoring, and decision support for SME internationalisation.

To realise this goal MULTISENSOR addresses a number of scientific (SOs), technological (TOs) and business objectives (BOs):

- SO1. To perform **mining and content distillation of unstructured heterogeneous and distributed multimedia and multilingual data**.
- SO2. To perform a user- and context-centric analysis of **heterogeneous multimedia and multilingual content**.

**SO3.** To enable the semantic integration and context-aware interpretation over the spatiotemporal and psychological dimension of heterogeneous and spatiotemporally distributed multimedia and multilingual data as audio, video, text and social content interaction.

**SO4.** To provide semantic reasoning and intelligent decision support services.

**SO5.** To facilitate context-aware multimodal aggregation and multilingual summarisation and adequate presentation of the information to the user.

**TO1.** To develop crawlers and data delivery channels necessary for the collection of data.

**TO2.** To validate the technologies developed in MULTISENSOR by implementing a number of representative scenarios of pilot use cases that target journalism, international commercial media monitoring and international investment decision support.

**TO3.** To deliver an operational demonstrator

**BO1.** To define a business model for the exploitation of MULTISENSOR results by the partners and the consortium as a whole.

During the first two years of its lifetime three different applications have been developed based on MULTISENSOR technologies. The first provides search and exploratory functionalities for journalists, the second aims at supporting a media monitoring company to monitor specific profiles for their clients, while the third one provide decision support for SME internationalisation.

The MULTISENSOR Consortium consists of nine partners, each of them contributing their competence and experience in one or several fields crucial for the success of the project.

| Participant no. | Participant organisation name   | Part. short name | Country  |
|-----------------|---|------------------|----------|
| 1 (CO)          | Centre for Research and Technology Hellas, Information Technologies Institute | CERTH            | Greece   |
| 2               | Universitat Pompeu Fabra  | UPF              | Spain    |
| 3               | Fundació Eurecat  | EURECAT          | Spain    |
| 4               | Linguatec   | LT               | Germany  |
| 5               | Everis Spain SLU  | EVERIS           | Spain    |
| 6               | Pressrelations  | PR               | Germany  |
| 7               | Ontotext  | ONTO             | Bulgaria |
| 8               | Deutsche Welle  | DW               | Germany  |
| 9               | PIMEC   | PIMEC            | Spain    |

### **3.1.2 Summary of activities during the first and second years**

During the first and second years of its lifetime, MULTISENSOR achieved considerable progress with respect to all of its objectives. During the first year, empirical studies of the material were performed and the technical infrastructure was set up for all modules. The user requirements were defined, based on which the MULTISENSOR technical specifications were set and the platform architecture was designed. This architecture was realised by an operational prototype, which integrates the skeleton versions of the research modules and served as fundament for the work in the Project in the second year. Specifically, during the second year, the first and second prototype versions have been developed, integrating the initial and advanced versions of the research modules, respectively. Moreover, a detailed exploitation plan for the project has been drafted. Finally, in the context of each individual objective, further targeted actions have been performed.

For **SO1** the developed technologies during the 1<sup>st</sup> year concerned the development of the baseline technologies for the name entity extraction module, the deep syntactic parser, the speech

recognition and the machine translation module, as well as the concept extraction mechanisms from multimedia. In the 2<sup>nd</sup> year, the advanced versions for all the aforementioned modules have been developed.

With respect to **SO2**, during the 1<sup>st</sup> year, the implemented technologies included the context extraction module, the infrastructure for sentiment classification and a module for contributor analysis in social media. During the 2<sup>nd</sup> year, the activities concerned the extraction of advanced context features, the development of sentiment extraction techniques, the definition of indicators for media monitoring and SME internationalisation, as well as the development of the influential user detection and community detection modules.

During the 1<sup>st</sup> year, with respect to **SO3**, we have implemented a framework for ontology alignment, a baseline module for topic-based classification and a representation model for socially interconnected and multimedia enriched objects (SIMMO). During the 2<sup>nd</sup> year, we have researched and developed an advanced framework for ontology alignment, which combines different matchers (string, structural and visual). We have also developed a topic detection technique by extending the DBSCAN clustering algorithm. Finally, we have realised the SIMMO model as a retrieval system.

With respect to **SO4**, during the 1<sup>st</sup> year, the basic semantic representation and infrastructure was defined, covering a variance of ontologies and linked open data loaded in the semantic repository of the infrastructure. During the 2<sup>nd</sup> year, MULTISENSOR developed and implemented four different types of reasoning techniques – Parallel reasoning, Hybrid reasoning, GeoSPARQL and SPARQL-MM. In addition, the RDF Storing service was implemented, the main purpose of which is to convert the SIMMO object to RDF and store it into GraphDB. Finally, the first version of the Decision Support System and the first version of the Recommendation System were developed.

In the context of **SO5**, during the 1<sup>st</sup> year, the developed technologies included the basic summarisation infrastructure and the extractive summarisation pipeline, which integrates a set of Simple Multi-document Summariser (SUMMA) modules. During the 2<sup>nd</sup> year of the project, we implemented advanced methods for extractive summarisation and linguistic generation of abstractive summaries from text plans. Moreover, a text planner for abstractive summarisation was implemented. Finally, we implemented an advanced hybrid rule-based and statistical linguistic generation module.

In addition, substantial progress has been reported in the technological objectives. In the context of **TO1** during the 1<sup>st</sup> year, a prototype social media crawler has been implemented, while the crawler of a commercial media monitoring company (PR) has been integrated into the crawling framework of MULTISENSOR. During the 2<sup>nd</sup> year, the crawling framework integrating different crawlers (PR, EURECAT and CERTH) for news and social media has been finalised.

Following the workplan, with respect to **TO2**, during the 1<sup>st</sup> year three main use cases have been designed relevant to journalism, commercial media monitoring and SME internationalisation and based on these the user requirements have been extracted. During the 2<sup>nd</sup> year, the activities concerned the evaluation design, as well as the evaluation of the First Prototype.

Towards the achievement of **TO3**, in the 1<sup>st</sup> reporting period the operational prototype of the MULTISENSOR platform has been implemented integrating the skeleton of the content analysis services developed in the context of SO1-SO5. During the 2<sup>nd</sup> year, the 1st and the 2<sup>nd</sup> prototype of the MULTISENSOR platform has been implemented integrating the baseline and the advanced services respectively in the context of SO1-SO5.

During the 1<sup>st</sup> year, with respect to **BO1**, a detailed analysis of media monitoring and SME markets has taken place. During the 2<sup>nd</sup> year, a detailed exploitation plan to be followed for the rest of the project has been drafted and the market analysis has been updated.

In general, the project progresses as planned, with continued collaboration and good team spirit among the partners. The objectives of the first and second year were achieved successfully and corrective actions took place, in order to address any deviations.

### **3.1.3 Future work**

In the final year of its lifetime, MULTISENSOR will carry on the research and development tasks related to the individual scientific, technological and business objectives and the accomplishment of the milestones.

Specifically, for the third year, MULTISENSOR will work towards the achievement of the fifth milestone (MS5), which concerns the development of the final MULTISENSOR platform and its evaluation, as well as the exploitation of the project results.

### **3.1.4 Expected final results and potential impact**

The expected final result of MULTISENSOR is a platform that will support a) the journalists in mastering large and heterogeneous content in order to prepare articles and identify topics, relevant and contradictory content, as well as have access to multilingual multidocument summaries; b) the commercial media monitoring companies to summarise the opinions of people for specific products and c) the SMEs that want to internationalise by providing market analysis, product reports and decision support services. This platform will integrate and make use of innovative modules, which could be separately exploited. These include modules for name entity extraction, speech recognition, machine translation, concept detection from text and multimedia, topic classification and detection, multimedia retrieval, semantic search, ontology alignment, reasoning and decision support, as well as summarisation. During the first two years of its lifetime, MULTISENSOR has already developed three applications integrating the advanced and baseline versions of the aforementioned modules.

In its final stage of development, MULTISENSOR is expected to have a big impact from several perspectives. First, it will support actively the journalists (professional and amateurs), commercial media monitoring companies and the international investments by SMEs. Second, the SMEs in the ICT domain will benefit from the open source tools and technologies developed in MULTISENSOR, in order to improve their existing products and offer new services to their clients. Third, the development of such tools will boost the competitiveness specifically in the media monitoring domain and in general in Europe, since the mobility of SMEs will be facilitated. Finally, the social impact of MULTISENSOR will be the production of cross-validated news articles and the presentation of news stories from different cultural, political and linguistic perspectives.

### **3.1.5 Contacts and Information**

For further information visit the project web site <http://www.multisensorproject.eu/> or send e-mail to the Project Coordinator Dr. Ioannis Kompatsiaris, [ikom@iti.gr](mailto:ikom@iti.gr) or to the deputy Project Coordinator & Scientific Manager Dr. Stefanos Vrochidis, [stefanos@iti.gr](mailto:stefanos@iti.gr). You are also welcome to join us in social media:

Facebook: <https://www.facebook.com/pages/Multisensor/1481238748826033>

Twitter: <https://twitter.com/multisensor>

LinkedIn: <http://de.linkedin.com/in/multisensor>

