Intelligent Synthesis and Real-Time Response
using Massive Streaming of Heterogeneous Data
www.insight-ict.eu

1 Second Year Report

Project INSIGHT, "Intelligent Synthesis and Real-Time Response using Massive Streaming of Heterogeneous Data", is a Specific Targeted Research Project that runs since September 1st 2012 involving the following partners: i) National and Kapodistrian University of Athens (UoA), ii) IBM Ireland (IBM), iii) Fraunhofer Gesellschaft zur Foerderung der angewandten Forschung e.V IAIS (Fraunhofer), iv) Technische Universitaet Dortmund (TUD), v) Israel Institute of Technology (Technion), vi) Federal Office of Civil Protection and Disaster Assistance (BBK), vii) Dublin City Council (DCC). This documents summarizes the objectives of the project and highlights the accomplishments achieved during the *second* year of the project.

1.1 INSIGHT Context

Today's Environment Today's technological advances are fuelling a virtual explosion on the quantity, quality, and variety of information that is becoming available. As a consequence, new challenges have risen due to three recent revolutionary technologies: a) Advances in sensor networking and the availability of low-cost sensor-enabled devices in transportation, healthcare and emergency response, b) the widespread adoption of smart-phones and c) the extent of social networks such as Twitter and Facebook. Each of these revolutionary technologies is driving the development and adoption of applications where mobile devices are used as tools for continuous data sensing, collection and analysis.

The INSIGHT goal The goal of INSIGHT is to exploit and develop further the new capabilities that the combined use of these technologies is offering, in order to bring forth fundamental advances in the Smart City paradigm, focusing especially on emergency situations. We aim to achieve significant improvement in the utility of automated systems to manage resources and put new, more capable tools in the hands of disaster planners and city personnel when responding to emergencies in smart cities and countries.

Current systems are limited in several important elements: (i) There is a lack of methods for handling heterogeneous data streams in real-time; (ii) there is limited integration of big data analytics and social computing, (iii) real-time prediction and alarm capabilities have not yet been incorporated into the infrastructure for intelligent management.

To truly achieve quantum improvement, we have to integrate information that comes from the analysis of complex data of very different kinds, for example combine quantitative data from physical sensor with qualitative data from social networking media. The inherent complexities are immense, necessitating novel work both on algorithmic and systems aspects. We plan to create an infrastructure that will provide a long-term support for building, maintaining, and improving such systems. Our efforts will be realized in the context of two carefully

selected use cases that aim to develop new capabilities in managing emergency situations:

1. *City Level Use Case:* The City of Dublin is a Smart City already equipped with an infrastructure of sensor that collects a wide variety of data from the urban environment. We leverage the availability of such rich data, in combination to exploiting social media data, to enable traffic managers to detect with a high degree of certainty unusual events throughout the network. For example, we aim to develop novel methods to ameliorate the impact of urban flooding on the capacity of the road system, through monitoring, analysing and combining sensor data and social data to improve emergency response and evacuation and provide the necessary information to stakeholders and end-users.

2. *Nation-wide Disaster Monitoring:* The aim is to achieve early warning of experts at the national joint situation and situation center and monitoring of disasters with nation-wide impact, such as severe weather conditions, floods, evacuations and their subsequent events, such as fires, power outages. The INSIGHT project will develop a system to provide big data monitoring and reliable detection of disasters in near real-time, operating on combined data from stationary, mobile and social sensor networks to detect, to verify the existence and to visualize multiple facets of a disaster, and manage its consequences, thus improving our capability to cope with natural disasters. Our ambition is to gain three-dimensional insights into large scale disaster situations: social / socio-emotional, physical and activity dimension.

1.2 INSIGHT Objectives

The INSIGHT project aims at creating an open, flexible and powerful solution and developing the capabilities for fundamental advances in coping with emergencies in the Smartcities of the future. To achieve our goal, the project has the following innovation objectives:

1. To develop an adaptive, scalable, real-time infrastructure for emergency monitoring.
2. To develop radically new methodologies for monitoring, processing, analyzing and synthesizing massive amounts of heterogeneous data for improving our ability of coping with emergencies.
3. A major objective of INSIGHT is to facilitate the adaptation and use of the technologies that we will develop: To ensure re-usability and facilitate faster adaptation of the proposed methodology, we will provide appropriate customized user-interfaces, as well as release code in the public domain.

1.3 Activities and Results of Years 1 and 2

From Sept 2012 to August 2014 the work of the consortium focused on the the development of advanced algorithms and software components for event detection and event recognition, defining an ontology of disastrous events, a generic information fusion framework as well as on disseminating project results towards the scientific community and to the public. The work of the first two years demonstrated the strong collaborations between the partners, and the focus on designing realistic use-cases and evaluation scenarios. In summary, the following are the main activities and results during year one and year two of the project:

- *Algorithmic and Systems Progress:* The consortium made important progress at all algorithmic and system components that will be employed in realizing INSIGHT. Novel techniques were developed for: (i) Traffic modelling; recent work has focused on scaling novel modelling techniques to city- and country-scale. (ii) Complex event processing; investigated the scalability of complex event processing techniques to the size of the datasets we expect, and the ability to monitor for events online. (iii) Twitter and social media monitoring and analysis; developed novel techniques for geo-locating twitter messages, techniques for monitoring emotional response in social media in real time, novel visual analytics techniques for spatio-temporal analysis of geo-located tweet datasets and real-time event detection. (iv) Monitoring heterogeneous data sources; investigated techniques for merging heterogeneous data sources, both merging traffic data from different sensors in real-time, and by combining social media data with static data to provide real-time feedback. We have also developed a generic framework for fusing information (Round Table). (iv) Active learning and crowdsourcing; we developed novel algorithms for user selection and an approach for using crowd-sourcing to discover and track events. (v) Defined an ontology of disastrous events thaw will be utilized for communicating information between the aforementioned components.
- *Integration:* First versions of these components have been integrated into a proof of concept prototype that showcase the communication between them in the system level and the information level.
- *Collaboration:* A major outcome of the first two years is the definition of the INSIGHT architecture, the INSIGHT communication mechanism, and the evaluation mechanisms through the use of concrete use cases. This work was done by a collaboration of all partners, leading to a common understanding of the use cases from end-users, algorithmic and system perspective. This process provided guidance for the architectural design phase through the mapping of the framework to the use cases, and led to the development of a proof of concept of the INSIGHT system that focuses in the city-level use case. In addition, this process has highlighted the extensive collaboration between the INSIGHT partners.
- *Dissemination Efforts:* In the second year of the project, the consortium provided 30 Oral Presentations to Non-Technical audience, 24 Research Papers, 27 Invited Talks, Keynotes, Tutorials on topics related to INSIGHT and 5 articles to websites, magazines or newspapers. In addition, INSIGHT very proudly received 4 awards! This effort adds up to the activities of the first year (15 scientific publications and several talks in scientific and technical conferences and forums). Finally, we have successfully run a workshop on the topic "Mining Urban Data". The benefit from this event was strong project and research interactions as well as dissemination of INSIGHT.
- *Data Collection and Availability:* During the first two years of the project, INSIGHT developed the required technical frameworks and interfaces in order to provide data access to a broader community. More specifically SCATS data (traffic volume information) and Bus data (GPS signals transmitted from buses moving around the city) are made available at Dublinked[1], one of the largest data repositories of urban information. We strongly believe that the data availability will significantly benefit the research

[1] `www.dublinked.com`

community and will encourage the industry to develop application with social impact.

1.4 Expected final results and their potential impact and use

The expected final exploitable results of the project include:

1. The first exploitation direction is directly linked to the city-wide and the nation-wide monitoring scenarios. Fraunhofer aims to transform the INSIGHT research prototype in a fully working system, jointly with BBK. Simultaneously, centers at the European level (and beyond) will be targeted. We plan to explore the use of the system to other cities as well, taking advantage of the similarity of the traffic management problem, while building an infrastructure with the flexibility to address differences. IBM plans to use the Inteligent Sensor Agents in its smart transportation software solutions. IBM also plans to use the distributed optimization algorithm developed in the project in its Internet of Things solutions. IBM will rollout the INSIGHT project in other test cities on the basis of a commercial project. IBM is leveraging its world-wide Sales Network to position INSIGHT in its "Smarter Cities" Initiative and the developed components as plug-ins in its products around Traffic Prediction and Stream Computing.

2. The techniques developed in the project will advance the state-of-the-art in monitoring individual data sources as well, and in exploitation we will build stand-alone modules for monitoring one or more data sources. These will provide reinforced ability for a wide range of innovators to tap data infrastructures and to add value beyond the original purpose of the data through data analysis.

Key expected results for the following one year-period of the INSIGHT project include:

- Continuation and finalization of the integration.
- Work on validation by working closely with our user partners and by running the designed nation-wide and city-wide experimental scenarios.
- Dissemination of research results and encouragement of the community to use the open data we provided.
- Focus on the exploitation and adoption of the developed technologies.

1.5 Public Website and Additional Resources

INSIGHT's website can be accessed from the following URL: `http://www.insight-ict.eu`.