

Smart cities management by integrating sensors, models and user generated contents

A. Candelieri¹, F. Archetti^{1,2}, I. Giordani², G. Arosio²
& R. Sormani¹

¹*Department of Computer Science, Systems and Communication,
University of Milano-Bicocca, Milan, Italy*

²*Consorzio Milano Ricerche, Milan, Italy*

Abstract

This paper presents a further innovation of the collaborative decision support network LENVIS, developed in the homonymous European FP7 project (Localized ENVironmental Services for all, www.lenvis.eu). The system is aimed at supporting water and air quality management and assessing the impact on health computed as emergency hospital admissions. LENVIS has been designed according to the Internet of Services paradigm: several web services, such as sensors data integrators, geographical information systems, predictive and simulation models, and data analytics tools, have been suitably combined to provide complex applications to different stakeholders. LENVIS has been already integrated with popular social networking platforms as Facebook, Twitter and LinkedIn ensuring that its information can be assessed seamlessly during a user's regular internet browsing activities and that the user generated contents can be processed by LENVIS analytics. The experiences referred to in this paper are from Milan, a city in Northern Italy with severe pollution problems, mostly due to Particulate Matters (PM), and three cities in Portugal, where the quality of bathing water affects the way in which population uses water resources.

Keywords: smart cities, environment and healthcare management, decision support network, Markov models, social innovation, social networks.

1 Introduction

This paper outlines the evolution of LENVIS from the prototype of the European project towards a design of a framework for smart cities management. This



evolution is driven by the growing importance and diffusion of user's generated contents on environmental and health issues and the need to integrate this information with the readings of sensor networks and the results of simulation models. The authors are particularly indebted to a paper (Sivastrava and Vakali [1]) in which the relevance of "narrative aware design framework" in the design of smart urban environment is pointed out. According to the LENVIS experience and the results obtained in the project, here we think to enlarge the vision of Srivasta and Vakali, based on the combination of information from sensors data and users' storytelling, by introducing also predictive and simulation models in the smart cities management.

Therefore the evolution of LENVIS, proposed in this paper, takes into account not only technological innovations, moving towards an Internet of Things paradigm, but goes well beyond that towards an environment which is both sensors- and models- driven and socially aware. In Figure 1, we propose an analogy about our vision of an effective and efficient smart cities management: smart city paradigm has to be essentially based on information and knowledge, and three are the pillars sustaining it.

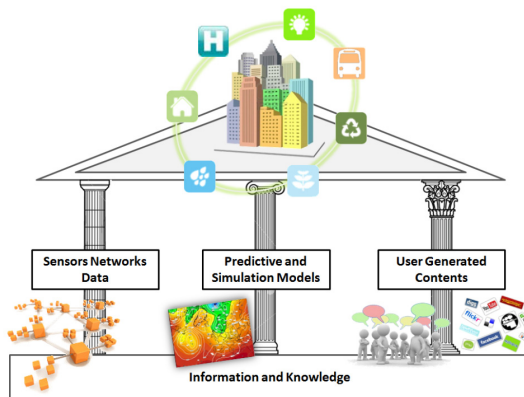


Figure 1: The Smart City Temple has (at least) three pillars leaning on information and knowledge and enabling effective and efficient smart cities management. (Doric, Ionic and Corinthian styles are not casual; they provide an indication about the appearance of each pillar during the time).

The main challenges of the LENVIS project have been to develop:

- cross-platform techniques and an uniform data model to gather and fuse time series from several heterogeneous data sources, in particular those acquired by sensors, provided by simulation models or generated by the users (in well structured form).
- mechanisms enabling two way communication and collaboration among different stakeholders, such as local environment protection agencies, healthcare providers, general practitioners, and citizens in general;

- technical solutions for composing services according to the needs of each LENVIS user;
- strategies to guarantee individual privacy in sharing data and information.

Our proposal outlines a new design framework for the smart city which considers quantitative sensor acquired data as well as qualitative human generated data through web platforms.

The applications in this framework can already benefit from the combination of sensors and data from users on the move. Also residents could use, in their mobility related decisions, socially aware services which place emphasis on the collection of information that generates real time adaptive recommendations for residents to improve navigation within the city and assist in the selection of appropriate routes based on various parameter (pollutant concentrations, noise level, parking, public transport routes). According to this possible innovation of LENVIS, some further steps have to be performed:

- design and develop methodologies and services for cross-media contents analysis, such as sentiment and opinion mining strategies (Pang and Lee [2]) for monitoring perceived quality on environmental issues (e.g., air and water quality) and city services (e.g., transportation, healthcare, water distribution, waste management);
- design and develop modules able to integrate information provided by the three different “pillars” and process more complex, complete and reliable decision support.

The main issues already addressed by the LENVIS project have been air quality, water quality and assessment/prediction of pollution impact on citizens’ health/living (Popescu *et al.* [3]). LENVIS is able to provide warnings and decision support functionalities for each kind of stakeholder. The three main groups of LENVIS users are:

- local health and environment authorities, healthcare providers;
- general practitioners (GPs) and, in particular, consultants, pneumologists and allergologists;
- user groups and European citizens in general.

This paper is focused on the two main use cases:

- Air quality in Milano. In particular, the system is able to support:
 - local authorities in monitoring pollution level and evaluating effectiveness of environmental policies;
 - healthcare providers, both hospital and general practitioners, in planning effective resources allocation and improve the diagnosis of pollution related conditions (respiratory allergies), through a forecasting service;
 - citizens in planning outdoor activities or receiving selective alerts/warnings toward groups at risk of the population.
- Water quality in Portugal. In particular, the system is able to provide information about water quality to surfing groups who are already using social networks to meet on line.



The rest of the paper is organized as follows: in Section 2 a general description of LENVIS is reported, summarizing its architecture and services already developed. Section 3 outlines the Portuguese and the Italian use cases. Section 4 outlines the design framework of the next generation LENVIS for smart city management. Finally, Section 5 provides conclusions and discussions.

2 LENVIS as a network of services

Information Networks, composed by several services (e.g., GIS mapping, data warehousing and Business Intelligence, simulation models for environment and health, etc) and developed by different companies or research institutes, may represent an effective strategy to achieve relevant results in the environmental health domain (Sonntag [4]), such as:

- describing health trends in response to environmental conditions,
- improving quantitative risk assessment by reducing uncertainties in the source-to-adverse outcome continuum,
- identifying more reliable methods and predictive models for quantitative risk assessment.

Indeed heterogeneity, both at data and information level, is particularly strong in the field of environment and healthcare management:

- data are stored in different formats;
- tools to analyse data are very different – and stakeholders usually want to use their own;
- different decision makers have different and typically conflicting objectives;
- the societal need, particularly relevant in environmental issues, is strongly dependent on risk perception, which may be really different among stakeholders.

According to the aforementioned aspects, the most suitable approach consists of a collaborative network able to integrate and provide, as services, different systems and functionalities and to combine them in order to provide more complex and stakeholder-specific decision support.

2.1 LENVIS architecture

LENVIS consists of a wide set of services and data sources, acquired through sensors and related to contents generated by the users, all distributed on different locations, and fully agrees with the collaborative network vision. In order to guarantee high integration and interoperability, LENVIS has been designed and developed according to the Internet of Services paradigm, in particular it is a collection of several web services that can be suitably combined for providing users with more complex and advanced decision support functionalities.

In the following Figure 2, a schematic representation of the network architecture of LENVIS, the data sources and the services involved is proposed.



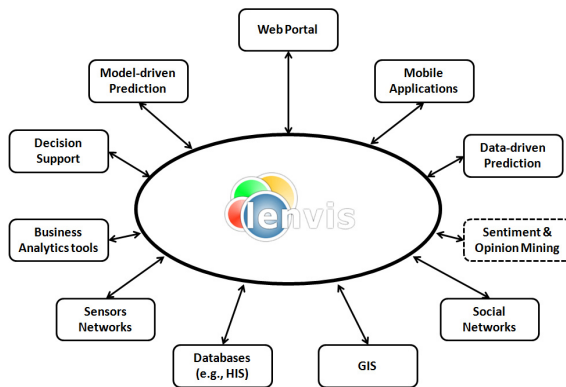


Figure 2: Data sources and services in LENVIS (Sentiment and Opinion Mining is a functionalities set that will be developed in the Future LENVIS; this shall require updating some other modules, such as the decision support).

Sensors Networks, Databases and Social Networks are the basic modules devoted to the data/contents acquisition and storage. Through the unique data-model defined within LENVIS the system is able to integrate sources that use different data-formats. The GIS (Geographical Information System) modules are aimed at geo-referring data and information. Business Analytics tools, Data-driven and Model-driven Predictions and Decision Support are the modules devoted to provide the most suitable decision support functionalities to each stakeholder or LENVIS user in general. The Web Portal, the Mobile Applications, and the Social Networks as well, are responsible for the interaction mechanisms between LENVIS users and the system. Finally, the Sentiment and Opinion Mining module – which will be developed in the “Future LENVIS” – is aimed at providing cross-media contents analytical functionalities in order to mine relevant information from the users’ storytelling and complete, subsequently, the knowledge coming from structured data (sensors and models).

Developing a cross-platform set of web services has been a crucial challenge of the project: LENVIS partners have defined the interfaces between the web services and designed suitable solutions to guarantee interoperability among them and usability by other platforms developed under different technologies.

LENVIS web services follow the interoperability requirements of the INSPIRE Directive (2007/2/EC) - *Infrastructure for Spatial Information in the European Community*-, in order to avoid multiple collection and maintenance of geographic data and to encourage sharing of data between different users and application.

LENVIS fits into the European Open Data experience where data sets available through government agencies can be accessed and used, according to the Open Innovation paradigm, by developers, students, and other innovators to create innovative valuable ideas and business opportunities that improve

ordinary citizens' lives in a variety of way (Hoyer [5]). The project had faced technical problems, due to the heterogeneity in data formats and models, as well as the integration of third parties software and services. Where possible, adopting standards and defining a meta-model for data have permitted us to successfully deal with these issues.

2.2 Environmental predictive models

LENVIS integrates a number of computational services dedicated to now-casting and forecasting. Some of these models are mechanistic and based on modelling the emissions and fine-tuning the parameters through the concentration levels provided by the sensors network. In the following Figure 3, a map of the PM₁₀ levels, as predicted by the emissions simulation models, is depicted for the city of Milan.

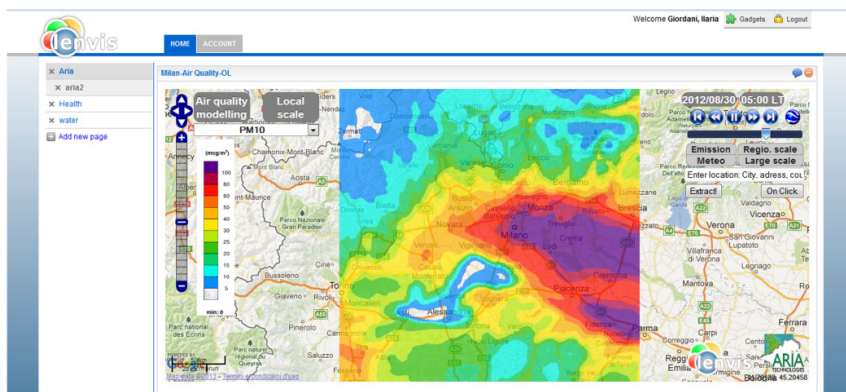


Figure 3: A frame of the prediction of PM₁₀ concentration in the city of Milan.

Some other models have been obtained through statistical learning; their output is basically a sequence of numerical, categorical or ordinal values (increase, decrease or persistence of pollutants concentrations rather than the trends of emergency hospital admissions, with their joint probability). In particular, inference engines based on Markov models, which can be easily adapted to represent the evolution of pollutants' concentrations, have been developed within the LENVIS project (Candelieri *et al.* [6]).

The modelling issue is very complex due mostly to the uncertainties which characterize it and can be related to:

- data whose quality and availability are in some instances rather poor, coming from different sources often annotated by heterogeneous users (e.g. clinical data);
- decision workflow characterized by the interaction of decision makers, different, often without a precise hierarchical structure, and having conflicting objectives;

- models applied for simulation and forecasting which must allow an intuitive characterization of uncertainty and automatic scenario generation even at the price of losing accuracy with respect to highly specialized ones.

2.3 Environment-related decision support for health risk assessment

A report of the World Health Organization (WHO [7]) and a recent result of (Sun *et al.* [8]), indicate a 0.11% increase in mortality due to cardiovascular or respiratory diseases, for a $10\mu\text{g}/\text{m}^3$ increase in the daily average of PM_{10} .

With respect to the correlation between pollution and emergency hospital admissions, at the European level, (Blangiardo *et al.* [9]) reported that a $10\mu\text{g}/\text{m}^3$ daily increase in PM_{10} concentration produces a 0.9%, 0.8% and 1.3% increase in admissions, in London, due to cardiac diseases, myocardial infarction and respiratory disease, respectively. Although traditional studies are useful in better understanding the epidemiological aspects linking environment and health risk, LENVIS has been designed to overcome their limits about the possibility to provide forecasts in the very short term (e.g., 1 to 5 days). A specific web service (namely HIDSS: Health Impact Decision Support System) has been designed, developed and deployed within LENVIS, aiming at analysing relations between pollutant concentrations and health impact and providing health indicators (i.e., emergency hospital admissions trend in the short term) as an integral part of the environment and healthcare management (Candelieri *et al.* [10]).

In the following Figure 4 a picture of the first mechanism of decision support is reported, that is a gadget devoted to visualize emergency hospital admissions and pollutant concentrations (i.e., time series data), contemporary. This service allows both environmental and healthcare managers to perform a preliminary exploratory analysis about possible correlations, causalities and cyclic/periodic behaviours, in specific time windows and at a given location (i.e., pollutant level near a monitoring station, or in the entire city as median value of concentrations for that pollutant).

LENVIS is fully compliant with the requirements of the most up-to-date *European Directive on ambient air quality and cleaner air for Europe* [11] which mandates sustained and structural modifications to the management of the environmental information and suggests, for the main pollutants, different thresholds to trigger alerts and alarms, both general and for specific frail segments of population (e.g., children, elderly, and individuals affected with respiratory diseases).

The decision support provided by the system goes well beyond a simple data aggregation-and-visualization task. As aforementioned, its analytical core consists of Markov-based inference strategies used to provide the following set of specific decision support services (Candelieri *et al.* [6, 10]):

- supporting environment related authorities to assess, more objectively, the air quality and evaluate the effectiveness of emissions limitation policies;
- allowing hospital managers to plan resources in advance according to admissions trend forecast in the short term, enabling a more rational resources allocation;

- assisting General Practitioners to predict adverse events and prevent hospital admissions, reducing costs while improving patient quality of life.

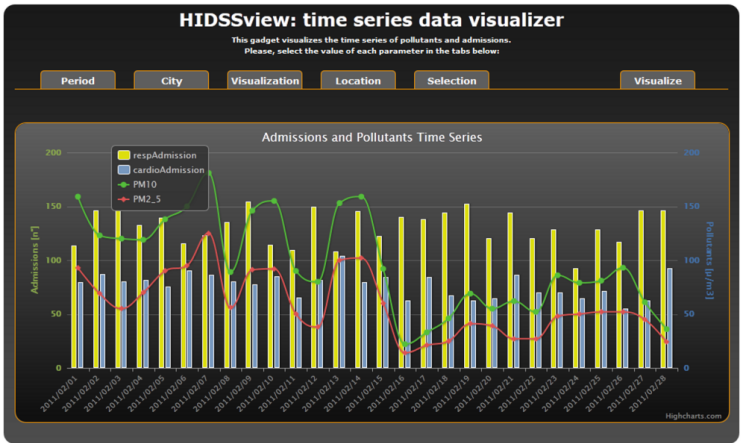


Figure 4: Time series data visualizer (HIDSS gadget), as shown in the LENVIS portal. Daily hospital admissions, due to respiratory events, and PM₁₀ and PM_{2.5} concentrations in February 2011.

3 LENVIS case studies

Two different case studies of the LENVIS project are considered in this paper: Milan (Italy), for the air quality and health risk assessment, and three Portuguese coastal cities, for water quality and social involvement of the surfers' community.

3.1 Case study #1: Milan (Italy)

For this pilot two different datasets have been used: daily concentrations of several pollutants and emergency hospital admissions due to cardiovascular and respiratory adverse events. Emergency hospital admissions have been acquired as “packets” of daily observations and gathered on demand (i.e., not automatically). The entire dataset has been split in training and test set to perform reliability evaluation of the predictions. Local environmental authorities, in Milan, typically publish daily average data. According to our vision this rate will have to become hourly as required to correlate emissions with activity levels (e.g. traffic or some industrial carbon sensitive activities as well as meteorological data).



3.2 Case study #2: Carcavelos, Torre and Santo Amaro de Oeiras (Portugal)

The Portuguese case study addresses bathing water quality and related health issues on three bathing waters nearby Lisbon: Carcavelos, Torre and Santo Amaro de Oeiras, located west of Lisbon in a densely urbanised region. Bathing water quality is under threat due to sporadic pollution events in the small streams discharging along the coast that might affect bathing waters quality and impair bathers' health. LENVIS makes available information about water forecasts, wave forecasts, water quality and quantity forecasts, data from meteorological and hydrometric automatic stations. In particular, multi-scale atmospheric and ocean models are used in this case study, going from global to local scale, and different processes are coupled (e.g., atmospheric forcing, wind wave, hydrodynamics and water quality processes).

During the last phase of the LENVIS project, mechanisms devoted to collected structured feedbacks from bathers and surfers (social participations) have been developed in order to improve forecasts through the local information provided by the "human sensors"; the following Figure 5 shows this loop.

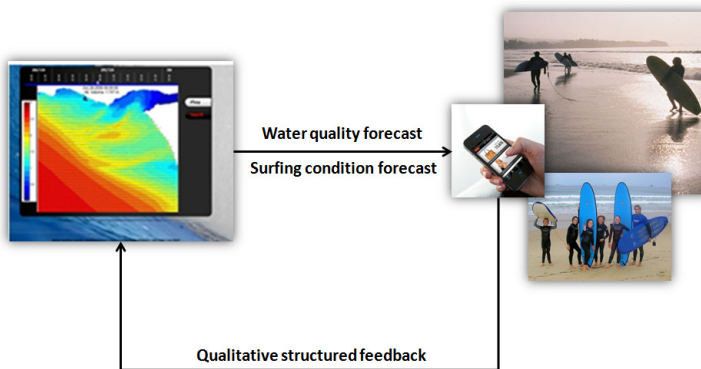


Figure 5: LENVIS' strategy to enforce water quality and surfing condition forecasts through the structured feedbacks provided by the surfers community

4 The future LENVIS

The fusion of sensor readings, simulation models and user generated contents makes for an improved performance of the overall LENVIS system. In this way the difference in information needs and the perceived value of environmental and health information are brought into the loop. Specific instances of the value of social networks are:

- highly localized urban conditions which would require micro-simulations or sensor readings solutions beyond what is technologically or economically feasible can be captured by the networking activity of citizens in that area;

- emergencies captured by peak activity recorded by the sensors or predicted by simulation models are “fortified”/“explained” by simultaneous social burst (heavy tagging);
- the user generated content might also be useful to verify whether some sensors are malfunctioning as their data can be refined or even contradicted by social storytelling.

Therefore, analysis of the user generated contents may be successful in reducing uncertainty which may arise from many sources:

- Process (e.g. uncertain on travel times).
- Data (GPS based location, natural language ambiguity in social data).
- Model (simulation models).

With the right analytical methods these data can be used to predict when special conditions arise (e.g. in atmospheric pollution). This information may assist authorities in taking short term (e.g. inform/alert residents and vulnerable social groups) or long term actions (understanding the progress of the phenomenon in order to improve the city’s conditions).

The so called “urban social pulse” can be gleaned through applications like Flickr, Facebook, Twitter and location based services such as 4Squares and Google Latitude. For instance residents might express throat disturbances and this might serve to override sensor data or simulation generated data such as PM₁₀ level which might not be enough to trigger an alarm (spatial distribution of sensors and simulation grid too large to allow proper monitoring of micro-areas).

Similarly to the recent paper of Srivastava and Vakali [1], we propose a three levels schema for the Future LENVIS, but revising it according to our experience with models and data driven decision support. In the following Figure 6, the proposed schema is depicted. In particular, the integration of methodologies and services for cross-media contents analytics is the key point for the innovation: it will enable a deeper understanding about the users’ urban experiences, than structured feedbacks, improving the impact of social participation in the decision making activities related to smart cities management. According to the new opportunities, the current Decision Support of LENVIS needs to be updated, evolving towards an integrated one able to provide suggestions according to all the sources of information and knowledge.

Finally, with respect to the different sources that will be integrated, the authors will take remarkably into account the characterization recently proposed by IBM for Big Data (Zikopoulos *et al.* [12]). Data in particular can be characterized, beyond quality, according to three different dimensions:

- **Variety:** while environmental and clinical data are structured, user generated contents may include a significant audio-video component and unstructured text which may contain relevant information.
- **Velocity:** it ranges from very low, e.g. hospital admissions data, to very high, e.g. sensors data with high acquisition frequency (10 Hz).
- **Volume:** the main contributions are sensors data and user generated contents. For the first, the volume is predictable according to the structure of the deployed network and the acquisition rate while the second usually depends on “social waves” related to specific events or news.

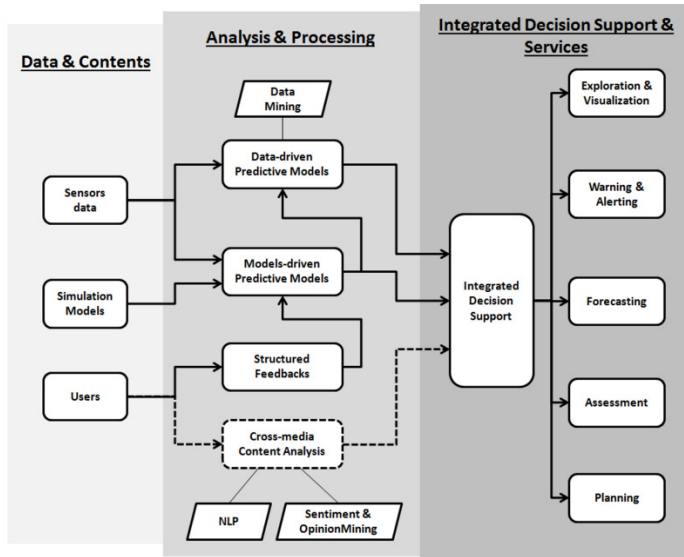


Figure 6: A three-level schema of the Future LENVIS system. Dotted lines refer to the methodologies enabling a deeper analysis of the users' generated contents.

5 Conclusions and discussion

The proposed innovation of the collaborative decision support network LENVIS has relevant societal implications that can be understood bearing in mind that LENVIS puts, more and more, the interaction between public sector actors (institutional care givers, GPs and local authorities) as well as citizens at the very centre of the innovation process. This is obtained correlating the sensors data and simulation models based decisions with perceptions as relayed by social media of the environmental condition in a given area at same time window.

Thus the user is put at the very centre of the innovation process allowing a service pull model of the environmental innovation. Innovation becomes a co-creative collaborative process between service providers and users.

More interesting, LENVIS enables not only a traditional top-down but also a bottom-up approach. Different environmental legislations and risk perception – as for the reach diversity in Europe – are regarded as a valuable framework for the development and deployment of LENVIS which offers to this purpose an open access and open functional platform architecture enabling 2-way communications between the user and service provider via crowd-sourcing tools to capture valuable ideas produced by the communities.

Finally, LENVIS fits into the European Open Data experience where data sets available through government agencies have been used by developers, students, and other innovators to create a number of programs and applications that improve ordinary citizens' lives in a variety of ways. These include community

health maps and dashboards, health data integrated with web search engines, and tools for citizens to find the best healthcare providers. The environmental protection agency EPA developed AIRNow.gov to provide the public with easy access to air quality information wherever they live.

References

- [1] Srivastava, L., Vakali, A., Towards a narrative aware design framework for smart urban environment, F. Álvarez *et al.* (Eds.): FIA 2012, LNCS 7281, 166–177, 2012.
- [2] Pang, B., Lee, L., Opinion Mining and Sentiment Analysis, *Foundations and Trends in Information Retrieval*, 2(1, 2): 1–135, 2008.
- [3] Popescu, I., Archetti, F., van Andel, S.C., Giordani, I., “Lenvis: a user centric, web services based system to retrieve, analyze and deliver environmental and health information. *Environmental Engineering and Management Journal*, 11(5): 889–897, 2012.
- [4] Sonntag, W., Health & Ecoinformatics, U.S. Environmental Protection Agency, 2004. http://www.cdc.gov/nceh/tracking/conf04/pdfs/wed/ses1a/b_sonntag.pdf
- [5] Hoyer, B.M, Unlocking the digital future through open innovation an intellectual capital approach – A critical analysis of open innovation as structural capital, European Commission, Information Society and Media, 2011.
- [6] Candelieri, A., Giordani, I., Testa, P., Arosio, G., Archetti, F., A Markov-based model to forecast emergency hospital admissions due to air pollution: the LENVIS project approach, *Environmental Engineering and Management Journal*, 11(5): 999–1008, 2012.
- [7] World Health Organization (WHO) Europe, Health Aspects of Air Pollution – Results from the WHO Project “Systematic Review of Health Aspects of Air Pollution in Europe”, 2005.
- [8] Sun, Q., Hong, X., Wold, L.E., Cardiovascular Effects of Ambient Particulate Air Pollution Exposure, *Circulation*, 121: 2755–2765, 2010.
- [9] Blangiardo, M., Richardson, S., Gulliver, J., Hansell, A., A Bayesian analysis of the impact of air pollution episodes on cardio-respiratory emergency hospital admissions in the Greater London area, *Stat Meth Med Res*, 20(1): 69–80, 2011.
- [10] Candelieri, A., Archetti, F., Giordani, I., Arosio, G., Markov-based approaches to support policies makers in environment and healthcare, In *Proceedings of International Environmental Modeling and System Society (iEMSs) Conference*, 2012.
- [11] Directive 2008/50/EC of the European Parliament and of the Council, of 21 May 2008, on Ambient Air Quality and Cleaner Air for Europe, 2008.
- [12] Zikopoulos, P.C., Eaton, C., Deroos, D., Deutsch, T., Lapis, G., *Understanding Big Data: Analytics for Enterprise Class Handoop and Streaming Data*, McGraw-Hill, 2012.

