

# A Pilot Web-based System for Environmental Health Impact Assessment of Air Pollution

Marius Daniel BONTOS, Daniela VASILIU

University Politehnica of Bucharest,  
Splaiul Independenței 313, Bucharest, 060042, Romania,  
bontosmarius@gmail.com, vasilu1958@gmail.com

**Abstract:** This paper presents an on-line pilot system for environmental health impact assessment of air pollution with the support of georeferenced relational databases collection for disseminating information to a wide range of users with the view to enabling them to make appropriate decisions. This tool integrates modern technologies for processing, computing, and information dissemination such as automatic acquisition, web crawling, relational databases and Geographic Information System (GIS).

**Keywords:** automatic web data acquisition, relational databases, health impact assessment, Geographic Information System (GIS), database driven website.

## 1. Introduction

The quality of the environment is the result of interaction of natural elements - earth, air, water, climate, biosphere and elements resulting from human activity. In turn, the environment influences the conditions of existence of society, its development possibilities and the quality of human life. Consequently, environmental and environmental health protection is a public priority in Romania and on a global scale as well, aiming to achieve a clean and healthy environment and preserving natural resources, in accordance with the requirements of sustainable economic and social development. Achieving these goals requires better education and awareness amid global informatization of the society (WHO, 2011; EHJ, 2011).

Several studies (Ezzati, 2004; WHO, 2009) made on the impact of environmental pollutants on health have been performed. These show that about 25-30% of all diseases in industrialized countries are attributed to environmental factors. They mainly affect the children (in their case the proportion of environmental mortality being 37% for 0-4 years-group and 36% for 0-14 years (WHO, 2006) and vulnerable groups like poor population and future mothers. This percentage is likely to have an upward trend if nothing is done to reduce environmental pressures on human health.

Information technologies are now the support of all components for basic and applied scientific research in environmental protection. They are utilized for the numerical simulation of complex processes, for monitoring and

controlling the experiments, as well as in all applications for processing and transmitting specific information. (Vasilu, 2007). At the same time data mining technology enable finding useful information within immense amounts of data ( Peng et al ,2008).

At the same time decision support systems (DSS) are meant to help the managers, planners and others knowledge workers to evaluate the complex situations with the view to choosing the most adequate causes of action in various application domains (Filip, 2008, 2002). In the case of complex application systems, the process of designing and building information system can be viewed as a series of decision making activities (Filip, 2011).

In this respect, the main objective of this paper is to present ISEH ("Integrated Information System for Environment and Health"), an on-line pilot system meant for the environmental health impact assessment of air pollution with the support of geo-referenced relational databases collection for disseminating information to a wide range of users. This tool aims to provide timely and adequate high quality data and knowledge in order to inform the public and support decision making in environmental and public health at national, regional and local level.

For applications with multiple and diverse sources for large amounts of data, like the purposed one, there are on the market products like "Data Warehouse / Business Intelligence". Their disadvantage is the high cost and functional opacity. In addition, a solution made directly from the basic components can take advantage not only of a low cost but also of a good adaptation of the application.

This paper is organized as follows. Section 2 presents the architecture of the proposed environmental health impact evaluation system. Each component is presented in the following sections. Section 3 describes the data acquisition module with its two components: meteorological data acquisition based on ADCON system (ADCON, 2011) and pollution data acquisition with web crawling / web spidering technology. Section 4 refers to the design, creation and management of specific environment and health databases required for system implementation. Section 5 presents the health impact assessment of air pollution module and the case study. Section 6 presents the implementation of GIS technology for study areas selection and for the creation of thematic maps. Section 7 describes the integration of all aspects above mentioned on a common web platform which aims at disseminating the information collected and also of the calculated indicators in an accessible format. Last section concludes the paper.

## 2. System Architecture

The possibility of accessing in a timely manner the relevant information is an essential element in the attempt of improving efficiency of environmental assessment practices and strengthening the capacity to develop and manage specific systems and processes.

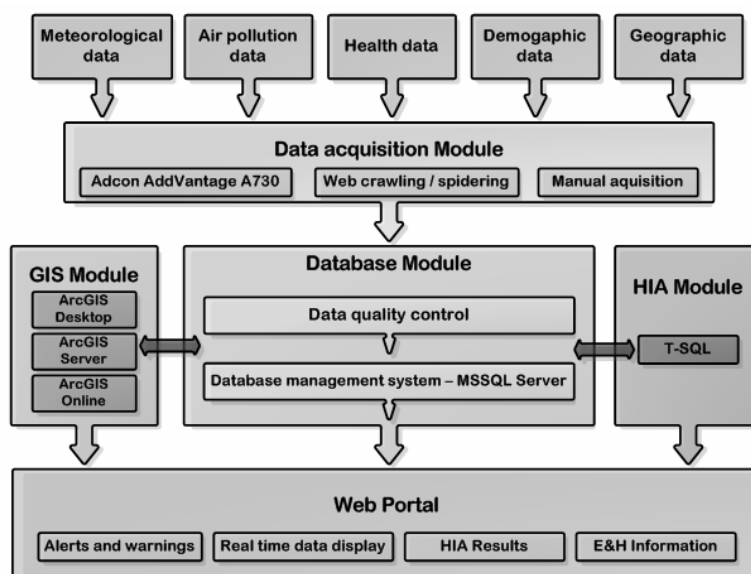
Figure 1 shows the diagram of the environmental health impact assessment system

proposed in this paper. It is a large scale complex multilayer system (Filip, Leiviska, 2009) and it is based on modern technologies for environmental monitoring, data analysis, data processing and dissemination of information using the World Wide Web network (Vasiliu, 2007). The system contains four interconnected modules (acquisition module, database module, GIS module, health impact assessment module) that provide information to a display area which is implemented with WWW technology. They will be described in the following sections of the paper.

## 3. Data Acquisition Module

Monitoring is the central activity of IISEH. To achieve health impact assessment for a chosen area large series of data are required from various fields such as meteorological data, data on air pollution and major pollutants, demographics information and health indicators. These can be collected either by the systems itself or acquired from the main official data providers from environmental and health fields.

For the system described in this paper meteorological data are acquired using an automatic acquisition system, produced by ADCON Telemetry (ADCON, 2011) and pollution data were acquired from the National Network of Air Quality Monitoring (RNMCA, 2011).



**Figure 1.** The architecture of IISEH (“Integrated Information System for Environment and Health”)

With the development of Internet and automated acquisition systems, ever more data providers publish information on their web sites. The data published on those sites can be automatically acquired using the web crawling / spidering technology. One of the main advantages of this technology is the lack of security problems because it does not require a direct connection to acquisition system or website databases.

Due to the large number of data and web pages contained on their web sites, the extraction of the useful information can be difficult and may require the development of Web crawlers or Web spiders programs to facilitate the access to desired data available online (Castillo, 2004).

The web crawling acquisition module developed in this paper consists of three major components which perform data acquisition, file parsing and data saving. They are described in the sequel.

### 3.1 Achievement of the acquisition parameter files

A focused crawling algorithm (Chung et al., 2003) was chosen for air pollution data acquisition. To acquire the available information in a consistent way, it was necessary to create acquisition files for each data collection site (DCS) and for each parameter.

An example of acquisition file for one DCS and one parameter can be seen below (Figure 2).

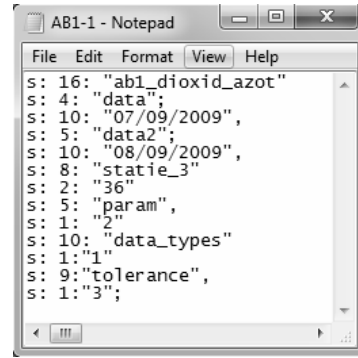


Figure 2. Example of acquisition file

Where:

- s: 16: "ab1\_dioxid\_azot" - represents the file name;
- s: 4: "data"; s: 10: "07/09/2009", s: 5: "data2"; s: 10: "08/09/2009"- is the first acquisition period;
- s: 8: "statie\_3" s: 2: "36", - represents the code of the measure station;
- s: 5: "param", s: 1: "2" - is the measured parameter code;
- s: 10: "data\_types" s: 1: "1" - represents the type of data;
- s: 9: "tolerance", s: 1: "3" - is the number of decimal places for the acquired data.

### 3.2 Parsing the files

The parsing script (lib\_parser.php) is used to extract from acquisition parameter files described above, the necessary data (descriptors) for sending data requests to the target website. A code sequence from the parsing program is shown in Figure 3.

```
<?php
set_time_limit(3600*10);error_reporting(E_ERROR&~E_NOTICE);
function parser_toFieldName($string){
.....
function parser_toFieldValue($string){
.....
function parser_parse($fileName,$level=1){
$f=fopen($fileName,"r");$input=fread($f,filesize($fileName));fclose($f);$input=str_replace("<br>","|",$input);
$input=str_replace("<trclass=\"tbody\" align=\"center\">","",$input); $input=str_replace
('colspan="1"',"$input");$input=str_replace("\t","",$input); $input=str_replace('class="tbody" align="center"',"$input");
$input=str_replace("\r\n","",$input); $input=str_replace("\r","",$input);
$input=str_replace("\n","",$input);$doc = new DOMDocument();$err=error_reporting();error_reporting (E_ERROR);
$doc = new DOMDocument("1.0","utf-8");$doc->loadHTML($input); error_reporting($err);
$output=array();$output["header"]=array();$output["items"]=array();
$tables=$doc->getElementsByTagName("table");$l=0;
for($i=0;$i<$level;$i++){$table=$tables->item($i);$lstl=$table->childNodes;$header=array();
    $l=0;$indicator="";$statie="" $data_intrare="";
for($j=0;$j<$lstl->length;$j++){$lstj=$lstl->item($j)->childNodes;$l++;$dataLine=array();$startLine=1;
for($k=0;$k<$lstj->length;$k++){$obj=$lstj->item($k);$value=$obj->nodeValue;
    if($l==$startLine){$lValue=trim(parser_toFieldName($value));$index=$k;
    if ($index==0){$t=explode("_",$lValue);$data_intrare=$t[2]."-".$t[1]."-".$t[0];
```

```

        $iValue="data_colectare";}
    if ($iValue!= ""){$header[$k]=$iValue;}}elseif($i!=$startLine+1){
    if ($indicator!= "") continue;$t=explode(" | ",parser_toFieldValue($value));$stati=$t[0];
        $indicator=$t[1];}elseif($i>$startLine+1){
    if (array_key_exists($k,$header)){
    if ($k==0){list($h,$m)=explode(":",$value);
    if ($h==24){$h=23;$m=59;}$value=$data_intrare. "" . $h. ":" . $m. ":00";}
    $key=$header[$k];$dataLine[$key]=parser_toFieldValue($value);}$output["items"][]=$dataLine;}$output["header"]
    =$header;return $output;}
    ....
    function parser_addtodb
    .... ?>

```

**Figure 3.** The parsing script “lib\_parser.php”

### 3.3 Acquisition and data saving

Acquisition and data saving is performed automatically by a program (post.php) that loads the descriptors and sends them as http request to the target server. The next step consists in the capturing of the content. This is done by extracting elements from web pages saved locally. The data that will be extracted is in a table format which is easily recognized by the browser program.

In addition to serialized data, the script contains the login information to the target website and to local databases where the information collected will be stored. All these procedures are presented below, broadly (Figure 4):

## 4. Structuring, Creating and Managing Specific Databases for Environment and Health (Database Module)

Based on the relational database model and taking into account the inputs from the acquisition system, the containing databases and tables have been properly structured to meet the system future needs.

In case of the proposed theme, application objectives are the dissemination of collected information, the creation of environment - health connections by calculating specific indicators and their dissemination in a more

```

//The procedure for the inclusion of the parsing data from the scrip files:
Include_once ("lib_parser.php")
//The procedure for sending the request data to the web address and local saving:
Do_post_request2function ($url, $ data)
//The procedure for connecting to MSSQL database:
$link = mssql_connect('PC\SQLEXPRESS','username','password');
if (!$link || !mssql_selectdb ('MediuDatabase', $link)) {die('Unable to connect or select database!);}
//The procedure for connecting to MySQL database:
mysql_connect("localhost","root","");
mysql_select_db("mediu");
//The procedure for saving the downloaded information into MSSQL database:
parser_addtodb_mysql ($data, "DATE_POLUARE_AER" ,...)
//The procedure for saving the downloaded information into MSSQL database:
parser_addtodb_mysql ($data, "DATA_NEW" ,...)

```

**Figure 4.** SQL Procedures for acquisition and data saving

Where: *lib\_parser.php* - is the script designed to parse the acquisition files; *\$ url =*”address”;  
*\$ data* - is the acquisition parameters.

accessible format. In this case, the existence of a database provides the background information in a quality structure for an easier reporting of the results. The database should allow the obtaining of detailed information and the calculation and presentation of synthetic indicators.

The Integrated Information System for Environment and Health contains four main databases (“AERData”, “SANATATEData”, “GISData” and “WORKData”), needed to assess the health impact of air pollution. The diagram of the “AERData” database is presented in the Figure 5:

In this context, in 1997 InVS (“France Sanitary Surveillance Institute”) implemented a model of epidemiological surveillance in nine French cities in order to measure and monitor the relationship between short-term urban pollution and human health (Quenel, 1997; InVS, 2002). The model has been continuously improved. It was expanded

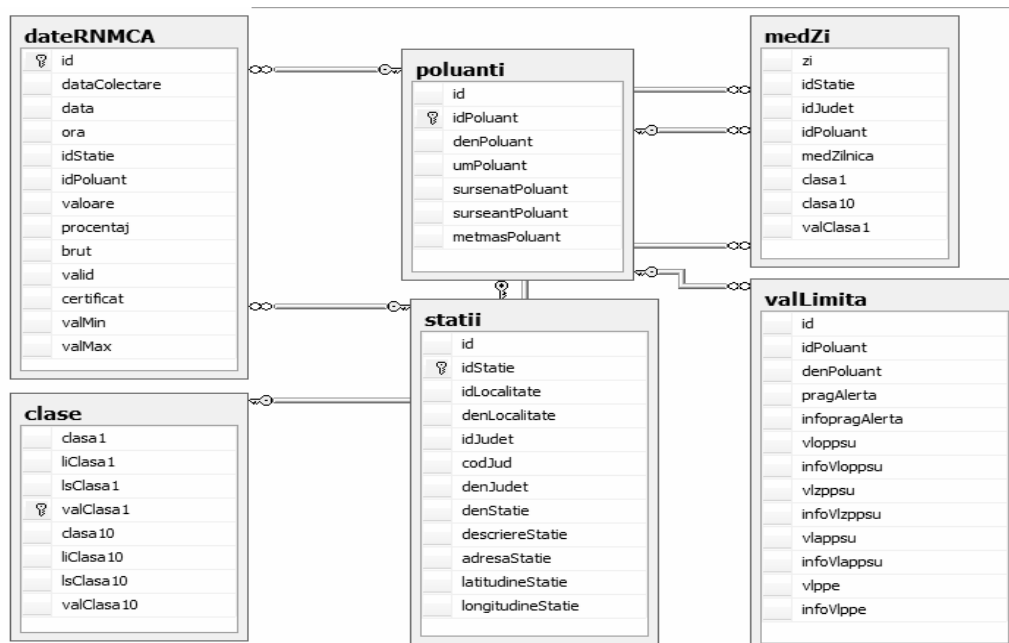


Figure 5. “AERData” database diagram

## 5. Health Impact Assessment of Urban Air Pollution (Module)

Currently, air pollution effects on human health are determined more accurately, especially with preference to short-term effects that can occur, due to the large number of human experimental (Schlesinger, 1995) and epidemiological studies (Dockery, Pope, 1994; Schwartz, 1994; ATS, 1996; Bates, 1996, Thurston, 1996) performed. They enable the establishment of the role of air pollution in the occurrence or exacerbation of a wide range of health events, ranging from early mortality to changed parameters of respiratory function and increased cardio-respiratory diseases. Because not all the problems related to air pollution effects on human health are solved, it is necessary to correlate "epidemiological evidence" to a better understanding of the etiopathogenic mechanisms which govern the appearance of air pollution adverse health effects.

to 26 cities from 12 European countries (InVS, 2002; APHEIS, 2003) and currently used by WHO (“World Health Organization”) and EEA (“European Environment Agency”) as the main model of health impact assessment for air pollution. Based on this model InVS has created an Excel spreadsheet (InVS, 2003) for the health impact assessment.

In order to reduce the time required to calculate the health impact and to simplify the process, we opted for the implementation of the calculation model directly in the DBMS by creating a SQL program (script).

As an example, the results obtained by running the script for PM<sub>10</sub> (particulate matter less than 10 micrometers of diameter) pollution health impact (mortality) in Targoviste (Romania) city area are presented below (table 1).

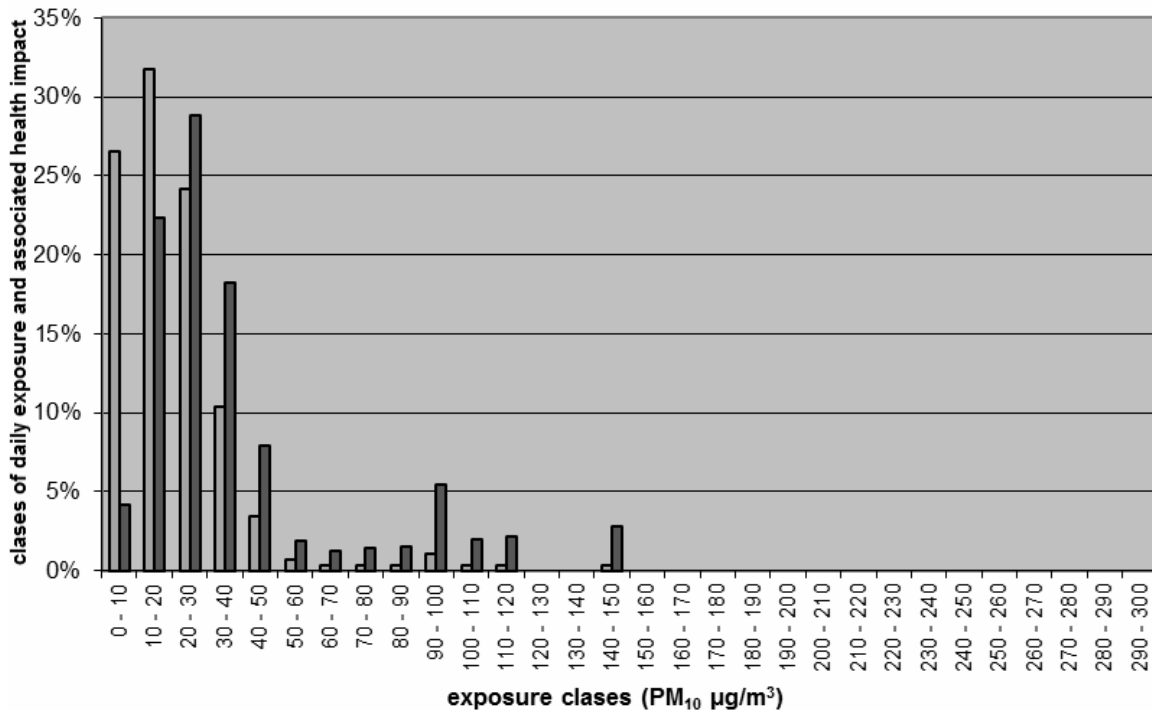
**Table 1.** Health impact assessment of PM<sub>10</sub> pollution

RESULTS		For the considered period		Annually	
		NA	IC 95%	NA	IC 95%
Scenario 1: Calculation of health impact for exposure to low pollution levels	10	3.36	2.23 - 4.49	3.36	2.23 - 4.49
Scenario 2: The health gain associated with the elimination of peaks (reference level = standard limit value)	20	1.69	1.12 - 2.26	1.69	1.12 - 2.26
Scenario 3: Health gain obtained for an exposure equal to the annual average reduced by 25%	25	1.45	0.97 - 1.93	1.45	0.97 - 1.93

In the case of the first scenario, it can be concluded that if the daily average value of pollution was equal with the chosen level, in this case 10 µg/m<sup>3</sup>, approximately 3 deaths would be avoided, with interval of confidence between 2 and 4 deaths. The distribution chart on classes of daily exposure and associated health impact is represented in the Figure 6. The grey series represents the distribution of days depending on levels of exposure to air pollution and series in black corresponds to health impact associated with each class of pollution indicator.

## 6. Using the Geographic Information System for Environment and Health

A GIS (“Geographic Information System”) can be a useful tool for researchers and health planners: “Health and ill-health are affected by a variety of life-style and environmental factors, (Scholten, Lepper, 1991). Characteristics of including the places where people live” these locations (including socio-demographic and environmental exposure) offer a valuable source for epidemiological research studies on

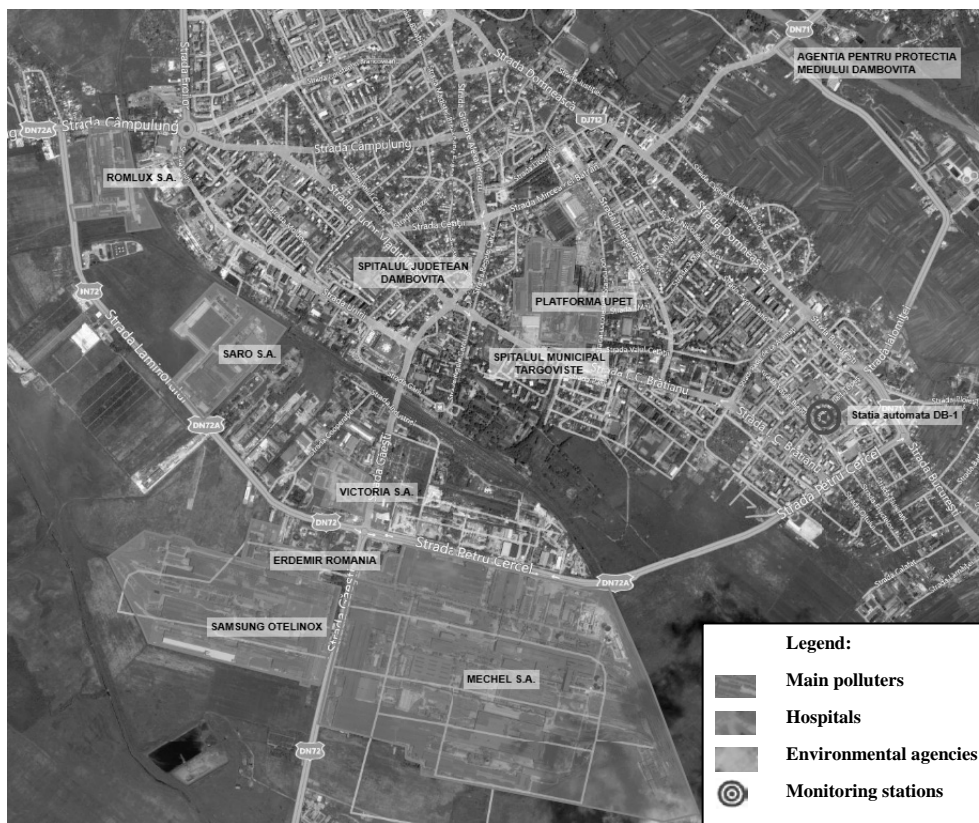


**Figure 6.** The distribution chart on classes of daily exposure and associated health impact on PM<sub>10</sub> pollution

health and environment. Therefore, health and ill-health always have a spatial dimension.”

In IISEH, GIS technology was used for several purposes such as: a) selecting the study areas, b) identifying the main polluters (figure 7) and c) disseminating information.

For a better understanding of the data contained in the databases and the health impact assessment results, we chose to create choropleth maps using GIS technology. Graduated color ramp were used for better visualization of the values distribution, taking



**Figure 7.** Point of interest identification from selected area using GIS technology



**Figure 8.** The representation of PM10 pollution [ $\mu\text{g}/\text{m}^3$ ] per counties

into account the limit values for each pollutant according to European standards (Figure 8).

## 7. Creating the Web Platform for Integrated Information System

Creating a Web portal that enables the integration of information from various sources by processing programs, automatic evaluation of the impact of pollution on health and visualization of major interest parameters, in a format available for a wide range of users is a necessity for specialists from both areas.

The IISEH portal was developed using the DotNetNuke Platform (DNN, 2011) for creating the structure and adding the textual information, while Microsoft Visual Web Developer Express was used for database driven web pages.

The portal is divided into seven major sections:

- Section "Homepage" contains general information about Integrated Information for Environment and Health System;
- Section "Data Acquisition" contains the description of the two automatic data acquisition models used in this study;
- Section "Databases" presents the structure for the main used databases and provides an easy way to observe the daily averages of pollutants;
- Section "HIA" includes automatic calculation module for health impact assessment;
- Section "GIS" includes maps made using GIS technology;
- Section "Info & News" presents a collection of news and alerts purchased using RSS feeds.

## 8. Conclusions

The system described in this paper can be easily implemented and used for any urban area and can be extended to a wider range of pollutants and health indicators by acquiring the necessary data.

To fulfill its role as a decision-making support, which describes the level of pollution and its influence on the population health, the Integrated Information System for Environment

and Health (IISEH), has the following characteristics:

**Opening:** the system offers to everyone the opportunity to consult the data they are interested in.

**Accessibility:** the system was designed to allow any user with minimal computer knowledge to learn the operating mode and to use it effectively.

**Complexity:** the system covers all designed functions for protection of the environment and public health.

**Extensibility:** new features can be added without major investments in the system.

**Modularity:** the system is organized into five levels (acquisition, database, health impact assessment, GIS, web).

**Integration:** the various parts (levels) of the system forms, from the user's point of view, a coherent system. This integration is performed at a technical level, but it ranges more to the conceptual level and involves the following aspects: a) Presentation: a uniform interface for all levels; b) Data: users see a common database, even if the distribution is made between levels; c) Communication: the transfer of data between levels is standardized.

**Maintainability:** the system can be corrected, adapted and / or improved easily.

**Reliability:** the system has been designed so that it can perform all functions required continuously. Also, recovery procedures were provided to allow the highest possible operating time.

**Cost efficiency:** for a small scale, the system can be implemented with minimum costs.

## REFERENCES

1. ADCON Telemetry, (<http://www.adcon.at/>, accessed on 05.12.2011), (2011)
2. APHEIS Report, **Health Impact Assessment of Air Pollution and Communication Strategy**, (2005), (<http://www.apheis.org/vfbisnvsApheis.pdf>, accessed on 22.09.2011)
3. ATS (American Thoracic Society), **Health Effects of Outdoor Air Pollution**, American Journal of Respiratory and



- Critical Care Medicine (1996); vol. 153, pp. 3-50.
4. BATES, D. V., **Particulate Air Pollution, Thorax** (1996), 51:S3-S8.
  5. CASTILLO, C., **Effective Web Crawling**, (2004), ([http://chato.cl/research/crawling\\_thesis](http://chato.cl/research/crawling_thesis), accessed on 23.09.2011)
  6. CHUNG TSOI, A., D. FORSALI, M. GORI, M. HAGENBUCHNER, F. SCARCELLI, **A Simple Focused Crawler**, (2003), (<http://cs.brynmawr.edu/Courses/cs380/fall2006/www12DanielePoster.pdf>, accessed on 01.10.2011)
  7. EHJ (Environmental Health Journal), **Environmental Health**, (<http://www.ehjournal.net/>, accessed on 16.12.2011), 2011.
  8. EZZATI, M., A. LOPEZ, A. RODGERS, C. MURRAY, **Comparative Quantification of Health Risk: Global and Regional Burden of Disease Attributable to Selected Major Risk Factors**, World Health Organization Library , ISBN 9241580313, 2004.
  9. DNN Platform (DotNetNuke), (<http://www.dotnetnuke.com/>, accessed on 14.12.2011), 2011.
  10. DOCKERY, D. W., C. A. III. POPE, **Acute Respiratory Effects of Particulate Air Pollution**, Annual Review Public Health, vol. 15, 1994, pp. 107-132.
  11. FILIP, F. G., F. MOISIL, **Decision Support of Environment**, in Encyclopedia of Life Support Systems – ELOSS, Thematic level article, vol. VIII, ([www.eolss.net/ebooklib](http://www.eolss.net/ebooklib)), 2002.
  12. FILIP, F. G., K. LEIVISKA, **Large-scale Complex Systems**, in Springer Book of Automation, (S.Y Nof, Ed.), Springer, New York, 2002, part D., pp. 619-638.
  13. FILIP, F. G., **Decision Support and Control of Large-scale Complex Systems**, Annual Reviews in Control, vol. 32, Issue 1, 2008, pp. 61-70.
  14. FILIP, F. G., **Designing and Building Modern Information Systems: A Series of Decisions To Be Made**, Computer Science Journal of Moldova – CSJM, vol. 19, no. 2, 2011, pp. 119-129.
  15. InVS (Institut de Veille Sanitaire), **Programme de surveillance air et santé – 9 villes**, Rapport Phase II. Institut de Veille Sanitaire, 2002, ([http://www.invs.sante.fr/publications/2002/psas\\_020624/programme.html](http://www.invs.sante.fr/publications/2002/psas_020624/programme.html), accessed on 23.09.2011)
  16. InVS (Institut de Veille Sanitaire), **APHEIS. Air Pollution and Health: a European Information System. Evaluation de l'impact sanitaire de la pollution atmosphérique dans 26 villes européennes. Synthèse des résultats européens et résultats détaillés des villes françaises**, 2002, ([http://opac.invs.sante.fr/index.php?lvl=notice\\_display&id=6028](http://opac.invs.sante.fr/index.php?lvl=notice_display&id=6028), accessed on 23.09.2011).
  17. InVS (Institut de Veille Sanitaire), **Évaluation de l'impact sanitaire de la pollution atmosphérique urbaine - Actualisation du guide méthodologique**, 2003, ([http://www.invs.sante.fr/publications/2003/eis\\_pollution/index.html](http://www.invs.sante.fr/publications/2003/eis_pollution/index.html), accessed on 22.09.2011).
  18. PENG, Y., K. GANG, Y. SHI, Z. CHEN, **A Descriptive Framework for the Field of Data Mining and Knowledge Discovery**, International Journal of Information Technology & Decision Making, Vol. 7, Issue 4, 2008, pp. 639-682.
  19. PRÜSS-ÜSTÜN, A., C. CORVALÁN, **Preventing Disease Through Healthy Environments Towards an Estimate of the Environmental Burden of Disease**, World Health Organization Library , ISBN 924593822, 2006.
  20. QUENEL, P., C. Le GOASTER, C. CASSADOU, D. EILSTEIN, L. FILLEUL, L. PASCAL, H. PROUVOST, P. SAVIUC, A. ZEGHNOUN, A. Le TERTRE, S. MEDINA, M. JOUAN, **Surveillance des effets sur la santé de la pollution atmosphérique en milieu urbain: étude de faisabilité dans 9 villes françaises.– Pollution Atmosphérique** (1997), (<http://vertigo.revues.org/4595>, accessed 24.09.2011).

21. RNMCA (Air Quality National Monitoring Network), ([www.calitateaer.ro](http://www.calitateaer.ro), accessed on 7.10.2011) in Romanian, 2011.
22. SCHOLTEN, H. J., M. J. C. LEPPER, **The Benefits of the Application of Geographical Information Systems in Public and Environmental Health**, US National Library of Medicine, 1991.
23. SCHLESINGER, R. B., **Toxicological Evidence for Health Effects from Inhaled Particulate Pollution: Does It Support the Human Experience?**, Inhalation Toxicology Journal, Vol. 7, No. 1, 1995, pp. 99-109.
24. SCHWARTZ, J., **Air Pollution and Daily Mortality: A Review and Meta-Analysis**, Environmental Research, vol. 64, 1994, pp. 36-52.
25. THURSTON, G. D., **A critical review of PM<sub>10</sub> Mortality Time-series Studies**. – Journal of Exposure Analysis and Environmental Epidemiology, vol. 6, 1996, pp. 3-21.
26. VASILIU, D., **Monitorizarea Mediului** (Environment Monitoring), Editura Tehnică, București, 2007, pp. 41-61, 90-105, 268-329, in Romanian.
27. WHO (World Health Organization), **Global Health Risks Mortality and Burden of Disease Attributable to Selected Major Risks**, World Health Organization Library, ISBN 978 9241563871, 2009.
28. WHO (World Health Organization), **Environmental Health**, ([www.who.int/topics/environmental\\_health/en/](http://www.who.int/topics/environmental_health/en/), accessed on 10.12.2011), 2011.