



"This project has received funding from the European Union's Seventh Programme for research, technological development and demonstration under grant agreement N°603946"



HEALS

Health and Environment-wide Associations
based on Large population Surveys

FP7-ENV-2013-603946

<http://www.heals-eu.eu/>

D8.1 Report on quality assessment and quality control of environmental data collected

WP8 Environmental data mining

Version 1

Lead beneficiary: UOWM

Date: 30/9/2014

Nature: R

Dissemination level: PU




 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

TABLE OF CONTENTS

1. ABSTRACT	5
2. INTRODUCTION	5
3. ENVIRONMENTAL DATA MANAGEMENT SYSTEM (EDMS)	7
3.1. Data categories of EDMS	12
3.1.1. Land use/Land cover	13
3.1.2. Meteorological Data.....	15
3.1.3. Comfort data indoors (temperature-relative humidity).....	17
3.1.4. Air emission data	18
3.1.5. Air emission data – consumer products	20
3.1.6. Pollutant concentration data – air	21
3.1.7. Pollutant concentration data – soil.....	26
3.1.8. Pollutant concentration data – dust	27
3.1.9. Pollutant concentration data – water	27
3.1.10. Pollutant concentration data – drinking water	28
3.1.11. Food concentration data	29
3.1.12. Population data (age/gender/etc)	30
3.1.13. Noise data.....	31
3.1.14. Building characteristic data.....	33
3.2. Template of EDMS input data.....	34
3.3. Dataset selection procedure	38
4. DATA DERIVATION METHODOLOGY ASSESSMENT	39
4.1. Land use/Land cover	39
4.2. Meteorological Data.....	40
4.3. Comfort data indoors (temperature-relative humidity).....	42
4.4. Air emission data	42
4.5. Air emission data – consumer products	43
4.6. Pollutant concentration data – air	44
4.7. Pollutant concentration data – soil	53

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

4.8.	Pollutant concentration data – dust	58
4.9.	Pollutant concentration data – water	58
4.10.	Pollutant concentration data – drinking water	59
4.11.	Food concentration data	60
4.12.	Population data (age/gender/etc)	61
4.13.	Noise data.....	61
4.14.	Building characteristic data.....	62
5.	DATA QUALITY CONTROL	62
5.1.	Land use/Land cover	63
5.2.	Meteorological Data.....	64
5.3.	Comfort data indoors (temperature-relative humidity).....	65
5.4.	Air emission data	65
5.5.	Air emission data – consumer products	65
5.6.	Pollutant concentration data – air	66
5.7.	Pollutant concentration data – soil	74
5.8.	Pollutant concentration data – dust	75
5.9.	Pollutant concentration data – water	75
5.10.	Pollutant concentration data – drinking water	77
5.11.	Food concentration data	77
5.12.	Population data (age/gender/etc)	78
5.13.	Noise data.....	79
5.14.	Building characteristic data.....	80
6.	CONCLUSIONS/RECOMMENDATIONS	81
7.	CONTRIBUTION	82
8.	REFERENCES	83
9.	ANNEXES	84

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

Document Information

Grant Agreement Number	ENV-603946	Acronym	HEALS
Full title	Health and Environment-wide Associations based on Large population Surveys		
Project URL	http://www.heals-eu.eu/		
EU Project Officer	TuomoKarjalainen,- Tuomo.KARJALAINEN@ec.europa.eu		


Deliverable	Number	8.1	Title	Report on quality assessment and quality control of environmental data collected
Work Package	Number	8	Title	Environmental data mining

Delivery date	Contractual	M12	Actual	30/9/2014
Status	Draft <input checked="" type="checkbox"/>		Final <input type="checkbox"/>	
Nature	Demonstrator <input type="checkbox"/> Report <input checked="" type="checkbox"/> Prototype <input type="checkbox"/> Other <input type="checkbox"/>			
Dissemination level	Confidential <input type="checkbox"/> Public <input checked="" type="checkbox"/>			

Author (Partners)	University of Western Macedonia (UOWM)			
Responsible Author	John Bartzis		Email	bartzis@uowm.gr
	Partner 16	UOWM	Phone	0030 24610 56710

Document History

Name	Date	Version	Description
Report on quality assessment and quality control of environmental data collected	30/9/2014	1	

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100


1. ABSTRACT

In the framework of the WP 8 of the HEALS project an extensive data mining were execute on emissions of stressors, concentrations of toxic substances in environmental media (outdoor and indoor air, soil, water), in food and in drinking water, various meteorological conditions, external exposures to environmental hazards and the impact of environmental and microenvironment modulation. The data mining comprised by exploration of the existing European and worldwide organizations databases of the aforementioned parameters, various European and national project and literature published review. The data investigation consist on location, location characteristics, the time period and the time resolution, units of measurement, measurement methodology, equipment used and the quality assurance and quality control of the related data. Data on the following categories: Land use/Land cover, Meteorological Data, Comfort data indoors (temperature-relative humidity), Air emission data, Air emission data – consumer products, Pollutant concentration data – air, Pollutant concentration data – soil Pollutant concentration data – dust Pollutant concentration data – water Pollutant concentration data – drinking water Food concentration data Population data (age/gender/etc) Noise data Building characteristic data have been collected and assessed. The quality assurance and quality control procedures of the above mentioned data are described thoroughly. Finally some recommendations derived from the data mining quality assurance/quality control that might be adopted while collecting the data through the EXHES study of the WP17, is presented.

2. INTRODUCTION

HEALS project represents a comprehensive applied methodology focusing on the different aspects of individual assessment of exposure to conventional and emerging environmental stressors and on the prediction of the associated health outcomes. HEALS adopt one defined by complex and dynamic interactions between DNA sequence, epigenetic DNA modifications, gene expression and environmental factors that all combine to influence disease phenotypes.


In the frame of WP8 HEALS project, an environmental data management system is developed to permit the integration of data on emissions of stressors, concentrations of toxic substances in environmental media (outdoor and indoor air, soil, water), in food and in drinking water and external exposures to environmental hazards. The following specific assignments have been identified:

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

- Data collection on specific environmental stressors on different environmental media. Environmental data sources needed to perform Environment-wide application studies focusing in the areas covered by the population studies addressed in Stream 5 of the project. A detailed review including past and ongoing research and survey projects both at National and European level and EU-wide monitoring systems such as the ones managed by EEA and ESA, national and regional monitoring networks in the areas of interest to the Stream 5 population studies have been identified until now. Contributions on data provision from every HEALS partner will be assessed to verify the data directly available within the consortium. The main objective of this Task is to gather collect and mining the environmental data from the information sources identified through the above review process and to be successively stored in the environmental management system developed in Task 8.3. The data collected, relevant to the groups of substances identified in Stream 5, will comprise but will be not limited to the following variables:
 - Emission data and emission factors
 - Satellite data for estimation of air pollution levels (data from and in collaboration with the GMES initiative)
 - Pollution levels in different media (outdoor air, indoor air, soil, surface water, ground water)
 - Pollution levels in food and drinking water
 - Meteorological data to be used in input of air quality modeling
 - Land use/land/cover for estimation of emission inventory

Since the data collected at this stage should serve WP11 and WP12, the dataset will be completed and provided in a standard format, compliant with the INSPIRE Directive.

- Quality of the collected data. The data collected until now have been evaluated against their quality and applicability through the activities foreseen in the frame of the Task 8.2 of the project plan. Careful checking of all the data used, whether obtained from outside the project or derived and provided through the project partners has been performed. A wide range of methods have been used for this purpose, building on techniques already extensively applied on existing data set. These includes:
 - consultation with data suppliers and past data users to identify any known gaps or uncertainties;
 - scrutiny of relevant metadata, describing the source and genealogy of the data (e.g. sampling methods, measurement procedures, analytical procedures, reporting);
 - establishment of clear data standards and criteria prior to their acquisition and use, so that data can be rejected if they are not of a sufficient quality (e.g. in terms of coverage, sampling density, measurement accuracy, timeliness);

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

- statistical screening of the data – e.g. to check for outliers, impossible values, anticipated correlations/patterns;
- manual checking of subsamples of the data (e.g. to check for formatting errors);
- intercomparison and triangulation against independent data sets and sources.

An important corollary of the above QA/QC activities is the identification of gaps, which will need to be covered. Moreover this helps the project team guide the optimally design the Pilot European Exposure and Health Examination Survey (EXHES) carried out in WP17.

The objectives of the above procedure are the following:

- To identify, mine, collect and review existing datasets including environmental and food contamination variables for different environmental health stressors considered in Stream 5 and of ancillary data such as spatially resolved land use/cover data.
- To analyse the collated data in terms of collection method and data quality, availability, applicability domains.
- To store the collated datasets within an environmental data management system to render the data readily available for the WP9 and WP11 and finally for health impact assessment in population studies executed in Stream 5.


3. ENVIRONMENTAL DATA MANAGEMENT SYSTEM (EDMS)

The environmental data management system has been created through the WP8 of the HEALS project in order to gather, collect and mine the environmental data from other information sources and data that will be produce through the project. The first approach on creating the EDMS had three alternatives:

- Link to existing databases
- Incorporate these application to the environmental management system
- Extract datasets and import them to the environmental management system

The last is characterized as the best alternative; however the storage capacity of the system will increase in huge number. Also the design of the database includes the storage of the collected data in a coherent environmental data management system (EDMS) for further use within the project.

The design and implementation of the Environmental Data Management System or EDMS, is one of the three specific tasks in WP 8 and is in straight relation to the other two tasks, namely Data Collection and Data Quality and Assessment. In short, the EDMS has to be a

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

solid and robust application in order to host the data processed or pre-processed by the aforementioned tasks. Pre-processing involves Environmental Data Source Identification suitable for Environment-wide application studies in specific areas. This task will be carried out through detailed review of past and ongoing research and survey projects.

Moreover, this data have to be gathered as well and possibly stored in its existing format before they are stored in the EDMS. This means that primary data, when available, will come in a wide variety of forms as long as its structure and file format is concerned. Ultimately, since the collected data will serve other WPs (11 and 12) the dataset, after having been evaluated in terms of quality and applicability, will have to be transformed in a standard format, compliant with the INSPIRE Directive.


On top of the already referred cross – task issues the EDMS must be designed and implemented following certain procedures and standards. Critical issues that must be taken into account here is that the HEALS database should be able to retrieve data from existing Databases. Additionally,

a) All data must be geo-referenced, by specifying the geographic coordinates of each single observation, both for point-form and for polygonal spatial information and

b) All data will be univocally coupled to a time reference (instant, hour, day, month, etc.) and as such they must be ready to be investigated, by using time series based statistics. Also, they must be easily aggregated on the basis of different time scales. There also must exist queries for automatic updating, importing data, exporting data, selecting specific subsets and grouping records.

Furthermore the EDMS is in compliance with the INSPIRE Directive. The Directive 2007/2/EC of the European Parliament aims to create a European Union (EU) spatial data infrastructure. It addresses 34 spatial data themes which are based on the following common principles:

- Data should be collected only once
- Combine seamless spatial information
- Should be possible for information collected to be shared
- Geographic information readily available
- Geographic information easy to find
- Implementing Rules (IR).
 - Metadata
 - Data Specifications
 - Network Services
 - Data and Service Sharing

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

– Monitoring and Reporting


- The Implementing Rules are adopted as Commission Decisions or Regulations and are binding in their entirety.

Building the EDMS is perceived as a 4 step process. First, we must define the web framework and application architecture which will be used. The application, in general will consist of a web component, various application programming interfaces for interacting with databases and the database itself. Then, we have to choose an appropriate Database Management System and design of the relational model or models. Secondly, we must populate the EDMS via a commonly accepted template for entering data sets. The third step is to establish links to existing data bases or more appropriately to their web interfaces in order to provide a head start for the steps to come. At the same time we can work on the matter of connecting to other databases though queries for automatic updating, importing data, exporting data, making specific selections or grouping records. The last step involves data which come in formats that require further processing.

EDMS - Design, implementation and data population

- STEP 1
 - Web framework / application architecture
 - Database Architecture
- STEP 2
 - Data Population via (Common) Data Template(s)
- STEP 3
 - Data directly retrieved from other databases
- STEP 4
 - Other data transfer

The design of the Database structure which besides accommodating **HEALS** owns datasets should be able to retrieve data from **existing Databases** identified in Task 8.1 through suitable query scripts. The Database in the EDMS, will be implemented in a standard database package or using a standard database management system, for interoperability in data storage, management and exchange with WP12. Interoperability involves both the database for the stored datasets and any other databases necessary for building the web interface of the stored data as well as the Application Programming Interfaces (API's) for data exchange. Relational database management systems such as MySQL or PostgreSQL seem the straightforward choice for this part of the application, being open source and widely tested, supported and of course deployed in numerous applications, not to mention their ease of use in web applications. PostgreSQL has the additional attribute of being object – oriented, bringing the advantages of concepts such as objects, classes and inheritance in data base schemas and is probably the best alternative in cases MySQL cannot cope with

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

the datasets that are to be stored and queried. The implication here is that it is not necessary to use only one Database Management System. Should the application dictates a different approach additional systems should be used without losing compatibility with the existing ones.

It is obvious that one of the first tasks to address must be the design of the relational models and table schemas which will clearly depict how the datasets will be stored and how the integrity constraints, which are set to provide and maintain the accuracy and consistency of data over its entire life-cycle, will be defined and then expressed in the relational model. Having in mind the diversity of available datasets this will be a very complex task and numerous issues that have to be addresses individually, will most certainly arise during the implementation process.


The Database structure which will accommodate HEALS datasets will also store data from existing Databases. These databases will have already been identified and suitable query scripts, depending on each situation, must be implemented in order to fulfill the task of retrieving and converting the desirable data.

Additional tasks to fulfill while implementing the database include a) automatic updating, b) importing form other data formats or software (such as IPCHeM and ToxHub), c) exporting into other data formats, d) selecting specific subsets of data, e) grouping records by means of aggregation functions.

The structure of the EDMS must be compatible with the **IPCHeM** database of the JRC and the **ToxHub** platform of the HEROIC project, so that the collected data can feed into the above databases during project execution and in the future, contributing thus to environmental data integration across Europe.

Data base architecture design STEP 1

- Review existing databases' schema
- Define DB structure (tables and relations). This structure should be compatible with the IPCHeM database of the JRC and the ToxHub platform of the HEROIC project.
- Define queries to import data from other databases
- Define update queries
- Define selection queries
- PHP + MySQL
- All data will geo-referenced (integration with WP12)
- The environmental database management system will be available as a **module** to the **GeoDatabase** platform.
- EDMS will be populated by insert queries with data that will be collected using the specified template.

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100


- Develop scripts to collect and store data
- Convert to our own db format (schema)
- Define queries to import data
- Define update queries
- Define selection queries

The web framework is the software which we can use to implement the web application and the APIs to it and its underlying databases. Some of the most prominent programming languages which support web frameworks are PHP, Python, Java and Ruby. For the EDMS we opted for PHP as it is backed by widely tested frameworks that assist our effort. Such frameworks would also impose the Model–View–Controller or MVC software pattern a general application architecture.

In MVC, the model represents the data. The model or data is independent of the model and view components. The view displays the model data, and sends user actions (e.g. button clicks) to the controller. That is, after it has been formed and presented to the user, the user can use it to invoke the controller which will possibly retrieve data or manipulate the data and alter the view.

The controller as already said, provides model data to the view and interprets user actions. It can also send commands to its associated view to change the view's presentation of the model. The controller generally depends on the view and the model.

The MVC model is quite straightforward and can be used to model the EDMS reliably. In the EDMS data have to be stored independently. This represents the model. An interface must be present to retrieve or alter the data which is the controller and finally the data must be presented to the end user. The latter can be perceived as the view.

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

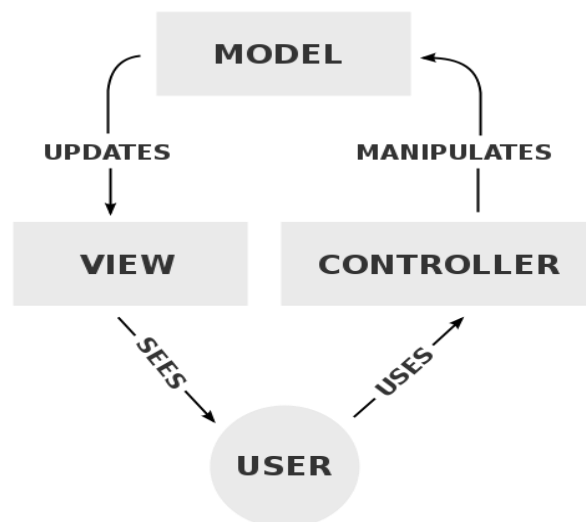


Figure 1. Model – View - Controller


Data Retrieval

- Data retrieval through direct filtering option
 - Use filters to select the parameters of the data to be retrieved
- Data retrieval through map representation
 - Use filters to select Data Class, Country, Pollutant/Stressor and Data Source
 - Have available stations displayed on map
 - Click stations to retrieve the data

3.1. Data categories of EDMS

Humans exposed to different environments through their entire life. In order to have a comprehensive picture of the human exposure to various environmental media the EDMS has an integrated list of media that human can be exposed, containing data on the environmental stressors that can potential affect the human health. These categories are:

- Land use/Land cover
- Meteorological Data
- Comfort data indoors (temperature-relative humidity)
- Air emission data
- Air emission data – consumer products
- Pollutant concentration data – air
- Pollutant concentration data – soil

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

- Pollutant concentration data – dust
- Pollutant concentration data – water
- Pollutant concentration data – drinking water
- Food concentration data
- Population data (age/gender/etc)
- Noise data
- Building characteristic data

3.1.1. Land use/Land cover

Europe is one of the most intensively used continents on the earth, with the highest share of land used for settlement, production systems (including agriculture and forestry) and infrastructure. How land is used constitutes one of the principal reasons for environmental change, with significant impacts on quality of life and ecosystems. In addition, data about land cover/land use indicating potential sources of environmental contaminations.

In the EDMS DataBase there are the following data sets:

A

Data for various European countries on land use/land cover for the period of **1992-1994**. The global land 1-km AVHRR dataset consists a network of 29 high resolution picture transmission (HRPT) stations, along with data recorded by the National Oceanic and Atmospheric Administration (NOAA), has been acquiring daily global land coverage since 1992 until 1994. Under the guidance of the International Geosphere Biosphere Programme, processing standards for the AVHRR data have been developed for calibration, atmospheric correction, geometric registration, and the production of global 10-day maximum normalized difference vegetation index (NDVI) composites.

The corresponding web page of the land use/land cover data described above is: <http://edc2.usgs.gov/1km/1kmhomepage.php>.

B


The Corine database on land use/land cover for Europe (an area of 62500 km²). Maps are provided for a specific time point (year 2006) and are updated if necessary on future time points. Maps are provided with 250x250 m resolution.

The corresponding web page of the land use/land cover data described above is: <http://www.eea.europa.eu/data-and-maps/data/corine-land-cover-2006-raster-3#tab-gis-data>.

C

Data describes annual average temperature in region of Catalunya (Spain) in C° from **23/12/1995 to 23/12/1996**.

The corresponding web page of the meteorological data described above is: <http://www20.gencat.cat/portal/site/territori/menuitem.2a0ef7c1d39370645f13ae92b0c0e1a0/?vgnextoid=f331e7c9a149b310VgnVCM1000008d0c1e0aRCRD&vgnnextchannel=f331e7c9a>

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

[149b310VgnVCM1000008d0c1e0aRCRD&vgnnextfmt=default](http://www20.gencat.cat/portal/site/territori/menuitem.2a0ef7c1d39370645f13ae92b0c0e1a0/?vgnextoid=1ce2e7c9a149b310VgnVCM1000008d0c1e0aRCRD&vgnnextchannel=1ce2e7c9a149b310VgnVCM1000008d0c1e0aRCRD&vgnnextfmt=default). It is uploaded to the EDMS as a single file.

D

Annual average precipitation for region of Catalunya (Spain) in mm. Data cover the period of **01/01/1961 to 01/01/1990**.

The corresponding web page of the meteorological data described above is: <http://www20.gencat.cat/portal/site/territori/menuitem.2a0ef7c1d39370645f13ae92b0c0e1a0/?vgnextoid=1ce2e7c9a149b310VgnVCM1000008d0c1e0aRCRD&vgnnextchannel=1ce2e7c9a149b310VgnVCM1000008d0c1e0aRCRD&vgnnextfmt=default>. It is uploaded to the EDMS as a single file.

E

Difference between the potential evapotranspiration and real evapotranspiration, defining the water deficit, in mm, for Spain. Data cover the period of **23/12/1995 to 23/12/1996**.

The corresponding web page of the meteorological data described above is: <http://www20.gencat.cat/portal/site/territori/menuitem.2a0ef7c1d39370645f13ae92b0c0e1a0/?vgnextoid=6764e7c9a149b310VgnVCM1000008d0c1e0aRCRD&vgnnextchannel=6764e7c9a149b310VgnVCM1000008d0c1e0aRCRD&vgnnextfmt=default>. It is uploaded to the EDMS as a single file.

F

Values of annual wind speed average wind to a height of 60 m distributed with a resolution of 200 x 200 m. Values are in km/h, m/s. Data cover the region of Catalunya (Spain) for the period of **26/11/2003 to 26/11/2004**.

The areas are distributed in ten categories, from 0 m/s to 9,5 m/s


The corresponding web page of the meteorological data described above is: <http://www20.gencat.cat/portal/site/territori/menuitem.2a0ef7c1d39370645f13ae92b0c0e1a0/?vgnextoid=b2fc569d1939b310VgnVCM1000008d0c1e0aRCRD&vgnnextchannel=b2fc569d1939b310VgnVCM1000008d0c1e0aRCRD&vgnnextfmt=default>. It is uploaded to the EDMS as a single file.

G

Wind annual average speed at 80 m of height. The parameters are quantified in m/s, km/h and the area are distributed in 10 categories from 0 m/s to 9.5 m/s. Data cover the region of Catalunya (Spain) for the period of **26/11/2003 to 26/11/2004**.

The corresponding web page of the meteorological data described above is: <http://www20.gencat.cat/portal/site/territori/menuitem.2a0ef7c1d39370645f13ae92b0c0e1a0/?vgnextoid=c98d9f7ae839b310VgnVCM1000008d0c1e0aRCRD&vgnnextchannel=c98d9f7ae839b310VgnVCM1000008d0c1e0aRCRD&vgnnextfmt=default>. It is uploaded to the EDMS as a single file.

H

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

Daily global Irradiation with a resolution of 100m x100m and the values are quantified in mJ/m2. There's 12 categories, from < 12 (MJ/m2) to 17,5 (MJ/m2). Data cover the region of Catalunya (Spain) for the period of **01/01/2001 to 31/12/2001**.

The corresponding web page of the meteorological data described above is: <http://www20.gencat.cat/portal/site/territori/menuitem.2a0ef7c1d39370645f13ae92b0c0e1a0/?vgnextoid=fb13e7c9a149b310VgnVCM1000008d0c1e0aRCRD&vgnnextchannel=fb13e7c9a149b310VgnVCM1000008d0c1e0aRCRD&vgnnextfmt=default>. It is uploaded to the EDMS as a single file.

3.1.2. Meteorological Data

Meteorological data cover all the monitoring data that have been collected from various meteorological stations located around Europe. Also in EDMS are integrated data from modelling processes like air dispersion models.

In the EDMS DataBase there are the following data sets:


A

The ECA dataset consists of daily station series obtained from climatological divisions of National Meteorological and Hydrological Services and station series maintained by observatories and research centres throughout Europe and the Mediterranean. ECAD covers the following elements: mean temperature (0.1 °C), maximum temperature (0.1 °C), minimum temperature (0.1 °C), sunshine (0.1 hours), snow depth (1cm), precipitation amount (0,1 mm) , sea level pressure (0.1 hPa), humidity (1%) , wind gust (0.1 m/s), wind speed (0.1 m/s) , wind direction (degrees), cloud cover (oktas).

Data regarding the above parameters exists on daily basis from **1850 to 2013**. The corresponding web page of the meteorological data described above is <http://eca.knmi.nl/dailydata/index.php>.

B

NASA, through its' Science Mission Directorate, has long supported satellite systems and research providing data important to the study of climate and climate processes. These data include long-term estimates of meteorological quantities and surface solar energy fluxes. These satellite and modeled based products have been shown to be accurate enough to provide reliable solar and meteorological resource data over regions where surface measurements are sparse or nonexistent, and offer two unique features - the data is global and, in general, contiguous in time. Surface meteorology and Solar Energy (SSE) dataset covers data for the following parameters: Daylight cloud amount, Cloud amount at 3-hourly intervals, Frequency of cloud amount at 3-hourly, Air Temperature, Daily Temperature

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

Range, Cooling Degree Days above 18° C, Heating Degree Days below 18° C, Arctic Heating Degree Days below 10° C, Arctic Heating Degree Days below 0°, Earth Skin Temperature, Daily Mean Earth Temperature minimum, maximum and Frost Days Dew/Frost Point Temperature, Wind speed at 50 m, Percent of time for ranges of wind speed at 50, Wind Speed at 50 m at 3-hourly intervals, Wind Direction at 50 m, Wind Direction at 50 m at 3-hourly intervals, Wind Speed at 10 m for terrain similar to airports, Wind Speed adjustments for height and vegetation type.

Data regarding the above parameters exists on daily basis from **1983-2005**. The corresponding web page of the meteorological data described above is: <https://eosweb.larc.nasa.gov/cgi-bin/sse/daily.cgi?email=skip@larc.nasa.gov>.

C

Data describes annual average temperature in region of Catalunya (Spain) in C° from **23/12/1995 to 23/12/1996**.

The corresponding web page of the meteorological data described above is: <http://www20.gencat.cat/portal/site/territori/menuitem.2a0ef7c1d39370645f13ae92b0c0e1a0/?vgnextoid=f331e7c9a149b310VgnVCM1000008d0c1e0aRCRD&vgnnextchannel=f331e7c9a149b310VgnVCM1000008d0c1e0aRCRD&vgnnextfmt=default>. It is uploaded to the EDMS as a single file.

D

Annual average precipitation for region of Catalunya (Spain) in mm. Data cover the period of **01/01/1961 to 01/01/1990**.

The corresponding web page of the meteorological data described above is: <http://www20.gencat.cat/portal/site/territori/menuitem.2a0ef7c1d39370645f13ae92b0c0e1a0/?vgnextoid=1ce2e7c9a149b310VgnVCM1000008d0c1e0aRCRD&vgnnextchannel=1ce2e7c9a149b310VgnVCM1000008d0c1e0aRCRD&vgnnextfmt=default>. It is uploaded to the EDMS as a single file.


E

Difference between the potential evapotranspiration and real evapotranspiration, defining the water deficit, in mm, for Spain. Data cover the period of **23/12/1995 to 23/12/1996**.

The corresponding web page of the meteorological data described above is: <http://www20.gencat.cat/portal/site/territori/menuitem.2a0ef7c1d39370645f13ae92b0c0e1a0/?vgnextoid=6764e7c9a149b310VgnVCM1000008d0c1e0aRCRD&vgnnextchannel=6764e7c9a149b310VgnVCM1000008d0c1e0aRCRD&vgnnextfmt=default>. It is uploaded to the EDMS as a single file.

F

Values of annual wind speed average wind to a height of 60 m distributed with a resolution of 200 x 200 m. Values are in km/h, m/s. Data cover the region of Catalunya (Spain) for the period of **26/11/2003 to 26/11/2004**.

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

The areas are distributed in ten categories, from 0 m/s to 9,5 m/s

The corresponding web page of the meteorological data described above is: <http://www20.gencat.cat/portal/site/territori/menuitem.2a0ef7c1d39370645f13ae92b0c0e1a0/?vgnextoid=b2fc569d1939b310VgnVCM1000008d0c1e0aRCRD&vgnnextchannel=b2fc569d1939b310VgnVCM1000008d0c1e0aRCRD&vgnnextfmt=default>. It is uploaded to the EDMS as a single file.

G

Wind annual average speed at 80 m of height. The parameters are quantified in m/s, km/h and the area are distributed in 10 categories from 0 m/s to 9.5 m/s. Data cover the region of Catalunya (Spain) for the period of **26/11/2003 to 26/11/2004**.

The corresponding web page of the meteorological data described above is: <http://www20.gencat.cat/portal/site/territori/menuitem.2a0ef7c1d39370645f13ae92b0c0e1a0/?vgnextoid=c98d9f7ae839b310VgnVCM1000008d0c1e0aRCRD&vgnnextchannel=c98d9f7ae839b310VgnVCM1000008d0c1e0aRCRD&vgnnextfmt=default>. It is uploaded to the EDMS as a single file.

H

Daily global Irradiation with a resolution of 100m x100m and the values are quantified in mJ/m². There's 12 categories, from < 12 (MJ/m²) to 17,5 (MJ/m²). Data cover the region of Catalunya (Spain) for the period of **01/01/2001 to 31/12/2001**.


The corresponding web page of the meteorological data described above is: <http://www20.gencat.cat/portal/site/territori/menuitem.2a0ef7c1d39370645f13ae92b0c0e1a0/?vgnextoid=fb13e7c9a149b310VgnVCM1000008d0c1e0aRCRD&vgnnextchannel=fb13e7c9a149b310VgnVCM1000008d0c1e0aRCRD&vgnnextfmt=default>. It is uploaded to the EDMS as a single file.

3.1.3. Comfort data indoors (temperature-relative humidity)

This section concerning data related to micro-environments climate relates to indoor spaces such as places of residence, workplace, kindergarten or school.

In the EDMS DataBase there are integrated the following data sets:

A

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

Data concerns urban indoor environment and particular schools, for Portugal and the region of Porto. The data are measurements on temperature (°C), relative humidity (%) and CO2 concentration (ppm) for the period of **01/11/2011 to 20/12/2012**. Air sampling was performed concurrently in two heating season periods: November 2011 to March 2012 and November 2012 to December 2012. The air sampling covered a 5-day period (school week, from Monday morning until Friday afternoon) during teaching periods, with a time step of 5 minutes.

The data are uploaded to the EDMS as a single file.

B

Data concerns urban indoor environment at homes and particular bedrooms, for Portugal and the region of Porto. The data are measurements on temperature (°C), relative humidity (%) and CO2 concentration (ppm) for the period of **01/11/2012 to 31/03/2013**. Air sampling took place from November 2012 to March 2013.

The air sampling covered a period of 7 days with a time step of 5 minutes. In each investigated home the child's bedroom were selected for air sampling.

Homes were identified from C01 to C68.

The data are uploaded to the EDMS as a single file.


3.1.4. Air emission data

Air emission data includes the anthropogenic and biogenic emissions (point and non-point sources) of pollutants for being used in many different applications and particularly on air quality modelling systems. Emission inventories are useful information for prediction the dispersion and concentration of pollutants in the atmosphere (outdoor and indoor). This is complementary to land cover data, but for the indoor air can specifically related to point-like sources of pollutants whose emissions can be linked to a specific location diverted from building materials and consumer products.

In the EDMS DataBase there are the following data sets:

A

The BUMA project aims to thoroughly assess the human exposure to air hazards emitted by building materials commonly used in Europe. Indoor organic compounds are released from a variety of building materials including vinyl tiles and coving, carpets, particleboards, wood products, paints, adhesives etc. VOCs associated with paints, varnishes, sealing caiks, adhesives, carpets and other materials are likely to be a major cause of health complaints associated with indoor air in new buildings.

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

The database includes the Emission Rates ($\mu\text{g}/\text{h}/\text{m}^2$) from materials in the following categories: Adhesives, floorings, gypsum products and plaster boards, internal and external wall ceiling finishes, thermal insulation products, wood based panels, carpets, paints and varnishes. It contains data for emission from literature (from **1995 and on**) as well as from project's experiments. Data refer to the following environmental stressor: VOCs, organic compounds and aldehydes.

The corresponding web page of the air emission data described above is: <http://www.mech.uowm.gr/bumaproject/>.

B

Data from CEIP (the EMEP Centre on Emission Inventories and Projections). Data cover the period from **1980** until now with a projection to **2050**. The spatial resolution of the data covers the area of 2500 Km² (dimensions 50km x 50km). The environmental stressors that data existed are: PM₁₀, PM_{2.5}, VOCs, POPs, NO₂, SO₂, NH₃ and CO.

The corresponding web page of the emission data described above is: http://www.ceip.at/ms/ceip_home1/ceip_home/webdab_emepdatabase/.

C

Data obtained from the United Nations Framework Convention on Climate Change; National Inventory Submissions 2014. Data cover the period from **1990** until **2012** with time resolution of 1 year. Data refer to the following environmental stressor: VOCs, POPs, NO₂, SO₂, CO, N₂O, CH₄ and CO₂.

The corresponding web page of the emission data described above is: http://unfccc.int/national_reports/annex_i_ghg_inventories/national_inventories_submissions/items/8108.php.

D


Data from the University of Stuttgart (Germany) for the 27 EU countries plus the EFTA Countries. Data cover the period from **2010** until now with a projection to **2030**.

-USTUTT Emission Data -Model Resolution: 5km*5km – grid cell; calculations also possible for 1km*1km

-Temporal Resolution: hourly

-Key sources: anthropogenic emission sources in NFR structure

The environmental stressors that data covered are: PM₁₀, PM_{2.5}, VOCs, toxic metals, POPs, NO₂, SO₂, CO₂, CH₄, NH₃ and CO.

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

Model Data sets are available offline from Institute for Energy Economics and Rational Use of Energy (IER), University of Stuttgart.

E

Emission data cover the region of Catalunya (Spain). Data corresponds for average CO2 emission for one year period (**01/01/2009 – 31/12/2009**). Values are in tons/year. There's 5 categories divided from <50000 tons/year to >1000000.

The corresponding web page of the emission data described above is: <http://www20.gencat.cat/portal/site/territori/menuitem.2a0ef7c1d39370645f13ae92b0c0e1a0/?vgnextoid=0c811aea2789b310VgnVCM2000009b0c1e0aRCRD&vgnnextchannel=0c811aea2789b310VgnVCM2000009b0c1e0aRCRD&vgnnextfmt=>. It is uploaded to the EDMS as a single file.

3.1.5. Air emission data – consumer products

Product emissions are major contributors to the indoor air quality in dwellings. Several international studies have reported on the occurrence of ingredients of health concern in consumer products. In the EDMS DataBase there are the following data sets:


The BUMAC database was set-up as a compilation of data on the current state-of-the-art on consumer product compositions and emissions, on test chamber experimental results, data on ventilation patterns and reported indoor condition values (test conditions such as temperature, relative humidity, etc) exposures, risks and health end points. Qualitative data were assured by using only data outcomes from procedures derived from standardized emission test protocols. The environmental stressors that emission data referred are: PM10, PM2.5, VOCs, Phthalates, Organic compounds, Organoalogenes, POPs, NO2, O3 and NH3.

BUMAC classifies, organizes and presents in a user-friendly and comprehensive way, the emission data from commonly used consumer products and their emitted compounds. The BUMAC database is continuously updated with emission data from studies carried out by certificated institutions as well as by peer reviewed literature focusing more on data for health relevant compounds. The database also includes data from emission studies which were carried out in the frame of EPHECT on 15 selected consumer products.

The database includes the following categories:

Air fresheners, Appliances, clothes and fabrics, decoration, electronic equipment, fireplaces, flowers and plants, household products, personal care products/cosmetics, pest control, pest care products, printed materials, toys.

The corresponding web page of the air emission data for consumer products described above is: <http://bumac.uowm.gr/>.

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

3.1.6. Pollutant concentration data – air

Data about concentrations of pollutants/stressors in environmental media, e.g.:

- Outdoor air
- Indoor air (local measurement or modelled based on outdoor air)

In the EDMS DataBase there are the following data sets:

A

Data for outdoor air for several environmental stressors are uploaded to the AirBase database of EEA. Data came from monitoring stations throughout Europe (point data). Data presented in csv files for every stressor for every European country and time resolution (daily or hourly data).

Concerning Greece, data cover the municipalities of Thessaloniki, Aliartos, Athens, Agia Paraskevi, Elefsina, Galatsi, Heraclion, Ioannina, Kalamaria, Eleftherio-Kordelio, Koropi, Larissa, Ano Liosia, Likovrisi, Amarousio, Megalopoli, Nea Smyrni, Kalithea, Oinofyta, Panorama, Patra, Peristeri, Pireas, Sindos, Ptolemaida, Thracomakedones, Zografos and Volos. Data are presented in hourly and daily basis covering the period of **01/01/1984 to 31/12/2012**. Data referred to quantified stressor concentration in air. The environmental stressors that data covered are: PM10, PM2.5, VOCs, toxic metals, NO2, O3, SO2, CO, NOx, NO and CO.

Units are µg/m3 except for:

CO: mg/m3 and

Cadmium, nickel, arsenic: ng/m3


The corresponding web page of the pollutant concentration described above is: <http://www.eea.europa.eu/data-and-maps/data/airbase-the-european-air-quality-database-8>, and especially for Greece: http://www.eea.europa.eu/data-and-maps/data/airbase-the-european-air-quality-database-8/gr/airbase_gr_v8.zip.

B

OFFICAIR data cover indoor and outdoor atmosphere. In OFFICAIR project were monitored 37 office buildings in 8 countries: Greece, Spain, Italy, Netherlands, Hungary, Finland, Portugal and France.

The following measurements are included in the OFFICAIR database:

Aldehydes, VOCs, NO2 and O3, airflow measurement and continuous monitoring of temperature and relative humidity, at 5 locations per building (4 indoor and 1 outdoor) + PM2.5 (in 1 room for 5 days) in two seasons (summer 2012 and winter 2012-13).

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

Aldehydes, VOCs, NO₂ and O₃ were measured via passive sampling method for 5 days from Monday to Friday. Analysis was performed in central certified labs.

The corresponding web page of the pollutant concentration described above is: <http://web.jrc.ec.europa.eu/airmex/officair-db/viewdata/index.cfm>.

C

Data for outdoor air for several environmental stressors for the region of Toscana (Italy) for the municipalities: Arezzo, Firenze, Grosseto, Livorno, Lucca, Massa, Pisa, Prato, Pistoia and Siena. Data cover the period from **01/01/2009-31/12/2013**. Data referred to quantified stressor concentration in air. The environmental stressors that data covered are: PM₁₀, PM_{2.5}, VOCs, organic compounds, NO₂, O₃, SO₂, CO, H₂S and PAHs. The units and time resolution of the data are:

H₂S: hourly average; µg/m³ at 20°C

NO₂: hourly average; µg/m³ at 20°C

NO: hourly average; µg/m³ at 20°C

NO_x: hourly average; µg-NO₂/m³ at 20°C

O₃: hourly average; µg/m³ at 20°C

BTEX: hourly average; µg/m³ at 20°C

Alkane (n-hexane, n-heptane, n-octane): hourly average; µg/m³ at 20°C

SO₂: hourly average; µg/m³ at 20°C

CO: hourly average; mg/m³ at 20°C

PM₁₀: daily average, µg/m³ at ambient temperature


PM_{2.5}: daily average, µg/m³ at ambient temperature

The corresponding web page of the pollutant concentration described above is: <http://www.arpat.toscana.it/datiemappe/dati/qualita-dellaria-dati-orari>.

D

Data for outdoor air for the environmental stressor of pollen for the region of Toscana (Italy) for the municipalities: Firenze, Pistoia, Montecatini Terme (Pistoia) and Lido di Camaiore (Lucca). Data cover the period from **01/01/1996-31/12/2012**. The daily data of the concentration of the main pollen and fungal spores are expressed in particles/m³.

The Environmental Protection Agency of Tuscany Region (ARPAT, www.arpat.toscana.it) makes available the data of the pollens concentration from 1996 to 2012, in particular: Firenze (2000-2012), Pistoia (1996-2012), Montecatini Terme (1996-2012) and Lido di Camaiore (2001-2012).

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

Data come from Monitoring Stations installed on the regional territory. Data, divided into municipalities (Firenze: fi1; Pistoia: pt1; Montecatini Terme: pt2-montecatini; Lido di Camaiore: lu1-lido-di-camaiore) and years, are available for download in zipped xls. In every file there are 4 columns containing the following information: date (dd/mm/yyyy), station (location), inspected pollen, concentration (particles/m3).

The corresponding web page of the pollen concentration in air described above is: http://www.arpat.toscana.it/datiemappe/dati/dati-concentrazioni-pollini-e-spore-fungine-in-toscana-anni-1996-2012/attachment_download/csv

E

Data for indoor air at schools for several environmental stressors for the region of Porto (Portugal). The environmental stressors that data covered are: PM10, PM2.5, VOCs, mold formaldehyde, acetaldehyde, total bacteria and Gram negative bacteria. Data cover the period from **11/11/2011-20/12/2012**. The data units are:

PM2.5, PM10: mg/m3

VOCs: µg/m3

Aldehydes (formaldehyde and acetaldehyde): µg/m3

Total fungi, total bacteria and Gram negative bacteria: CFU/m3

Air sampling was performed concurrently in two heating season periods: November 2011 to March 2012 and November 2012 to December 2012.

PM2.5, PM10: The measurements were conducted 24 hours in each classroom at 1 minute intervals, but it was considered the data regarding teaching periods.

VOCs: The passive air sampling was conducted during a scholar week, from Monday morning to Friday afternoon.

Aldehydes (formaldehyde and acetaldehyde): Samplers were placed in each classroom on Monday morning and collected on Friday afternoon.


Total fungi, total bacteria and Gram negative bacteria: Air was drawn through the sampler at 100 l/min, and sequential duplicate air samples (duplicates of 100 and 250 litres) were collected indoors and outdoors between 9.30 a.m. and 12.00 a.m.

The data are uploaded to the EDMS as a single file.

F

Contains data for indoor air at residences (urban and suburban area) for several environmental stressors for the region of Porto (Portugal). The environmental stressors that data covered are: PM10, PM2.5, VOCs, mold formaldehyde, acetaldehyde, total bacteria and Gram negative bacteria. Data cover the period from **01/11/2012-31/03/2013**. The data units are:

PM2.5, PM10: mg/m3

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

VOCs: $\mu\text{g}/\text{m}^3$

Aldehydes (formaldehyde and acetaldehyde): $\mu\text{g}/\text{m}^3$

Total fungi, total bacteria and Gram negative bacteria: CFU/m³

Air sampling took place from November 2012 to March 2013.

PM_{2.5}, PM₁₀: The measurements were conducted 24 hours in each bedroom at 1 minute intervals.

VOCs and Aldehydes (formaldehyde and acetaldehyde): The passive air sampling was conducted over a total of 7 consecutive 24-hours periods, from Saturday to Friday.

Total fungi, total bacteria and Gram negative bacteria: Air was drawn through the sampler at 100 l/min, and sequential duplicate air samples (duplicates of 100 and 250 litres) were collected indoors on Saturday. In each investigated home the child's bedroom were selected for air sampling.

The data are uploaded to the EDMS as a single file.

G

Contains data for CO₂, temperature and RH% of outdoor air at school grounds for the region of Porto (Portugal). Data cover the period from **01/11/2011-20/12/2012**. The data units are:

Temperature: °C

Relative humidity: %

CO₂: ppm

Outdoor air sampling was performed concurrently in two heating season periods: November 2011 to March 2012 and November 2012 to December 2012. The air sampling covered a 5-day period (school week, from Monday morning until Friday afternoon) during teaching periods, with a time step of 5 minutes.

The data are uploaded to the EDMS as a single file.

H


Data for outdoor air at school grounds for several environmental stressors for the region of Porto (Portugal). The environmental stressors that data covered are: PM₁₀, PM_{2.5}, VOCs, mold, CO formaldehyde, acetaldehyde, total bacteria. Data cover the period from **01/11/2011-20/12/2012**. The data units are:

CO: mg/m³

PM_{2.5}, PM₁₀: $\mu\text{g}/\text{m}^3$

VOCs: $\mu\text{g}/\text{m}^3$

Aldehydes (formaldehyde, acetaldehyde): $\mu\text{g}/\text{m}^3$

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

Total fungi and total bacteria: CFU/m³

Outdoor air sampling was performed concurrently in two heating season periods: November 2011 to March 2012 and November 2012 to December 2012.

The data are uploaded to the EDMS as a single file.

I
Contains data for pollutant concentration of outdoor air at residences grounds (urban and suburban area) for several environmental stressors for the region of Porto (Portugal). The environmental stressors that data covered are: PM₁₀, PM_{2.5}, VOCs, mold, CO, formaldehyde, acetaldehyde and total bacteria. Outdoor air sampling took place from **November 2012 to March 2013**. The data units are:

CO: mg/m³

PM_{2.5}, PM₁₀: µg/m³

VOCs: µg/m³

Aldehydes (formaldehyde, acetaldehyde): µg/m³

Total fungi and total bacteria: CFU/m³

CO: The air sampling covered a period of 7 days with a time step of 5 minutes from Saturday to Friday.

PM_{2.5}, PM₁₀: The measurements were conducted 24 hours in each outdoor sampling site at 1 minute intervals.

VOCs and Aldehydes (formaldehyde and acetaldehyde): The passive air sampling was conducted over a total of 7 consecutive 24-hours periods, from Saturday to Friday.

Total fungi and total bacteria: Air was drawn through the sampler at 100 l/min, and sequential duplicate air samples (duplicates of 100 and 250 litres) were collected outdoors on Saturday.

The data are uploaded to the EDMS as a single file.


J
Satellite data on pollution concentration for air. Provide data for NO₂, O₃, SO₂, BrO, HCHO and aerosol in spatial resolution of 80x40km for the period of **10/2006 to 09/2014**.

Total ozone and SO₂column amount [DU] and [molecules/cm²]

Total H₂O column amount [kg/m²] and [molecules/cm²]

Other trace gas column amounts [molecules/cm²]

No Total ozone column product provides information about vertical column densities of ozone in the atmosphere

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

Total and tropospheric NO₂ column products provide information about vertical column densities of nitrogen dioxide in the atmosphere

SO₂ column product provides information about vertical column densities of sulfur dioxide in the atmosphere

Total BrO column product provides information about vertical column densities of bromine monoxide in the atmosphere

Tropospheric HCHO column product provides information about vertical column densities of formaldehyde in the atmosphere

The corresponding web page of pollutant concentration in air described above is: http://o3msaf.fmi.fi/offline_access.html stored.

3.1.7. Pollutant concentration data – soil

Soil contamination is the occurrence of pollutants in soil above a certain level causing a deterioration or loss of one or more soil functions. Also, Soil Contamination can be considered as the presence of man-made chemicals or other alteration in the natural soil environment.


In the EDMS DataBase there are the following data sets:

A

Geochemical Atlas of Europe is the European contribution to the IUGS/IAGC Global Geochemical Baselines Programme, and has been carried out by government institutions from 26 countries (Austria, Albania, Belgium, Croatia, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, The Netherlands, Norway, Poland, Portugal, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, UK) under the auspices of the Forum of European Geological Surveys (FOREGS) The main objectives of this European survey were: 1) to apply standardised methods of sampling, chemical analysis and data management to prepare a geochemical baseline across Europe; and 2) to use this reference network to level national baseline datasets.

Samples of stream water, stream sediment and three types of soil (organic top layer, minerogenic top and sub soil) have been collected at 900 stations, each representing a catchment area of 100 km², corresponding to a sampling density of about one sample per 4700 km². In addition, the uppermost 25 cm of floodplain sediment was sampled from 790 sites each representing a catchment area of 1000 km². Samples have the following characteristics:

- Stream water (filtered and unfiltered);

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

- Stream sediment - mineral sediment (<0.150 mm);
- Residual soil - upper 0-25 cm horizon (topsoil) without the top organic layer (<2 mm);
- Residual soil - lower (C) horizon (subsoil); a 25 cm layer within a depth range of 50-200 cm (<2 mm);
- Humus (where present);
- Overbank sediment - upper 0-25 cm horizon (<0.150 mm, optional);
- Overbank sediment - bottom layer (<0.150 mm, optional);
- Floodplain sediment - upper 0-25 cm horizon (<2 mm), and
- Floodplain sediment - bottom layer (<2 mm, optional).

Data cover the period of **1998 to 2001**.

The corresponding web page of the soil contamination data described above is:
<http://weppi.gtk.fi/publ/foregsatlas/index.php>.

3.1.8. Pollutant concentration data – dust

Unfortunately a large amount of data on pollutants concentration concerning dust covering all European countries not exists. On the other hand, there is scattered information in the literature (journals, European projects) that someone can withdraw the appropriate data for the purpose of the EDMS and HEALS project generally. For this reason and in the frame of WP17 the team of WP8 will gather all the data and populate the EDMS for this particular category.


3.1.9. Pollutant concentration data – water

Water is not a commercial product like any other but, rather, a heritage which must be protected, defended and treated as such, thus the protection of ground and surface water against pollution caused by certain dangerous substances , as part of an overall policy on freshwater protection.(DIRECTIVE 2000/60/EC)

In the EDMS DataBase there are the following data sets:

A

Waterbase contains timely, reliable and policy-relevant data collected from **EEA member countries** through the WISE-SoE data collection (formerly known as Eionet-Water and Eurowaternet) process. WISE-SoE data collection selects validated monitoring data from national databases and adds information on the physical characteristics of the water bodies being monitored and on the pressures potentially affecting water quality. Waterbase contain data on contamination of Europe's rivers, lakes, groundwater bodies and transitional, coastal and marine waters, and on the quantity of Europe's water resources. The added value of Waterbase is that data collected through the WISE-SoE data collection process are from

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

statistically stratified monitoring stations and groundwater bodies and are comparable at the European level. EEA covers data for the following parameters:

Benzo(a)pyrene, Alachlor, Atrazine, chlorfenvinphos, chlorpyrifos, Diuron, gamma-HCH (Lindane), Isoproturon, Simazine, Trifluralin, Aldrin, DDT o,p', DDT p,p', DDE p, p', DDD p,p', Dieldrin, Endrin, Bentazone, alpha- HCH, beta-HCH, alpha-Endosulfan, Desethylatrazine, Desisoproturon, Propazine, Prometryn, Terbutylazine, 2,4-D,MCPA, Mecoprop, Isodrin, Linuron, Mercury, Benzene, 1,2-Dichloroethene, cadmium, Lead, 1,1,2-trichloroethene, 1,1,2,2-tetrachloroethene

Data are in µg/m³ and the temporal coverage is **1960-2012**.

The corresponding web page of the water contamination data described above is: <http://www.eea.europa.eu/data-and-maps/data/waterbase-groundwater-10>.

3.1.10. Pollutant concentration data – drinking water

Whereas, in view of the importance of the quality of water intended for human consumption for human health, it is necessary to establish the essential quality standards with which water intended for that purpose must comply. The parametric values applicable to water intended for human consumption are based on public-health considerations and on a method of assessing risk. (COUNCIL DIRECTIVE 98/83/EC)

In the EDMS DataBase there are the following data sets:

A


Vitens is publishing the quality of the drinking water for the **North of The Netherlands**. The regions of the corresponding data are: Utrecht, Gelderland, Overijssel, Groningen. The environmental stressors that data covered are: toxic metals, organic compounds, organoalogenes, other environmental stressors such as SO₄²⁻, NO₃⁻ etc.

The included link guides the user to measurements of the drinking water for different sampling point through the region. The data corresponded to the latest data of the period (**summer 2014**)

The corresponding web page of the drinking water contamination data described above is: <http://www.vitens.nl/overvitens/water/waterkwaliteit/Paginas/kwaliteitthuis.aspx>.

B

Evides is publishing the quality of the drinking water for the **Middle of the Netherlands**. The regions of the corresponding data are: Zeeland, Zuid-Holland, Brabant. The environmental stressors that data covered are: toxic metals, organic compounds, organoalogenes, other environmental stressors such as SO₄²⁻, NO₃⁻ etc.

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

The included link guides the user to measurements of the drinking water for different sampling point through the region. The data corresponded to the latest data of the period **(June 2013 - June 2014)**.

The corresponding web page of the drinking water contamination data described above is: <http://www.evides.nl/drinkwater/Documents/Waterkwaliteit/Tabel%20Drinkwaters.htm>.

3.1.11. Food concentration data

A variety of chemicals such as dioxins and polychlorinated biphenyls (PCBs) PAHs, metals etc., are toxic that persist in the environment and accumulate in the food chain. They can cause adverse effects on humans (e.g. nervous, immune and endocrine system, impair reproductive function and may cause cancer). Because of potential high dietary exposure in some population groups, food contaminants represent an important public health issue at the European level.

In the EDMS DataBase there are the following data sets:

A


The EFSA Comprehensive European Food Consumption Database (Comprehensive Database) has been built from existing national information on food consumption at a detailed level. Competent organisations in the European Union's Member States provided EFSA with data from those most recent national dietary survey in their country, at the level of consumption by the individual consumer. This included food consumption data concerning infants (2 surveys from 2 Member States), toddlers (8 surveys from 8 Member States), children (16 surveys from 14 Member States), adolescents (14 surveys from 12 Member States), adults (21 surveys from 20 Member States), elderly (9 surveys from 9 Member States) and very elderly (8 surveys from 8 Member States) for a total of 32 different dietary surveys carried out in 22 different Member States.

EFSA's Founding regulation (Art. 33) states that EFSA shall search for, collect, analyze and summarize particularly data on

- food consumption,
- incidence and prevalence of biological risks, and
- occurrence of contaminants and chemical residues.

EFSA covers food consumption data for the following categories: Beverages alcoholic and non-alcoholic, bread and rolls, bakery products, cereals and products, flour, pasta, eggs, cheese, milk, milk products, total added lipids, fish and sea food, nuts, fruits, meat and products, potatoes, vegetables, pulses, sugar, sugar products, juices

Summary statistics from the database enable quick screening for chronic and acute exposure to substances that may be found in the food chain. In the database, dietary surveys and food consumption data for each country are divided by category. These include: age,

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

from infants to adults aged 75 years or older; food group (nearly 160) and type of consumption, covering both regular and high consumption thus allowing calculations to be tailored to each category of consumer.

The statistics on food consumption are reported in grams per day, as well as grams per day per kg of body weight.

EFSA covers data for the time interval: **1997-2008**.

The corresponding web page of the food consumption data described above is: <http://www.efsa.europa.eu/en/datexfoodcdb/datexfooddb.htm>.

B

Food consumption data for the region of Catalunya (Spain). Data cover the period of **01/01/2008 to 31/12/2008**. Data are presented as a table with the daily food intake data (gr/day) by type of product. Data are divided in gender and age groups.

The data are uploaded to the EDMS as a single file.

3.1.12. Population data (age/gender/etc)

Data about population density disaggregated by gender and age class. These data are non-individual data.


In the EDMS DataBase there are the following data sets:

A

Population data for Italy covering all regions and municipalities. Available themes are the following: Environment and Energy, Population and Households, Households Economic Conditions and Disparities, Health statistics, Social Security and Welfare, Education and training, Communication, culture, leisure and time use, Justice and Security, Citizens' opinions and satisfaction with life, Social participation, National Accounts, Enterprises, Agriculture, Industry and Construction Services, Public Administrations and Private Institutions, External Trade and Internationalisation, Prices and Labour. Statistics are organised by theme in a two-level hierarchical tree. Data are presented in aggregate form in multidimensional tables; acting on variables, reference periods and the arrangement of heads and sides users can create custom tables and graphs. The system can be searched by keyword, theme and region.

A wide range of standard metadata facilitates the retrieval and understanding of statistics by users.

I.Stat's contents are constantly upgraded. At present, the system is undergoing a population and a refinement process. Therefore, it includes a subset of data currently provided by

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

ISTAT in various forms (databases, datasets, etc.). I.Stat is the warehouse of statistics produced by Italian National Institute of Statistics (Istituto nazionale di statistica- Istat), a complete and homogeneous wealth of information unique for the Italian official statistics.

The corresponding web page of the population data described above is: <http://dati.istat.it/Index.aspx?lang=en&SubSessionId=1b60e645-021b-48ef-a6c3-59092f0aa373>.

B

Population data for Italy for the region of Tuscany and municipality of Pisa. Data came from two epidemiological studies performed in subjects living in Pisa (Tuscany): Pisa2 study (1991-93) and IMCA2 study (2008-2011). The studies were carried out by the Pulmonary Environmental Epidemiology Unit, CNR Institute of Clinical Physiology.

A questionnaire on socio-demographic characteristics, respiratory symptoms/diseases and risk factors was used. A selection of the databases containing demographic information is reported in the template. Additional information, collected through the questionnaire, about risk factors (e.g. smoking habits, occupational exposure, environmental exposure) and health status are available.

In every file there are columns containing the following information: list (participant ID), age, gender (1: male; 2: female), education (1: 0-8 years of study; 2: >8 and ≤ 13 years; 3: >13 years) and crowding (crowding index: total number of co-residents per household divided by the total number of rooms, calculated only in Pisa2 study).

The data are uploaded to the EDMS as two single files.

C

Population data for various European countries. Data cover the period of **01/01/1960** to **01/01/2013**. Population (number) presented by calendar year, single year of age and gender. Available data for EU28, EU27, EA18, EA17, EA16, separate countries of EU28, EEA31, EEA30, EFTA


The corresponding web page of the population data described above is: http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=demo_pjan&lang=en.

3.1.13. Noise data

EDMS wish to contained data about noise per location.

In the EDMS DataBase until now there are the following data sets:

A

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

Noise data for various European countries for outdoor urban and suburban location with the following characteristics: near highways (average speed vehicle >80km/h), railway, airport and agglomerations for the years **2007** and **2012**. Noise exposure data available in excel files, categorized in excel sheets (2012 & 2007)

Results are expressed in Nr of people exposed to different noise bands (Lden) (Lnight)

% of people exposed to different noise bands (Lden) (Lnight)

Lden : Day-evening-night level

Lnight : Night-time noise indicator

Data are available per various cities in European countries, Also summarized data are presented per country.

The data were reported by member states and member countries corresponding to 2005 and 2007 deliveries, and to 2010 (2008) and 2012 deliveries, updated up to 28th of August 2013.

Major roads, railways, airports and agglomerations follow the below thresholds:

Agglomerations > 100,000 inhabitants.

Major civil airport > 50,000 movements/year

All major roads > 3 million vehicles/year

All major railways > 30,000 trains/year

The corresponding web page of the noise data described above is:
<http://noise.eionet.europa.eu/downloads.html>.

B

The Noise Navigator is spreadsheet of sound levels for more than 1700 occupational, recreational, and military noise sources (outdoor and indoor locations).


The tabled values are primarily A-weighted sound levels, as opposed to time-weighted average levels or Leqs.

The data are separated into groups by categories as shown on the Worksheet tabs.

Category 1:

Construction, Farm, Firearms and Explosives, Hand tools, Hearing testing, Household, Industrial, Logging, Machinery, Marine, Medical, Mining, Music/Recording, Nature, Office, Personal, Power Tools, Public, Recreation, Rocket

Schools, Toys, Transportation, Warning, Welding

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

Category 2:

Aircraft, Appliances, Automotive, Boat, Concert, Construction, Disco/Music Club, Earphones, Exercise, Fishing, Hobbies, Live Show, Machinery, Military, Motorcycle

Movies/Theater, Music, Open pit (mining), Racing, Recreation, Restaurant/Bar, Snowmobile, Sporting event, Subway, Toys, Traffic, Train, Underground (mining), Yard & Garden

The values in this spreadsheet can be sorted according to noise source, category of noise, and sound pressure level, or by any of the other columns, using the sort function in Excel.

The data are uploaded to the EDMS as a single file.

3.1.14. Building characteristic data

This section concerning data related to micro-environments relates to indoor spaces such as places of residence, workplace, kindergarten or school.

In the EDMS DataBase until now there are the following data sets:


A

Building characteristics for school buildings, at the municipality of Porto (Portugal) for the period of **01/11/2011 to 20/12/2012**. Walkthrough inspection and checklist was performed for each school concurrently in two heating season periods: November 2011 to March 2012 and November 2012 to December 2012. A walkthrough inspection was completed for each school by a trained researcher to gather information related to school building characteristics involved in study. In the walkthrough inspection a standardized checklist was used that has been previously field-tested in SINPHONIE project and was completed with the assistance of principals and teachers. The checklist includes a large set of questions related to the following main topics: outdoor characterization, building construction characteristics, ventilation and heating systems, past occurrences or visible problems and building use and potential indoor sources.

The data are uploaded to the EDMS as a single file.

B

Building characteristics for classrooms, at the municipality of Porto (Portugal) for the period of **01/11/2011 to 20/12/2012**. Walkthrough inspection and checklist was performed for each classroom concurrently in two heating season periods: November 2011 to March 2012 and November 2012 to December 2012. A walkthrough inspection was completed for each classroom by a trained researcher to gather information related to characteristics of the selected classrooms involved in study. In the walkthrough inspection a standardized checklist

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

was used that has been previously field-tested in SINPHONIE project and was completed with the assistance of principals and teachers. A specific checklist form was filled in for each classroom identifying all relevant information such as the area, finishing materials and their conditions concerning floor, walls, and ceiling; windows; past occurrences and visible problems, heating and ventilation system, scholar activity products (paintings, glues, etc.) maintenance routines and cleaning procedures. The type of classroom furniture, the presence of chalkboards, copiers and plants were also noted, as well as information about environmental modifiers including air fresheners and insecticides.

The data are uploaded to the EDMS as a single file.

C

Building characteristics for homes and bedrooms, at the municipality of Porto (Portugal) for the period of **01/11/2012 to 31/03/2013**. Walkthrough inspection and checklist was performed for each home and bedroom between November 2012 and March 2013. A walkthrough inspection was completed for each home and bedroom by a trained researcher to gather information related to characteristics of the selected building involved in study. A well-trained technician completed a standardized walkthrough inspection and filled in a detailed checklist to collect information about outdoor characteristics, building characteristics and observations on the dwelling's interior and its possible moisture damage, mould problems, presence of pets and carpeting throughout the home, cleaning practices and characteristics of ventilation and heating, cooking devices, indoor smoking habits, emission sources such as candles, incense and room deodorizers, etc.

The data are uploaded to the EDMS as a single file.

3.2. Template of EDMS input data

There are two categories of potential data for the purpose of the HEALS project:


- The Existing Data
- HEALS created Data

The methodology followed for creating the HEALS EDMS was first to Transfer/Fill data into a common template in xls/xml format and then to feed these data into the database. As decided an xml template was created in order to fulfill the above methodology and covered the potential data.

The first thoughts on the EDMS Template contain the following aspects:

Template Field Attributes

- Data Label (Attribute Name)
- Data Type (Classification, Standard etc.)
- Availability (Always given, where applicable - voidable)

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

Common Template Conceptual Schema Design (the process of generating a description of the contents of a database in high-level terms that are natural and direct for users of the database)

- Try to produce a common template for all data –if possible:
 - better management
 - facilitating interrelationships
- Irrelevant entries are left empty or marked N/A
- To cope data diversity, degree of freedom in filling :
 - Data Information
 - Measurement Information
 - QA/QC Procedure

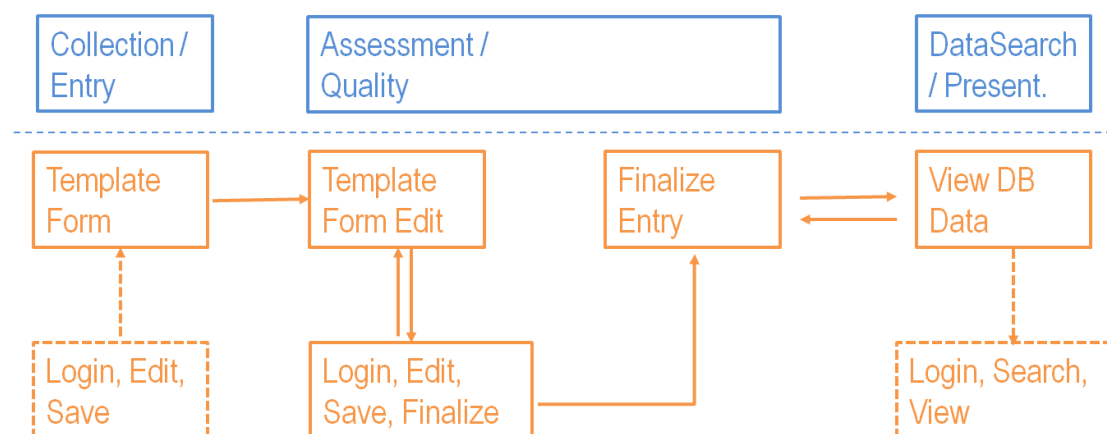



Figure 2. EDMS – STEP 2 – Insert Data via (Common) Data Template(s)

The EDMS features three user categories. There is the simple user type which logs into the system and can perform queries on the stored data, the partners which can also contribute by submitting data sets that could be potentially stored in the data base and the editors, which apart from contributing data sets, review and approve what partners have contributed. Figure 2 depicts the EDMSs structure and core functionality as a flowchart. Under the Collection / Entry label we have the actions the partners can perform (login, edit, save). The partners can fill a Template Form and submit it for consideration. Under the Assessment / Quality label we can see the actions an editor can perform. The editor, asynchronously, can review and potentially approve (finalize) entries submitted from any partner. Only then these data are available and can be retrieved by queries simple users make. The latter is described under the Data Search label where there is the interaction a simple user can perform with the EDMS (login, search and view).

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

Data Collection Entry – Template Form

The following sections are covered by the template form that every project partner need to fill in order to provide an environmental data set that is interested for the HEALS project:

1. **Data Class**
2. **Environmental Stressors**
3. **Location**
4. Medium/Medium Characteristics
5. **Data Information**
6. **Data Source / Web link / upload single file**
7. **Measurement methodology**
8. **QA/QC Procedure**
9. **SUBMITTING PARTNER**


In Annex I, is presented the manual provided to the submitted partners in order to fill correctly or clarify issues of the template form.

The EDMS of HEALS project contains data on the following environmental categories of interest (**Data class**) as it is mentioned previously:

- Land use/Land cover
- Meteorological Data
- Comfort data indoors (temperature-relative humidity)
- Air emission data
- Air emission data – consumer products
- Pollutant concentration data – air
- Pollutant concentration data – soil
- Pollutant concentration data – dust
- Pollutant concentration data – water
- Pollutant concentration data – drinking water
- Food concentration data
- Population data (age/gender/etc)
- Noise data
- Building characteristic data

Under the **environmental stressor** the EDMS includes the priority pollutants of the HEALS project or other environmental parameter relative to humans which are:

- PM10, PM2.5
- VOCs
- Phthalates
- Toxic metals

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

- Organic compounds
- Organoallogens
- Dust mites
- Pesticides
- Endocrine disruptors
- POPs
- NO₂
- O₃
- Mold
- Pollen
- Noise

Also data providers to the system can enter another environmental stressor that is interest to the project and cannot listed in the above list due to space limit.


The **location** part of the data insert template form provides information about the spatial distribution of the corresponding data such as country of origin of the data, geographical region, longitude, latitude etc., while the **medium characteristics** provides the specific medium (air, soil, water, food etc.) that the data refer.

In the **Data information** part of the template, is listed the temporal dimension, the time resolution, the measurement units of the corresponding data along with any useful information on the provided data for better understanding the way of the data presented in the EDMS.

The **Data Source / Web link** section contains the active web link for the data that is treated and accessed outside the HEALS EDMS as well as the umbrella web link that the data came from. Also in this section there is the option of upload to the HEALS database a single file containing data of interest that someone can provide to the HEALS project.

In the **measurement methodology** part it is written the method used for obtained the data and information on the whole procedure for obtaining the specific data such as equipment used, sampling techniques etc., while in the section of QA/QC the data that are imported or access through external web links are accompanied with all the quality control and quality assurance procedures that are available.

Finally in the **submitting partner** section is listed the identity of the data provider. In figure 3 is showed the xml form of the data template for the EDMS HEALS database. In Annexes III-VII there is an analytical presentation of the EDMS template web form.

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining		Security:
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

Choose from the options below.

A. Data Class

SELECT AN OPTION ▼

B. Environmental Stressors

<input type="checkbox"/> N/A	<input type="checkbox"/> PM ₁₀	<input type="checkbox"/> PM _{2.5}	<input type="checkbox"/> VOCs	<input type="checkbox"/> Phthalates
<input type="checkbox"/> Toxic Metals	<input type="checkbox"/> Organic compounds	<input type="checkbox"/> Organoallogens	<input type="checkbox"/> Dust mites	<input type="checkbox"/> Mold
<input type="checkbox"/> Pollen	<input type="checkbox"/> Pesticides	<input type="checkbox"/> Endocrine disruptors	<input type="checkbox"/> POPs	<input type="checkbox"/> NO ₂
<input type="checkbox"/> O ₃	<input type="checkbox"/> Noise	<input type="checkbox"/> Other (please specify)		

C. Location

Country	Region(s)	Municipality(ties)
SELECT AN OPTION ▼	<input type="text"/>	<input type="text"/>

Location Characteristics

Ambient Environment Type	Microenvironment Class
Click and select one or more values	<input type="checkbox"/> Outdoors <input type="checkbox"/> Indoors <input type="checkbox"/> In vehicle <input type="checkbox"/> Unspecified <input type="checkbox"/> Other (please specify)

Longitude	Latitude	Altitude	Above Ground Level
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>


Spatial resolution

SELECT AN OPTION ▼

Figure 3. Data Template form through the website of EDMS.

3.3. Dataset selection procedure

HEALS EDMS database was constructed under the HEALS project Work Package No. 8 “Environmental data mining”. The compilation of this database is based on an analytical description and classification of environmental stressors data that humans are exposed- (particulates, organic compounds, toxic metals, physical parameters etc) throughout the enlarged European Union. This was aided by national and European databases, consumer’s protection organizations and literature review as well as the existing data from other EU projects (EPHECT, OFFICAIR EXPOLIS, etc.). The literature review was conducted continuously during the work package duration, from scientific publications on environmental

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

stressors that potentially affecting the human health (directly and indirectly), relevant European project and the corresponding European database.

The fulfilment of the HEALS EDMS database with emission data was obtained by reviewing data of several sources. The scientific data on experimental studies about environmental stressors was reviewed by using search engines in the Internet and by searching from the web pages of relevant journals. In addition to electronic search, numerous study reports concerning emissions were also checked. Finally, other similar databases containing relevant data were reviewed and assessed, as well.

The criteria followed for assessed the data collected in Task 8.1 and consider them for populated the HEALS EDMS database are summarized in the following remarks:


- **The main focus of the review was on recent campaigns/data -studies (from 1995 and on for work on journals) in order to have the ability to evaluate more reliable data.**
- **The literature review limited to peer-reviewed journals after the year of 1995 in order to have newly and more reliable data.**
- **Only – emission- data based on standardized measurement methods described in ISO/ASTM/CEN guidelines were taken into consideration or when an alternative method is used this method should have at least been published in a peer reviewed journal.**
- **Data on various environmental stressors were assessed only from authorized organizations (European, international or global) and national public agents.**
- **Data were illustrated the description of sampling and analytical techniques will be the first priority to include them in the database. In that context the calibration procedure for the instruments/monitors used are preferably to be referred.**

4. DATA DERIVATION METHODOLOGY ASSESSMENT

In the next pages is presented for every data category listed in the EDMS platform the methodology assessment of the data mined until now and are populated into the EDMS database. The methodology assessment is based on the corresponding section of the EDMS template (**Part G** of the template form) filled by the partner searched/mined/provided the data.

4.1. Land use/Land cover

For the data set covering the global land 1-km AVHRR dataset consists a network of 29 high resolution picture transmission (HRPT) stations, along with data recorded by the National Oceanic and Atmospheric Administration (NOAA), has been acquiring daily global land

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

coverage since 1992 until 1994. Under the guidance of the International Geosphere Biosphere Programme, processing standards for the AVHRR data have been developed for calibration, atmospheric correction, geometric registration, and the production of global 10-day maximum normalized difference vegetation index (NDVI) composites.

Data acquisition and archive management of the HRPT data received in real time at the EDC, the NOAA LAC data received via the DOMSAT receiving systems, and the data acquired by the ground station network are conducted on a central computer system at the EDC. The system is capable of ingesting, archiving, creating metadata, and creating digital quicklooks for 100 images per day. The system is equipped with a wide variety of magnetic tape devices, a communication network to facilitate data ingest, and a sufficient processing capability for large volumes of data.


Acquisition of HRPT and LAC data is a standardized and routine process. However, nearly all of the ground receiving stations use different data and media formats. The EDC receives the large volume of data on a wide variety of media including 4- and 8-mm, 3480 cartridges, and 9-track tapes. Therefore, the EDC wrote ingest procedures specific to each station's format from sample AVHRR HRPT data obtained from each station. The ingest routines create a Level 0 archive format image file along with a CEOS Inventory Exchange Format (IEF) header file that has a three-letter code specific to each receiving station.

All data provided by the ground station network are logged in a data tracking system. Information such as the CEOS code of the contributor station, the number of images per media, and the media type are recorded. After the data are ingested the original contributor tape media are placed in an environmentally controlled archive.

The ingest routine creates a Level 0 archive copy of the data, metadata, quicklook microimage fiche, and digital quicklook images. The Level 0 archive copy is stored on 3480 tape cartridges for permanent archive. The metadata are checked for coordinate accuracy, satellite direction, day versus night determination, and acquisition date and time. Channel two (near-infrared) is used for digital quicklook images for daytime scenes while channel four (thermal) is utilized for night scenes. The digital quicklook image is scaled to 8-bits and subsampled every fourth line and fifth sample. Compression routines of the Joint Photographic Experts Group are implemented to realize between a 10 and 13 times data reduction using lossy compression (Nelson 1991). The size of a typical browse file is between 20,000 and 40,000 bytes after compression. The metadata and browse files are transferred to an **Information Management System (IMS)** where data base management routines are used to check for data quality, data gaps and inconsistencies in the metadata.

4.2. Meteorological Data

A

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100


The ECA dataset consists of daily station series obtained from climatological divisions of National Meteorological and Hydrological Services and station series maintained by observatories and research centres throughout Europe and the Mediterranean. ECAD covers the following elements: mean temperature (0.1 C), maximum temperature (0.1 C), minimum temperature (0.1 C), sunshine (0.1 hours), snow depth (1cm), precipitation amount (0,1 mm) , sea level pressure (0.1 hPa), humidity (1%) , wind gust (0.1 m/s), wind speed (0.1 m/s) , wind direction (degrees), cloud cover (oktas). Data regarding the above parameters exists on daily basis from 1850 to 2013.

The ECA dataset consists of daily station series obtained from climatological divisions of National Meteorological and Hydrological Services and station series maintained by observatories and research centers throughout Europe and the Mediterranean.

B)

Data from NASA, through its' Science Mission Directorate, has long supported satellite systems and research providing data important to the study of climate and climate processes. These data include long-term estimates of meteorological quantities and surface solar energy fluxes. These satellite and modeled based products have been shown to be accurate enough to provide reliable solar and meteorological resource data over regions where surface measurements are sparse or nonexistent, and offer two unique features - the data is global and, in general, contiguous in time. Surface meteorology and Solar Energy (SSE) dataset covers data for the following parameters: Daylight cloud amount, Cloud amount at 3-hourly intervals, Frequency of cloud amount at 3-hourly, Air Temperature, Daily Temperature Range, Cooling Degree Days above 18° C, Heating Degree Days below 18° C, Arctic Heating Degree Days below 10° C, Arctic Heating Degree Days below 0°, Earth Skin Temperature, Daily Mean Earth Temperature minimum, maximum and Frost Days Dew/Frost Point Temperature, Wind speed at 50 m, Percent of time for ranges of wind speed at 50, Wind Speed at 50 m at 3-hourly intervals, Wind Direction at 50 m, Wind Direction at 50 m at 3-hourly intervals, Wind Speed at 10 m for terrain similar to airports, Wind Speed adjustments for height and vegetation type. Data regarding the above parameters exists on daily basis from 1983-2005.

The global distribution of meteorological parameters in the SSE archive (e.g dew/frost point minimum, maximum and daily averaged temperatures, relative humidity, and surface pressure) are taken directly from or calculated based upon parameters in NASA's Global Model and Assimilation Office (GMAO), Goddard Earth Observing System global assimilation model version 4 (GEOS-4)

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

(<http://gmao.gsfc.nasa.gov/systems/geos4/>). Relative humidity is a calculated parameter based upon pressure, temperature and specific humidity, all parameters obtained from the assimilation model. Dew/frost point temperatures are calculated values based upon the relative humidity and air temperature which is obtained from the assimilation model. Precipitation data has been obtained from the Global Precipitation Climate Project (GPCP - <http://precip.gsfc.nasa.gov/>). The GPCP precipitation data product, Version 2.1, is a global 2.5° x 2.5° monthly accumulation based upon combination of observations from multiple platforms. The one degree SSE estimates of precipitation are based upon replicating GPCP values for SSE cells that overlap GPCP cells and averaging GPCP values when the SSE cell overlaps two or more GPCP cells. Monthly mean wind speed data is based upon the NASA/GMAO GEOS version 1 (GEOS-1) for the time period July 1983 –June 1993.

The meteorological data is on a 1 degree longitude by 1 degree latitude equal-angle grid covering the entire globe (64,800 regions). The image on the right is a detailed sample of the grid covering England. The data is generated using the NASA Goddard Earth Observing System - Version 4 (GEOS 4) Multiyear Assimilation Timeseries Data. The GEOS 4 data set has a spacing of 1.25 degrees of longitude by 1 degree of latitude. Bilinear interpolation is used to produce 1 by 1 degree regions.


4.3. Comfort data indoors (temperature-relative humidity)

There is no Data stored in the HEALS EDMS database so there are no quality assessment procedures yet.

4.4. Air emission data

A

The BUMA project aims to thoroughly assess the human exposure to air hazards emitted by building materials commonly used in Europe. Indoor organic compounds are released from a variety of building materials including vinyl tiles and coving, carpets, particleboards, wood products, paints, adhesives etc. VOCs associated with paints, varnishes, sealing caulks, adhesives, carpets and other materials are likely to be a major cause of health complaints associated with indoor air in new buildings. The database includes the Emission Rates ($\mu\text{g}/\text{h}/\text{m}^2$) from materials in the following categories: Adhesives, floorings, gypsum products and plaster boards, internal and external wall ceiling finishes, thermal insulation products, wood based panels, carpets, paints and varnishes.

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

The BUMA database is updated with emission data from studies carried out by certificated institutions during the project period, as well as by peer reviewed literature published after the year of 1995.

4.5. Air emission data – consumer products

A

The BUMAC database was set-up as a compilation of data on the current state-of-the-art on consumer product compositions and emissions, on test chamber experimental results, data on ventilation patterns and reported indoor condition values (test conditions such as temperature, relative humidity, etc) exposures, risks and health end points. Qualitative data were assured by using only data outcomes from procedures derived from standardized emission test protocols. Includes gaseous and particulate matter emissions, secondary reactions and degradations of coated surfaces. BUMAC classifies, organizes and presents in a user-friendly and comprehensive way, the emission data from commonly used consumer products and their emitted compounds. The BUMAC database is continuously updated with emission data from studies carried out by certificated institutions as well as by peer reviewed literature focusing more on data for health relevant compounds. The database also includes data from emission studies which were carried out in the frame of EPHECT on 15 selected consumer products.

The BUMAC database contains for every data entry the conditions of the experiment of the product the sampling technique and duration and the analytical principle and equipment used for the quantification of the environmental stressor released by each product.


For example of chamber test conditions:

Parameter	Value (range)	Unit
Chamber volume:	20.24	m ³
Wall type:	Steel	-
Temperature:	23±1	oC
Relative humidity:	18±1	%
Ozone concentration:	50±5	ppb
Air exchange rate:	0.5±0.1	1/h
Mean air velocity*:	4.80E-04	m/s

*) Calculated from air exchange rate and room dimensions

Product amount used: 0.07* g/h

*) Emission rate

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

Type of application: The unit was placed inside the chamber and turned on (maximum setting). Sampling was carried out continuously over a period of 5 hours - stable VOC concentrations were reached after 3 hours. The unit was weighed before and after the experiment. A total amount of 0.34 g was released from the unit during the experiment.

Equipment used for the BUMAC database values are GC/MSD, GC-FID, HPLC/DAD.

4.6. Pollutant concentration data – air

A

EEA database for the majority of European countries included the EU countries for outdoor air quality measurement on various environmental stressors.

Measurement methodology (Analytical/method principal)

SO₂: UV fluorescence

PM_x, TSP: Beta ray attenuation, gravimetric analysis, Tapered Element Oscillating Microbalance (TEOM)

NO, NO₂, NO_x: Chemiluminescence

BS: Reflectometry

O₃:Ultraviolet (UV) photometry

Toxic metals: Atomic absorption spectrometry (AAS)

VOC (Benzene,toluene, ethylbenzene, o-m-p-xylene): Gas chromatography with photo ionization detector

Equipment used:

SO₂:

Thermo model 43c-TL SO₂ analyser (UV fluorescence),

Horiba model APSA 360E SO₂ analyser,

Horiba model APSA 350 SO₂ analyser,

Horiba model APSA 300 SO₂ analyser,


DASIBI 4108 SO₂ analyser

PM₁₀:

Thermo Andersen ESM FH 62 I-R (Beta ray attenuation),

Low Volume sampler(gravimetric),

TEOM 1400A

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

NO, NO₂, NO_x:

Horiba model APNA 360 NO_x analyser (Chemiluminescence),
Thermo model 42C-TL (Trace level Nox), DASIBI 2108 NO_x analyser,
Thermo model 14B chemiluminescence NO-NO₂-Nox,
ENVIRONMENT ZC 32M

CO:

Horiba model APMA 360 CO analyser (Non-dispersive infrared spectroscopy (NDIR)),
DASIBI 3008 CO analyser,
Thermo model 48c CO analyser

O₃:

DASIBI 1008 O₃ analyser (Ultraviolet (UV) photometry),
Horiba model APOA 300 O₃ analyser,
Thermo model 49 O₃ analyser,
Environnement S.A. Model O342M UV Ozone Analyze

BS:

ENVIRONMENT SA FILTROMAT COLLECTOR


TSP:

High Volume Sampler (gravimetric),
WEDDING (Beta ray attenuation)

Toxic metals:

Low volume Sampler (AAS)

VOC:

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

SINTECH SPECTRAS BTX GC 855 series

Sampling characteristics:

The height of sampling point varies from sampling station from 3-8m and 10-35m. For BS data the sampling height varies between 400-1000m

The length of the sampling line is 200m for BS and 2m for the other measurements location of the sampling point varies: free air flow, courtyard, facade of building, kerbside (VOC)

B

In OFFICAIR project were monitored 37 office buildings in 8 countries: Greece, Spain, Italy, Netherlands, Hungary, Finland, Portugal and France.

Measurement methodology (Analytical/method principal)

For the sampling of aldehydes: DNPH passive sampler Radiello: blue diffusive body code 120-1 + cartridge code 165, Analysis: LC, ISO 16000-4

For the sampling of VOCs: passive sampler Radiello: diffusive body code 120-2 (reusable tubes) + cartridge code 145, Analysis: thermodesorption, GC, ISO 16017-2

For the sampling of NO₂: HDPE circular diffusive sampler Gradko using spectrophotometry analysis

For the sampling of O₃: passive sampler Radiello: diffusive body code 120-1+ cartridge code 172, Analysis: spectrophotometry, EN 14412 and EN 13528-3


For the sampling of PM_{2.5}: Gravimetric

Equipment used:

PM_{2.5} have been measured gravimetrically in one location per instrumented building (+ outdoor). Sampling via a low-flow pump and a size-selective device (impactor or cyclone) which operate at a constant flow rate to provide a 50% cut-off point at 2.5 µm.

Sampling characteristics:

Sampling procedure was according the project's detailed investigation manual and quality assurance plan.

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

C

The Environmental Protection Agency of Tuscany Region (ARPAT, www.arpat.toscana.it) makes available the data of the above mentioned pollutants from 2009 to 2012.

Measurement methodology (Analytical/method principal)

The official reference methods for the analysis (or sampling) of pollutants are given in Annex VI of the Legislative Decree 155 of 13 August 2010. As point B of the Annex VI, it is permissible to use others methods provided that they have appropriate certification of equivalence.

Reference:

<http://www.normattiva.it/uri-res/N2Ls?urn:nir:stato:decreto.legislativo:2010-08-13;155>

Equipment:

Every equipment interacts with other similar measuring and/or sampling devices within station that are thermostatically controlled at around 25 ° C.

The instruments for the analysis of gaseous pollutants are connected to a common sampling line which conveys the outdoor air inside the monitoring station. The sampling lines for PM analyzers/samplers are distinct from the previous one.

Within each monitoring station there is an acquisition system which allows both the transfer of raw data from the analyzer to the same system both the sending of the measurements to the ARPAT operation center by means of analog or GSM modem.


D

Data from the Environmental Protection Agency of Tuscany Region (ARPAT, www.arpat.toscana.it) makes available the data of the *pollens* concentration from 1996 to 2012, in particular: Firenze (2000-2012), Pistoia (1996-2012), Montecatini Terme (1996-2012) and Lido di Camaiore (2001-2012). Data come from Monitoring Stations installed on the regional territory.

Measurement methodology (Analytical/method principal)

The monitoring stations carry out the continuous sampling throughout the year and they are included in the Italian aerobiology monitoring network Pollnet (<http://www.pollnet.it>).

Equipment:

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

Lanzoni VPPS 2000 samplers (Hirst method).

Sampling characteristics:

ARPAT uses Lanzoni VPPS 2000 samplers with capture on strip for a period of one week, daily separation into segments, staining, detection and counting under an optical microscope.

E

Data on pollutants concentration-air indoor environment (schools) for Portugal.


Measurement methodology (Analytical/method principal)

PM2.5, PM10: Laser photometer based on light scattering technology. The measuring range of the equipment is 0.001 to 150 mg/m³ (1 to 150x10³ µg/m³) with accuracy of ± 0.1% of reading of 0.001 mg/m³ (1 µg/m³). Powered by an internal battery, the DustTrak DRX operated with a flow rate of 3.0 l/min using a built-in diaphragm pump.

VOCs: were collected using a stainless-steel sampling tubes containing Tenax® TA (60/80 mesh). Concentrations of target VOC that included benzene, toluene, m/p-xylene, o-xylene, d-limonene, α-pinene, trichloroethylene, tetrachloroethylene, naphthalene, styrene were investigated. After sampling, Tenax tubes were thermally desorbed (Dani STD 33.50) and quantified using a non-polar column by gas chromatography (Agilent Technologies 6890N) coupled to a mass spectrometry detector (Agilent Technologies 5973), according to ISO 16000, part 6 (2011). Total VOC concentration was quantified using the toluene response factor, and concentrations were calculated as the sum of volatile organic compounds eluting between hexane and hexadecane (included), expressed as toluene.

Aldehydes (formaldehyde and acetaldehyde): were sampled by passive devices Radiello® (RAD 165, Sigma Aldrich), consisting of a cartridge filled with 2,4-dinitrophenylhydrazine (2,4-DNPH). By reaction with 2,4-DNPH, aldehydes give the corresponding 2,4-dinitrophenyllhydrazones, which was desorbed in 2 ml of acetonitrile (Sigma-Aldrich) and manually stirred for 30 minutes. Then, the solution was passed through a polyvinyl difluoride 0.45 mm-syringe filter (13 mm Syringe Filter, Specanalítica) and determined using isocratic reverse phase High Performance Liquid Chromatography (HPLC) (Agilent Technologies, 1220 Infinity LC) with a UV detector operated at 360 nm according to ISO 16000-4 (2011). Aldehydes were identified and quantified by comparison of their retention times and peak areas with those of standard solutions. Each cartridge was sealed after sampling and then brought back to laboratory where it was stored in the refrigerator (<4 °C).

Total fungi, total bacteria and Gram negative bacteria: Air samples were obtained using a single-stage microbiological air impactor (Airideal™, bioMérieux SA) according to the NIOSH Method 0800 (1998). This is a handheld, battery-operated instrument that consists of a holder for a 90 mm-diameter nutrient agar plate and a sieve plate containing 220 holes. A

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

fan draws air through the sieve plate causing airborne particles to impact on the agar plate, and air is exhausted through the side of the sampler. MacConkey agar, Tryptic soy agar and malt extract agar supplemented with 1% of chloramphenicol were used as culture media for Gram negative bacteria, bacteria and fungi, respectively. Air was drawn through the sampler at 100 l/min, and sequential duplicate air samples (duplicates of 100 and 250 litres) were collected. Incubation conditions for bacteria were 37 ± 1 °C for 48 ± 3 hours. For fungi samples, the incubation period was 25 ± 3 °C for 72 ± 3 hours. The bacteria and fungi colonies were then enumerated and concentrations evaluated in colony-forming units per cubic meter of air (CFU/m³), taking into account a positive hole correlation factor that is used to correct for the possibility that more than one particle containing a cultivable microorganism, passes through the same hole (Andersen, 1958).

Equipment

PM2.5, PM10: TSI DustTrak DRX photometers (model 8533; TSI Inc.)

VOCs: gas chromatography (Agilent Technologies 6890N) coupled to a mass spectrometry detector (Agilent Technologies 5973)

Aldehydes (formaldehyde, acetaldehyde): High Performance Liquid Chromatography (HPLC) (Agilent Technologies, 1220 Infinity LC)

Total fungi, total bacteria and Gram negative bacteria: AirIdeal™, bioMérieux SA.

Sampling characteristics

Safe and childproof sampling sites were ensured and complied with the rules as prescribed by ISO 16000-1 (2004). Samples were collected at a height of about 1-1.5 m above the floor, which is the breathing zone. The selected place was not allowed to be closer than 1 m to a wall, a door or an active heating system. Furthermore, the indoor sampling site was selected as far away as possible from the blackboard (when applicable).


F)

Data on pollutants concentration-air indoor environment (homes) for Portugal.

Measurement methodology (Analytical/method principal)

PM2.5, PM10: Laser photometer based on light scattering technology. The measuring range of the equipment is 0.001 to 150 mg/m³ (1 to 150x10³ µg/m³) with accuracy of $\pm 0.1\%$ of reading of 0.001 mg/m³ (1 µg/m³). Powered by an internal battery, the DustTrak DRX operated with a flow rate of 3.0 l/min using a built-in diaphragm pump.

VOCs: were collected using a stainless-steel sampling tubes containing Tenax® TA (60/80 mesh). Concentrations of target VOC that included benzene, toluene, m/p-xylene, o-xylene, d-limonene, α -pinene, trichloroethylene, tetrachloroethylene, naphthalene, styrene were investigated. After sampling, Tenax tubes were thermally desorbed (Dani STD 33.50) and

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

quantified using a non-polar column by gas chromatography (Agilent Technologies 6890N) coupled to a mass spectrometry detector (Agilent Technologies 5973), according to ISO 16000, part 6 (2011). Total VOC concentration was quantified using the toluene response factor, and concentrations were calculated as the sum of volatile organic compounds eluting between hexane and hexadecane (included), expressed as toluene.

Aldehydes (formaldehyde and acetaldehyde): were sampled by passive devices Radiello® (RAD 165, Sigma Aldrich), consisting of a cartridge filled with 2,4-dinitrophenylhydrazine (2,4-DNPH). By reaction with 2,4-DNPH, aldehydes give the corresponding 2,4-dinitrophenyllhydrazones, which was desorbed in 2 ml of acetonitrile (Sigma-Aldrich) and manually stirred for 30 minutes. Then, the solution was passed through a polyvinyl difluoride 0.45 mm-syringe filter (13 mm Syringe Filter, Specanalítica) and determined using isocratic reverse phase High Performance Liquid Chromatography (HPLC) (Agilent Technologies, 1220 Infinity LC) with a UV detector operated at 360 nm according to ISO 16000-4 (2011). Aldehydes were identified and quantified by comparison of their retention times and peak areas with those of standard solutions. Each cartridge was sealed after sampling and then brought back to laboratory where it was stored in the refrigerator (<4 °C).


Total fungi, total bacteria and Gram negative bacteria: Air samples were obtained using a single-stage microbiological air impactor (AirIdeal™, bioMérieux SA) according to the NIOSH Method 0800 (1998). This is a handheld, battery-operated instrument that consists of a holder for a 90 mm-diameter nutrient agar plate and a sieve plate containing 220 holes. A fan draws air through the sieve plate causing airborne particles to impact on the agar plate, and air is exhausted through the side of the sampler. MacConkey agar, Tryptic soy agar and malt extract agar supplemented with 1% of chloramphenicol were used as culture media for Gram negative bacteria, bacteria and fungi, respectively. Air was drawn through the sampler at 100 l/min, and sequential duplicate air samples (duplicates of 100 and 250 litres) were collected. Incubation conditions for bacteria were 37±1 °C for 48±3 hours. For fungi samples, the incubation period was 25±3 °C for 72±3 hours. The bacteria and fungi colonies were then enumerated and concentrations evaluated in colony-forming units per cubic meter of air (CFU/m³), taking into account a positive hole correlation factor that is used to correct for the possibility that more than one particle containing a cultivable microorganism, passes through the same hole (Andersen, 1958).

Equipment

PM2.5, PM10: TSI DustTrak DRX photometers (model 8533; TSI Inc.)

VOCs: gas chromatography (Agilent Technologies 6890N) coupled to a mass spectrometry detector (Agilent Technologies 5973)

Aldehydes (formaldehyde, acetaldehyde): High Performance Liquid Chromatography (HPLC) (Agilent Technologies, 1220 Infinity LC)

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

Total fungi, total bacteria and Gram negative bacteria: AirIdeal™, bioMérieux SA.

Sampling characteristics

Sampling sites were ensured and complied with the rules as prescribed by ISO 16000-1 (2004). Indoor samples were collected from the rear of the bedroom, 1-1.5 m above the floor (breathing zone) and not closer than 1 m from a wall, a door or an active heating system under normal conditions regarding heating and airing of the home as well as the use of rooms.

G)

Data on CO₂ concentration, temperature and relative humidity for air outdoor environment (school grounds) for Portugal.

Measurement methodology (Analytical/method principal)

Temperature: thermistor

Relative humidity: thin-film capacitive

CO₂: infrared non-dispersive sensor

Equipment

Temperature, Relative humidity and CO₂: IAQ-CALC monitor (model 7545, TSI, Inc.)

Sampling characteristics

Outdoor samples were taken at places which provided electricity and a tamper-free environment. Outdoor samples were collected at heights of 1-2 m above the ground. All the samplers and equipments were mounted in a shelter protecting from direct sunlight and precipitation.


H)

Data on pollutants concentration-air outdoor environment (school grounds) for Portugal.

Measurement methodology (Analytical/method principal)

CO: electrochemical sensor for in a range from 0 to 500 ppm (0 to 572.8 mg/m³) with accuracy of $\pm 3\%$ of reading or ± 3 ppm (3.4 mg/m³). Data were downloaded into LogDat2™ downloading software and then exported for data management.

PM_{2.5}, PM₁₀: Laser photometer based on light scattering technology. The measuring range of the equipment is 0.001 to 150 mg/m³ (1 to 150x10³ µg/m³) with accuracy of $\pm 0.1\%$ of reading of 0.001 mg/m³ (1 µg/m³). Powered by an internal battery, the DustTrak DRX operated with a flow rate of 3.0 l/min using a built-in diaphragm pump.

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

VOCs: were collected using a stainless-steel sampling tubes containing Tenax® TA (60/80 mesh). Concentrations of target VOC that included benzene, toluene, m/p-xylene, o-xylene, d-limonene, α -pinene, trichloroethylene, tetrachloroethylene, naphthalene, styrene were investigated. After sampling, Tenax tubes were thermally desorbed (Dani STD 33.50) and quantified using a non-polar column by gas chromatography (Agilent Technologies 6890N) coupled to a mass spectrometry detector (Agilent Technologies 5973), according to ISO 16000, part 6 (2011). Total VOC concentration was quantified using the toluene response factor, and concentrations were calculated as the sum of volatile organic compounds eluting between hexane and hexadecane (included), expressed as toluene.

Aldehydes (formaldehyde and acetaldehyde): were sampled by passive devices Radiello® (RAD 165, Sigma Aldrich), consisting of a cartridge filled with 2,4-dinitrophenylhydrazine (2,4-DNPH). By reaction with 2,4-DNPH, aldehydes give the corresponding 2,4-dinitrophenylhydrazones, which was desorbed in 2 ml of acetonitrile (Sigma-Aldrich) and manually stirred for 30 minutes. Then, the solution was passed through a polyvinyl difluoride 0.45 mm-syringe filter (13 mm Syringe Filter, Specanalítica) and determined using isocratic reverse phase High Performance Liquid Chromatography (HPLC) (Agilent Technologies, 1220 Infinity LC) with a UV detector operated at 360 nm according to ISO 16000-4 (2011). Aldehydes were identified and quantified by comparison of their retention times and peak areas with those of standard solutions. Each cartridge was sealed after sampling and then brought back to laboratory where it was stored in the refrigerator (<4 °C).


Total fungi and total bacteria: Air samples were obtained using a single-stage microbiological air impactor (AirIdeal™, bioMérieux SA) according to the NIOSH Method 0800 (1998). This is a handheld, battery-operated instrument that consists of a holder for a 90 mm-diameter nutrient agar plate and a sieve plate containing 220 holes. A fan draws air through the sieve plate causing airborne particles to impact on the agar plate, and air is exhausted through the side of the sampler. Tryptic soy agar and malt extract agar supplemented with 1% of chloramphenicol were used as culture media for bacteria and fungi, respectively. Air was drawn through the sampler at 100 l/min, and sequential duplicate air samples (duplicates of 100 and 250 litres) were collected. Incubation conditions for bacteria were 37±1 °C for 48±3 hours. For fungi samples, the incubation period was 25±3 °C for 72±3 hours. The bacteria and fungi colonies were then enumerated and concentrations evaluated in colony-forming units per cubic meter of air (CFU/m³), taking into account a positive hole correlation factor that is used to correct for the possibility that more than one particle containing a cultivable microorganism, passes through the same hole (Andersen, 1958).

Equipment

CO: IAQ-CALC monitor (model 7545, TSI, Inc.).

PM2.5, PM10: TSI DustTrak DRX photometers (model 8533; TSI Inc.)

VOCs: gas chromatography (Agilent Technologies 6890N) coupled to a mass spectrometry detector (Agilent Technologies 5973)

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

Aldehydes (formaldehyde, acetaldehyde): High Performance Liquid Chromatography (HPLC) (Agilent Technologies, 1220 Infinity LC)

Total fungi, total bacteria and Gram negative bacteria: Airldeal™, bioMérieux SA.

Sampling characteristics

Outdoor samples were taken at places which provided electricity and a tamper-free environment. Outdoor samples were collected at heights of 1-2 m above the ground. All the samplers and equipments were mounted in a shelter protecting from direct sunlight and precipitation.

4.7. Pollutant concentration data – soil

The FOREGS Geochemical Baseline Mapping Programme was initiated in 1998 to provide high quality environmental geochemical baseline data in Europe. Geological surveys and related institutions from 26 countries have taken part in the mapping programme.

The FOREGS databases and material archives comprise: archived sample materials (topsoil, subsoil, floodplain and stream sediments, and humus) stored at the Geological Survey of the Slovak Republic. It contains the raw analytical data in Excel files. Each analytical method has a separate file for every solid sample material except stream water. All stream water results are included in a single file. All records include the GTN sample identifier, coordinates in decimal degrees (East coordinates are positive and West coordinates negative, so -14.36 mean 14.36 degrees West), in some cases there is the preprocessing code, country code of the field data, analytical results and another country code for analytical data.


Measurement methodology/Equipment

Determination of anions (Stream Waters)

The samples were analysed for major and trace anions, *i.e.* fluoride, chloride, nitrite, bromide, nitrate, orthophosphate and sulphate, by ion chromatography (IC). The determinations were made on a Dionex 4000i Ion Chromatograph using an AS14 analytical column with both conductivity and UV/visible absorbance detection. The method is accredited by the United Kingdom Accreditation Service (UKAS) according to the requirements of BS EN ISO 17025.

Determination of dissolved organic carbon (Stream Waters)

The samples were analysed using a Shimadzu TOC 5000 analyser with associated ASI 5000

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

auto-sampler. The method is accredited by the United Kingdom Accreditation Service according to the requirements of BS EN ISO 17025.

Determination of major cations, metals and other inorganic elements (Stream Waters)

The samples were analysed by both inductively coupled plasma quadrupole mass spectrometry (ICP-QMS) and inductively coupled plasma atomic emission spectrometry (ICP-AES), using Perkin Elmer Sciex ELAN 5000A and Spectro Flame M instruments respectively, in accordance with the German norms DIN 38406-29 (ICP-MS) and DIN 38406-22 (ICP-AES).

International certified reference materials (NIST 1640, NIST 1643d and SLRS-4) were included in every batch of 20 samples.

Determination of mercury (Stream Waters)

The samples were analysed for mercury by a cold vapour atomic absorption (CVAAS) technique using a CETAC M-6000A Hg Analyser. The method is accredited according to NS-EN ISO/IEC 17025. The method is based on reduction of Hg ions in solution to atomic Hg vapour using stannous chloride (SnCl_2) as a reducing agent. The Hg/ SnCl_2 emulsion is introduced into the top of a gas-liquid separator and is detected by CVAAS.

Determination of TOC (Stream and Floodplain Sediments)


Determination of the total carbon content of sediments was carried out by a pyrolysis technique with non-dispersive infrared detection using a Dohrmann-Rosemount DC-190 (USA) carbon analyser, with 183 boat sampling module and a Fuji model 3300 direct non-dispersive infrared gas detector. The method has been validated against EPA methods 415.1 and 9060, ISO 10694.

Determination of mercury (Stream and Floodplain Sediments)

Determination of total mercury was carried out using a cold vapour atomic absorption technique, involving preconcentration on a gold amalgam before detection with an Advanced Mercury Analyser (AMA-254, ALTEC) instrument.

Total multielement determination by XRF (Stream and Floodplain Sediments)

A range of elements were determined by wavelength dispersive x-ray fluorescence spectrometry (WD-XRFS) and energy dispersive polarised x-ray fluorescence spectrometry

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

(ED(P)XRFs). The instruments used were Philips PW1480 and PW2400 WD-XRFs, with W and Rh anode x-ray tubes respectively, and a Spectro X-LAB 2000 ED-XRF with a Pd anode x-ray tube.

Total multielement determination by ICP-MS (Stream and Floodplain Sediments)

In addition to XRF analysis, a range of major and trace elements were also determined by quadrupole ICP-MS using a VG Elemental (UK) model PQ3 instrument equipped with a water cooled sample introduction system, a PlasmaScreen™ torch, a high efficiency interface device AutoRange Plus™, a simultaneous detector system and PlasmaLab ICP-MS Software Suite.

Acid leachable multielement determination by ICP-AES (Stream and Floodplain Sediments)

In addition to the total concentrations of the elements, the acid leachable portion of selected elements were analysed after hot *aqua regia* leach by ICP-AES, using a J-Y 70 Plus Geoplasma ICP-AES instrument

Determination of TOC (Soils)


The total organic carbon content of soil samples was analysed using a LECO SC-DR144 instrument coupled to a Mettler AT 400 analytical balance. The method is based on total combustion of the samples after removal of the carbonate phase by hydrochloric acid. After combustion in a 99.99 % oxygen flow at 1350°C, the evolved CO₂ is passed through a water and halogen trap and detected with an infrared detector. Analytical grade CaCO₃ was used to construct a four point linear calibration.

Determination of mercury (Soils)

Mercury in soils was determined by the method described above for sediments.

Total multielement determination by XRF (Soils)

A range of elements was determined by wavelength dispersive x-ray fluorescence spectrometry (WD-XRFs) using Philips PW1480 and PW2400 WD-XRFs, with Cr and Rh anode x-ray tubes respectively. In addition to the routine laboratory quality control protocol, LUFA-B (a German residual soil reference sample) was analysed regularly throughout the batches of samples presented for analysis.

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

Total multielement determination by ICP-MS (Soils)

A range of rare earth elements (REE) and other trace metals were determined on soil samples using a Perkin Elmer Sciex Elan 5000 inductively coupled plasma mass spectrometer for the main method.

A simple cold dissolution method was applied for selected elements, viz: As, Cd, Cs, Cu, Mo, Ni, Pb and Sn. This method used a Perkin Elmer Sciex Elan 6000 inductively coupled plasma-mass spectrometer.

A reagent blank and two CRMs (Slovakian Residual Soil 982 and River Clay 981) were included as monitoring samples in every batch of 40 samples. In addition, CRMs were prepared within each batch of 80 samples for the main digestion method (syenite SY-2) and within each batch of 40 samples for the cold dissolution method (San Joaquin Soil NIST 2701).

Acid leachable multielement determination by ICP-AES (Soils)

The acid leachable metal content of soils was determined by the same method described above for sediments.

Determination of mercury (Humus)

Mercury in humus was determined by the method described above for sediments.


Total multielement determination by ICP-MS (Humus)

Humic samples were analysed for total metal content by inductively coupled mass spectrometry. Two instruments were used: a Fisons Plasmaquad PQ2 ICPMS during the early stages of the project and latterly an Argilent 7500i ICPMS instrument. The calibration regime for both instruments was very similar.

Some determinands were measured by more than one technique, e.g. a range of trace elements in soils and sediments were determined both by XRF and ICP-MS after mixed acid digestion

Analytical Requirements

Regional, large-scale geochemical projects demand special requirements for chemical analysis:

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

- (i) the analytical methods used must be sufficiently sensitive to allow detection of a wide range of determinands in all of the sample media at background levels;
- (ii) the analytical precision must be good, preferably significantly better than natural geochemical variation;
- (iii) the analytical accuracy must also be good, preferably even better than that achieved in most national geochemical surveys; and
- (iv) all data and other records pertaining to the analysis and testing must be fully documented and traceable.

Sampling characteristics:

The FOREGS sampling grid was based on GTN (Global Terrestrial Network) grid cells developed for the purpose of Global Geochemical Baseline mapping. The sampling procedures including the planning phase is described in detail in the field manual (Salminen, Tarvainen *et al.* 1998). A list of the GTN cells, which should be sampled was produced beforehand by the Geological Survey of Finland (GTK).

From each country at least one GTN cell was randomly selected for duplicate sampling. Countries with nine or more GTN cells collected duplicate samples from 2 or more cells. Duplicate samples of each material were taken from one geologically representative small catchment (1, 2, 3, 4 or 5) of the selected GTN cell, and its corresponding floodplain. The procedure of collecting the duplicate samples was identical with that of the normal samples.

Sampling was carried out in each country by national teams. Normally one team sampled all sites in each country during one field season between 1997 and 2001. However, in some cases the work was divided in two field seasons. Therefore, it can be stated that the FOREGS geochemical baselines mapping programme represents the end twentieth century state of the surficial environment in Europe.

Sample identifiers were formulated as indicated in the following example. The sample code of the stream sediment sample is:


N43E09S4, where

N43E09 = the 43rd GTN cell north of equator and the 9th cell east of 0 meridian;

S = Sample medium symbol for stream sediment, which is replaced by "W" for stream water, "T" or "C" for top or bottom residual soil, "H" for humus, "F" or "L" for top or bottom floodplain sediment respectively;

4 = Drainage basin number.

The letter D was used as a suffix for the duplicate sample identifier of each sampling medium: N43E09S4D.

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

The identifier for filtered blank (or zero) water samples was a “0” suffix: N43E09W40.

4.8. Pollutant concentration data – dust

There is no Data stored in the HEALS EDMS database so there are no quality assessment procedures yet.

4.9. Pollutant concentration data – water


Waterbase contains timely, reliable and policy-relevant data collected from EEA member countries through the WISE-SoE data collection (formerly known as Eionet-Water and Eurowaternet) process. WISE-SoE data collection selects validated monitoring data from national databases and adds information on the physical characteristics of the water bodies being monitored and on the pressures potentially affecting water quality. The added value of Waterbase is that data collected through the WISE-SoE data collection process are from statistically stratified monitoring stations and groundwater bodies and are comparable at the European level. EEA covers data for the following parameters:

Benzo(a)pyrene, Alachlor, Atrazine, chlorfenvinphos, chlorpyrifos, Diuron, gamma-HCH (Lindane), Isoproturon, Simazine, Trifluralin, Aldrin, DDT o,p', DDT p,p', DDE p, p', DDD p,p', Dieldrin, Endrin, Bentazone, alpha- HCH, beta-HCH, alpha-Endosulfan, Desethylatrazine, Desisotropilatrazine, Propazine, Prometryn, Terbutylazine, 2,4-D,MCPA, Mecoprop, Isodrin, Linuron, Mercury, Benzene, 1,2-Dichloroethene, cadmium, Lead, 1,1,2-trichloroethene, 1,1,2,2-tetrachloroethene

Data are in µg/m³ and the temporal coverage is 1960-2012

Measurement Methodology

The data requested through the WISE-SoE process derived from existing national and/or regional monitoring networks within each EEA Member Country. Member Countries are asked to provide data on transitional, coastal and marine waters according to criteria described in the (Eurowaternet) technical guidelines (EEA Technical Report No. 97), and as detailed in this data request. These data provide a general overview, based on truly comparable data, of water quality at a European level.

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

Data processing rules described in the “EU Directive on technical specifications for chemical analysis and monitoring of water status 2009/90/EC”

Sampling characteristic:

Each country follow Directive 2009/90/EC

4.10. Pollutant concentration data – drinking water

Data on Drinking water concentration parameters from the “Public company of drinking water of Thessaloniki”


Measurement Methodology

The verification is done by the application of certified analytical methods (Standard Methods of Analysis for Water and Wastewater and EPA) and performed as to the organoleptic, physico-chemical, toxic parameters related to undesirable substances and microbiological parameters in accordance with applicable law.

Equipment

The Quality Control Laboratory of Water Supply of Thessaloniki has quality assurance system ISO 9001: 2008. The ISO 9001: 2008 is an internationally recognized standard for quality assurance and its purpose is to organize and create a framework for the operation in the laboratory to provide services based on the needs of the consumer. Another process that is underway is the "Laboratory Accreditation according to ELOT EN ISO 17025». The accreditation of a laboratory is the identification of a competent official body recognized that the laboratory is able to perform accurate analyzes and therefore constitutes evidence and the credibility of its results. The main objective of accreditation as described and the requirements of international standard ISO / IEC 17025, is to reduce the likelihood of errors, both in the conduct of the analyzes and in the stages before and after doing so.

Sampling characteristics

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

Sampling was done on a daily basis, 7 days a week, 365 days a year and control includes, besides the distribution of water pumping stations, water intake sources, schools and hospitals.

4.11. Food concentration data

The EFSA Comprehensive European Food Consumption Database (Comprehensive Database) has been built from existing national information on food consumption at a detailed level. Competent organisations in the European Union's Member States provided EFSA with data from those most recent national dietary survey in their country, at the level of consumption by the individual consumer. This included food consumption data concerning infants (2 surveys from 2 Member States), toddlers (8 surveys from 8 Member States), children (16 surveys from 14 Member States), adolescents (14 surveys from 12 Member States), adults (21 surveys from 20 Member States), elderly (9 surveys from 9 Member States) and very elderly (8 surveys from 8 Member States) for a total of 32 different dietary surveys carried out in 22 different Member States.

EFSA's Founding regulation (Art. 33) states that EFSA shall search for, collect, analyze and summarize particularly data on

- food consumption,
- incidence and prevalence of biological risks, and
- occurrence of contaminants and chemical residues.


EFSA covers food consumption data for the following categories: Beverages alcoholic and non-alcoholic, bread and rolls, bakery products, cereals and products, flour, pasta, eggs, cheese, milk, milk products, total added lipids, fish and sea food, nuts, fruits, meat and products, potatoes, vegetables, pulses, sugar, sugar products, juices

Summary statistics from the database enable quick screening for chronic and acute exposure to substances that may be found in the food chain. In the database, dietary surveys and food consumption data for each country are divided by category. These include: age, from infants to adults aged 75 years or older; food group (nearly 160) and type of consumption, covering both regular and high consumption thus allowing calculations to be tailored to each category of consumer.

The statistics on food consumption are reported in grams per day, as well as grams per day per kg of body weight.

EFSA covers data for the time interval: 1997-2008

Measurement Methodology

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

Directive 2003/99/EC assigns EFSA the task of examining data submitted annually by Member States on zoonoses, zoonotic agents, food-borne outbreaks and antimicrobial resistance

4.12. Population data (age/gender/etc)

Data came from two epidemiological studies performed in subjects living in Pisa (Tuscany): Pisa2 study (1991-93) and IMCA2 study (2008-2011). The studies were carried out by the Pulmonary Environmental Epidemiology Unit, CNR Institute of Clinical Physiology.

A questionnaire on socio-demographic characteristics, respiratory symptoms/diseases and risk factors was used. A selection of the databases containing demographic information is reported in the template. Additional information, collected through the questionnaire, about risk factors (e.g. smoking habits, occupational exposure, environmental exposure) and health status are available.

Measurement Methodology and equipment use

Standardized questionnaire on socio-demographic characteristics, respiratory symptoms/diseases and risk factors

4.13. Noise data

Noise exposure data (<http://noise.eionet.europa.eu/>) available in excel files, categorized in excel sheets (2012 & 2007)

Results are expressed in Nr of people exposed to different noise bands (Lden) (Lnight)

% of people exposed to different noise bands (Lden) (Lnight)

Lden : Day-evening-night level

Lnight : Night-time noise indicator


Data are available per various cities in European countries, Also summarized data are presented per country.

The data were reported by member states and member countries corresponding to 2005 and 2007 deliveries, and to 2010 (2008) and 2012 deliveries, updated up to 28th of August 2013.

Major roads, railways, airports and agglomerations follow the below thresholds:

Agglomerations > 100,000 inhabitants.

Major civil airport > 50,000 movements/year

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

All major roads > 3 million vehicles/year

All major railways > 30,000 trains/year

Measurement Methodology

In Annex I of the European Directive 2002/49/EC, which is related to the assessment and management of environmental noise, Lden is defined by a formula based on Aweighted long term average sounds levels (ISO 1996-2 :1987).

Sampling characteristics


Each country member should follow Directive 2002/49/EC for sampling guideline

4.14. Building characteristic data

A walkthrough inspection was completed for each classroom by a trained researcher to gather information related to characteristics of the selected classrooms involved in study. In the walkthrough inspection a standardized checklist was used that has been previously field-tested in SINPHONIE project and was completed with the assistance of principals and teachers. A specific checklist form was filled in for each classroom identifying all relevant information such as the area, finishing materials and their conditions concerning floor, walls, and ceiling; windows; past occurrences and visible problems, heating and ventilation system, scholar activity products (paintings, glues, etc.) maintenance routines and cleaning procedures. The type of classroom furniture, the presence of chalkboards, copiers and plants were also noted, as well as information about environmental modifiers including air fresheners and insecticides.

5. DATA QUALITY CONTROL

In the next paragraphs is describing the quality control procedure used for the data set existing until now in the EDMS database of the HEALS project where the quality control procedures are known.

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

5.1. Land use/Land cover

A quality verification program is run immediately after each metadata transfer to the AVPDB to identify data and processing problems. Information that is output from subprograms includes:

- a. duplicate scene identifiers
- b. acquisitions per day (total, network, local, and 1 KM)
- c. scenes entered per day (total, network, local, and 1 KM)
- d. total scenes and total 1 KM scenes received from each station
- e. acquisition year error for DOMSAT and local acquire
- f. duplicate or overlapping coverage by same station
- g. duplicate data base record transfer from processing system
- h. scenes without reference aid id after elapsed time
- i. scenes without media storage location
- j. scenes with excessively short or long pass duration
- k. geographic coordinates beyond range of receiving station
- l. scenes with reference aid id with incorrect microfiche set code
- m. scenes with invalid receiving method
- n. receiving station mismatch between AVPDB, tape receipt, and processing logs.

Each subprogram will output specific information necessary to determine the cause of the anomaly. Independent programs have been developed to assist in resolving data and processing problems.


2. A coverage verification program is run periodically to report less than expected coverage by each receiving station. For dates with less than expected passes recorded the number and identification of scenes or passes archived is listed. The output is in the form of correspondence to the receiving stations and includes total scenes and tapes received and the range of acquisition dates.

3. Daily and Weekly LAC plots are plotted and distributed on a weekly basis.

4. Receiving station coverage circle plots are plotted periodically.

5. Plots by selected receiving station(s) are generated when necessary.

6. The number of scenes expected on each tape vs. the number of scenes acquired, ingested, and processed through reference aid generation is tracked by the DATE MANAGEMENT tape receipt data base. Anomalies are output to the common error file.

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

7. When a station's data is being ingested for the first time or their has been historical problems with a station's data, the microfiche images can be examined directly from the SGS4 disk using an X terminal. Systematic image data defects and linework misalignment are indicators of data problems.

8. A world plot of LAC and HRPT coverage by day will be useful for spotting gaps in coverage.

9. Line drops, linework placement, and image defects will be examined on the microfiche. The fiche will normally be available within 1 day of ingest.

10. DATE MANAGEMENT will delete 1 Kilometer scenes only if there is no visible land area or the scene is a duplicated by the same receiving station.

For all received tapes, multiple ingest attempts are done for each media type (8mm, 4mm, 3480, 9-track) that has initial ingest problems. The team's ingest procedures verify coordinate accuracy, satellite direction, correlate with Day/Night flags, correctly determine browse bands (band 2 for day, band 4 for night), acquisition date and time accuracy in the scene id, and that the facsimile header data matches the data base records.

5.2. Meteorological Data

A) The ECA dataset


Method validity

Series of the best possible quality are provided for ECA&D by the participating institutions. In addition, common quality control procedures are applied to all series using various algorithms. These quality control procedures lead to flags ("OK", "suspect" or "missing") assigned to individual data. Although data validation has been careful, it can never be excluded that some errors remain undetected. The risk for such errors is greatest in the recent data that stem from synoptical messages, because these data did not undergo the validation process in the participating institutions.

Imported data checks

For evaluation of the homogeneity of the time series in ECA&D a two step testing procedure was followed.

First, four common homogeneity tests were applied to evaluate the daily series in fixed time periods using the testing variables: (1) the annual mean of the diurnal temperature range DTR (= maximum temperature - minimum temperature), (2) the annual mean of the absolute day-to-day differences of the diurnal temperature range vDTR, (3) the annual wet day count RR1 (threshold 1 mm), (4) the annual number of snow days SD1 (threshold 1 cm), (5) the

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

annual mean of daily sea level pressure PP, (6) the annual sunshine duration SS, (7) the annual mean of daily average relative humidity RH, and (8) the annual mean of daily average cloud coverCC.

Second, the test results were condensed for each series into three classes: useful-doubtful-suspect. The four common homogeneity tests are: Standard Normal Homogeneity test, BuisHand Range test, PETtitt test and von NEUmann ratio test.

B) NASA data set

Method validity

Nasa provides estimates of the levels of uncertainty for temperature, surface pressure, relative humidity, and wind speed through comparisons with ground measurement data. SSE estimates were compared with ground site data on a global basis. Meteorological parameters were compared with data from the National Climate Data Center (NCDC). Wind speeds have been carried over from SSE Release 4 because newer data sets do not provide enough information about vegetation/surface types. The RETScreen Weather Database was used to test uncertainties in the SSE Release 4 wind speeds.

5.3. Comfort data indoors (temperature-relative humidity)

For the three data sets that are stored in the EDMS database concerning the SINPHONIE project, the QC procedure are the following:

Calibration of the instrument used

The instrument was calibrated annually at the factory.

5.4. Air emission data


A) The BUMA database

Method validity

The BUMA database is updated with emission data from studies carried out by certificated institutions as well as by peer reviewed literature.

5.5. Air emission data – consumer products

A) The BUMAC database

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

The BUMAC database is updated with emission data from studies carried out by certificated institutions following the international standard procedures or by published methodologies as well as by peer reviewed literature focusing more on data for health relevant compounds.

5.6. Pollutant concentration data – air

A) Data extract from AirBase (EEA)

Method validity

For PM10 and PM2.5 beta attenuation, method is validated for its equivalence to the reference method EN 12341 and EN 14907:2005 respectively

For SO₂, NO_x, O₃ the methods used are in line with the reference method of EU: EN14212, EN14211, and EN 14626 respectively

For VOC the method used complies with the reference method EN 14662:2005 part 1,2 and 3.

For CO the method used is the same with the reference method of EU EN 14626:2005

For toxic metals (Ni/Cd/Pb/As) the method used is the same as the reference method used (EN14902:2005)

Equipment used validity

The PM_x equipment is certified for EN 12341 and EN 14907 equivalent designations

SO₂, NO_x O₃ and CO are operate according to the reference measurement methods EN 14212, 14211, 14626, 14626 respectively

Toxic metals are measured according to the reference method EN 14902 using AAS.

Calibration

Calibration frequency; every month for all equipment used

Calibration method: For PM_x, standard mass


For O₃, O₃ generator stopped

For NO_x, SO₂, CO, compressed gases

For BS, black filter (reflectometry)

Imported data checks

The Airbase data sets use the DEM software for imported data check, which is the abbreviation of **D**ata **E**xchange **M**odule. The software can be used to exchange air quality

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

(AQ) data and meta information on operational AQ networks, stations and measurement configurations. The Data Exchange Module enables to:


- Add, modify and delete meta information on AQ networks, stations and measurement configurations; this can be done either on screen or by file entry;
- Select station-component combinations for which AQ data is to be exchanged, including subsequent linking of these combinations to AQ data files and import of these files with extensive error checking;
- **Check the imported data on outliers, on strange calculated annual values and on missing measurement data.**
- Visualise imported time series in the form of histograms and line graphs for error checking;
- Generate reports giving an overview of (imported) meta information contained in the application;
- Generate MS Excel files with DEM meta information, imported raw data and calculated statistics, an overview with imported raw data stored in the DEM, DEM AQ files with raw data, Google Earth KML files with station positions, XML and NASA Ames export files
- Transmit database through Internet (upload to EEA Central Data Repository (CDR)) for incorporation in the EEA database AIRBASE and the EMEP/CCC database EBAS.

DEM is pre-loaded with all the meta information on AQ networks, stations and measurement configurations currently contained in AIRBASE.

More specific there are several checks on the measurement data (Mol 2012):

- Check on outliers. For each component a default lower and upper value is defined, but it is also possible to change these values (per country) into a country dependent lower and upper value.
- “Check AQ Data With Statistics”. Overviews with statistics calculated from the imported measurement data are generated. The imported data can be checked on: zero or negative annual means or annual means which are 3 times lower or higher than the year before. The last annual means are preloaded in the DEM.
- “Check imported AQ-Data”. This module checks the data on missing imported data. The module gives information on data which have been reported the year before but not now or data which have been reported now but not the year before (there is a gap the year before).

In addition the raw data of every country in the Airbase data sets have the “Flag values” next to the pollutant concentration value which indicate the quality of the preceding measurement value. A quality flag value > 0 indicates valid measurement data. A quality flag ≤ 0 indicates invalid or missing data.

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

B) The OFFICAIR data sets

Method validity

LC, ISO 16000-4

thermodesorption, GC, ISO 16017-2

spectrophotometry, EN 14412 and EN 13528-3

Calibration

PM2.5: The flow rate was calibrated before sampling and verified after sampling using the whole sampling train (with the cyclone/impactor and representative filter in place).

Imported data checks

Data of blanks and duplicates are included in the data tables, where appropriate;


- For aldehydes, NO₂ and ozone (VOCs are not concerned), in addition to the temperature correction, the field blank must be subtracted in the reported concentration;
- Data not quantifiable and data not detectable will be reported as “LoQ” and “LoD”;
- The real values for ‘limit of quantification LoQ’ and ‘limit of detection LoD’ of each compound will be reported in the table (where applicable);
- For missing data insert a text ,“n.a.” or “n.v”
- Only units, will be reported:

Aldehydes, VOCs, PAHs, PMs, NO₂, O₃ in µg/m³;

Temperature, wall temperature in °C;

Relative humidity in %;

- The input data reporting files are prepared using Microsoft EXCEL 2010. Only this version will be used for data reporting into the database of the JRC;
- All data inserted into the EXCEL data reporting tables as ‘numbers’ with maximum 2 decimal places using a point ‘.’ (e.g. 10. 35 and NOT 10,35);
- The reporting table templates will be completed in the case of central analysis by the central laboratory and in case the partners execute the registration respectively the analysis by the partners;
- Data originated from central analysis will be first reported back to the partners by using the agreed input data templates. The partners collect all data of one measurement campaign, check these and send the complete dataset
- Once reported to the ftp server, the data will be automatically uploaded into the database;

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

C) Data from Toscana (Italy)

ARPAT is engaged in an ongoing management and validation of the collected data through the use of statistical tools useful for long data series analyses.

Data quality is also ensured guaranteeing the maintenance of quality standards required by the relevant legislation, in particular to the approval of the sampling and measurement instruments (Article 17 of Legislative Decree 155/2010); for this purpose, as required by L.R. 9/2010 art. 5, paragraph 4, the Region relies on the Regional Centre of Reference for the control and data quality assurance (Centro regionale di Riferimento per il controllo e l'assicurazione di qualità dei dati -CRRQA) of ARPAT.

D) Data for pollen from Toscana region (Italy)

Equipment validity

The samplers require a periodic control relative to the intake air flow.

Recently, the accuracy and precision of operators are checked internally and through participation in experimental ring tests organized by the network POLLnet in collaboration with the University of Perugia.

E) Data for indoor air pollutant concentration for Porto (Portugal)-schools

Method validity


VOCs: VOCs were analysed according to ISO 16000, part 6 (2011). To control contamination during transport and sampling, a field blank was employed in every school. All samples were taken in duplicate to test the reproducibility of measurements.

Aldehydes (formaldehyde, acetaldehyde): were analysed according to ISO 16000-4 (2011). As an internal quality, control duplicate samplings were collected in 1 school per set of 3 schools. Field blanks were collected and analyzed to assess possible contamination through the sample collection and analysis process.

Total fungi, total bacteria and Gram negative bacteria: Air samples were obtained according to the NIOSH Method 0800 (1998). For each sampling day, agar media blanks per culture were taken into the field but were not opened.

Calibration

PM2.5, PM10: The instrument was calibrated annually at the factory.

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

Total fungi, total bacteria and Gram negative bacteria: The impactor was calibrated before use.

E) Data for indoor air pollutant concentration for Porto (Portugal)-residences

Method validity

VOCs: VOCs were analysed according to ISO 16000, part 6 (2011). To control contamination during transport and sampling, for every set (week) of homes a field blank was employed. All samples were taken in duplicate to test the reproducibility of measurements.

Aldehydes (formaldehyde, acetaldehyde): were analysed according to ISO 16000-4 (2011). As an internal quality control duplicate samplings were collected in 1 bedroom per set of 3 homes. A blank sample was deployed in each set of 3-5 homes.

Total fungi, total bacteria and Gram negative bacteria: Air samples were obtained according to the NIOSH Method 0800 (1998). For each sampling day, agar media blanks per culture were taken into the field but were not opened.

Calibration

PM2.5, PM10: The instrument was calibrated annually at the factory.

Total fungi, total bacteria and Gram negative bacteria: The impactor was calibrated before use.

G) Data for outdoor air pollutant concentration for Porto (Portugal)-school grounds


Method validity

VOCs: VOCs were analysed according to ISO 16000, part 6 (2011). To control contamination during transport and sampling, a field blank was employed in every school. All samples were taken in duplicate to test the reproducibility of measurements.

Aldehydes (formaldehyde, acetaldehyde): were analysed according to ISO 16000-4 (2011). As an internal quality control duplicate samplings were collected in 1 school per set of 3 schools. Field blanks were collected and analyzed to assess possible contamination through the sample collection and analysis process.

Total fungi and total bacteria: Air samples were obtained according to the NIOSH Method 0800 (1998). For each sampling day, agar media blanks per culture were taken into the field but were not opened.

Calibration

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

CO: The instrument was calibrated annually at the factory.

PM2.5, PM10: The instrument was calibrated annually at the factory.

Total fungi and total bacteria: The impactor was calibrated before use.

H) Data for outdoor air pollutant concentration for Porto (Portugal)-residences

Method validity

VOCs: VOCs were analysed according to ISO 16000, part 6 (2011). To control contamination during transport and sampling, for every set (week) of homes a field blank was employed. All samples were taken in duplicate to test the reproducibility of measurements.

Aldehydes (formaldehyde, acetaldehyde): were analysed according to ISO 16000-4 (2011). As an internal quality control duplicate samplings were collected in 1 outdoor sampling site per set of 3 homes. A blank sample was deployed in each set of 3-5 homes.

Total fungi and total bacteria: Air samples were obtained according to the NIOSH Method 0800 (1998). For each sampling day, agar media blanks per culture were taken into the field but were not opened.

Calibration

CO: The instrument was calibrated annually at the factory.

PM2.5, PM10: The instrument was calibrated annually at the factory.


Total fungi and total bacteria: The impactor was calibrated before use.

I) Data for outdoor air pollutant concentration –Satellite Data

Method validity

There are different modules for quality control and monitoring, production control, archiving and order handling. The level 2 products generated at DLR are validated on a regular basis by the O3M-SAF partner institutes AUTH, BIRA, and FMI. The GOME-2/MetOp level 2 NRT products are disseminated primary with EUMETCast, WMO/GTS and internet. The off-line products are disseminated via Internet and media. The ordering of products can be done via dedicated user services.

The GOME-2/MetOp level 1 products are delivered in NRT (approx. 1:45 hours after sensing) in so called PDU files (product dissemination unit) containing 3 minutes of measurements via EUMETCast. The ground-segment at DLR needs less than 15

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

minutes for acquiring the input data, retrieving the trace gas total columns and disseminating the resulting level 2 products, that means the GOME-2/MetOp total column products are available to the users in less than 2 hours after sensing.

The GOME-2/MetOp operational processing is performed in a high availability virtualized environment on a blade center farm.

Calibration

Earth observation modes

Earth observation (or "earthshine") modes are those modes where the Earth is in the field of view of GOME-2. They are usually employed on the dayside of the Earth (sunlit part of the orbit). The scan mirror can be at a fixed position (static modes), or scanning around a certain position (scanning modes). All internal light sources are switched off and the solar port of the calibration unit is closed.

Nadir scanning

This is the mode in which GOME-2 will be operated most of the time. The scan mirror performs a nadir swath as described above. The swath width is commandable, its default value is 1920 km. Scanning can be performed either with constant ground speed, resulting in equally sized ground pixels (this is the default), or with constant angular speed ("GOME-1 mode"), resulting in larger ground pixels for the extreme swath positions as compared to the swath centre.

North polar scanning

The scan mirror performs a swath around the viewing angle $+46.696^\circ$ (default value) in order to cover the North Pole which would not be observable with the normal nadir scanning mode. This mode will typically be used during northern hemisphere spring.

South polar scanning

The scan mirror performs a swath around the viewing angle -46.172° (default value) in order to cover the South Pole which would not be observable with the normal nadir scanning mode. This mode will typically be used during southern hemisphere spring.


Other scanning

The scan mirror performs a swath around another off-nadir position.

Nadir static

The scan mirror is pointing towards nadir. This mode will typically be used during the monthly calibration. It is valuable for validation and long-loop sensor performance monitoring purposes.

Other static

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

The scan mirror is pointing towards an off-nadir position.

Calibration modes

In-orbit instrument calibration and characterisation data are acquired in the various calibration modes. They are usually employed during eclipse with the exception of the solar calibration which is performed at sunrise. Both internal (WLS, SLS, LED) and external (sun, moon) light sources can be employed. The various sources are selected by the scan mirror position.

Dark

The scan mirror points towards the GOME-2 telescope. All internal light sources are switched off and the solar port is closed. Dark signals are typically measured every orbit during eclipse.

Sun (over diffuser)

The scan mirror points towards the diffuser. All internal light sources are switched off and the solar port is open. Solar spectra are typically acquired once per day at the terminator in the northern hemisphere. The Sun Mean Reference spectrum will be derived from this mode.

White light source (direct)

The scan mirror points towards the WLS output mirror. The WLS is switched on and the solar port is closed. The WLS can be operated at four different currents (360, 380, 400, 420 mA). Etalon (and optionally Pixel-to-Pixel Gain (PPG) calibration) data will be derived from this mode.

Spectral light source (direct)


The scan mirror points towards the SLS output mirror. The SLS is switched on and the solar port is closed. Wavelength calibration coefficients will be derived from this mode.

Spectral light source (over diffuser)

The scan mirror points towards the diffuser. The SLS is switched on and the solar port is closed. Light from the SLS reaches the scan mirror via the diffuser. This mode is employed for in-orbit monitoring of the sun diffuser reflectivity.

LED

The scan mirror points towards the GOME-2 telescope. The LEDs are switched on and the solar port is closed. PPG calibration data will be derived from this mode.

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

The geophysical validation of the official GOME products is performed by independent partner organizations: AUTH, BIRA, and FMI. The validation is accomplished using ground-based measurements available through the World Ozone and Ultraviolet radiation

Data Centre, ground-based measurements from the NDACC network, other satellite data, as well as data assimilation techniques. The validation results are publicly available on the Internet: <http://o3msaf.fmi.fi/valreps.html>.

5.7. Pollutant concentration data – soil


1. Data on FOREGS-EuroGeoSurveys

For the data produced by the project, additional quality measures were used to provide extra assurance that the overall quality of data was to the high standards required. For all solid analyses, two reference materials (ISE 921 and ISE 982) were analysed at regular intervals (between 1 and 2%, depending on the method) to monitor long-term stability and to enable comparison of data from different methods and different laboratories. About 8% of the solid samples were also used for inter-laboratory comparison, e.g. about 50 soil samples were analysed by WD-XRF at BGS for comparison with the original BGR WD-XRF and GTK ICP-MS data.

Uncertainty in chemical analysis, although often significantly less than that associated with sampling and sample preparation, is nevertheless essential to understand and control. All of the participating laboratories were selected on the basis that they had established quality systems in place, they regularly participated in international proficiency testing schemes for the methods for which they were responsible, and their estimated analytical uncertainty and sensitivity for all determinands was likely to be acceptable.

In order to assess the precision of the chemical analysis relative to the variation due to sample preparation and sampling, duplicate samples and repeat measurements were taken. Approximately 5% of all sites were sampled in duplicate. The duplicate samples were prepared independently of each other and analysed randomly along with all other samples, with each duplicate sample also analysed in duplicate to allow estimation of sampling uncertainty by the ANOVA statistical interpretation method.

Missing values are marked with value -1.0 in most cases. Values below the detection limit (DL) are changed to half the detection limit value (DL/2). Example: In topsoil XRF data, the detection limit for arsenic (As) is 2 mg kg⁻¹; therefore, all values which were marked as <2 mg kg⁻¹ in the original file are transformed to the value 1.00. In the stream water data set there was more than one detection limit for anions. All anion concentrations below a detection limit were converted to half the value of the most common detection limit for an

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

anion. In some cases there was more than one detection limit also for cations. For cations: all values below any DL have been converted to half the detection limit value (DL/2).

5.8. Pollutant concentration data – dust

There is no Data stored in the HEALS EDMS database so there are no quality assessment procedures yet.

5.9. Pollutant concentration data – water

A) Data on Waterbase collected from EEA member countries

Method validity

Data on quality of water in groundwater are collected annually through the WISE-SoE data collection process. Dataset contains data selected from reporting of member and collaborating countries on chemical quality of groundwater, characteristics of groundwater bodies and sampling sites. Reported data have been assessed and processed by the ETC-Water and the EEA. Disaggregated records were annually aggregated by groundwater body, substance and year, and statistic value calculated. Results of quality assessment have been incorporated into the individual data tables.

Imported data checks

The Quality control tests have been performed on the Waterbase – Groundwater database provided in 05 May 2014 by ETC-ICM. Following main types of the tests have been performed on the data tables. Mandatory value and Measurement value tests, Primary key/Duplicate tests, Logical rules violation test, Outlier detection tests, Stations tests and Data definition compliance test (Waterbase – Groundwater Version 14).

Measurement values check:

Mean values in the Quality table are subject of this test. Detected issues are then stored as a code in a special QA field (QA_MVissues) as follows:

101 – the Mean value is missing


102 – the Mean value is negative and negative values are not allowed or possible

103 – the Mean value is equal 0 and 0 values are not allowed or possible

Records flagged with any of these flags either can't be used (101) or it is recommended that they are excluded from further use or analysis (102, 103).

Logical rules violation tests:

The following logical rules were tested in “NUTRIENTS” and “HAZSUBS” data tables:

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

201 – Mean \geq Minimum

202 – Mean \leq Maximum

203 – Median \geq Minimum

204 – Median \leq Maximum

205 – Minimum \leq Maximum

206 – If Minimum > 0 Then StandardDeviation $<$ Maximum

207 – If Minimum $<$ Maximum Then StandardDeviation > 0

210 – All measurement values ≥ 0 (exceptions: Alkalinity, Temperature)

211 – If NumberOfSamples = 1 Then (Mean = Minimum = Maximum = Median)

212 – If NumberOfSamples = 1 Then StandardDeviation = 0

213 – If NumberOfSamples = 0 Then all measurement values are Null

217 – NumberOfSamplesBelowLOQ \leq NumberOfSamples

Outlier detection tests:

A special QA field (QA_outlier) has been added to the tables and records, where the any of the situations mentioned above has been detected, have been flagged in this field as follows:

401 – Standard potential outlier - value is either higher/lower than limit value or is suspiciously high/low comparing to the rest of the time series or value change between two consecutive values is suspiciously abrupt or is marked as an outlier by a content expert

402 – Measurements are probably taken from a highly polluted locations but information was not confirmed

403 – The whole country delivery is considered as problematic because it contains too many quality issues


409 – The value can't be confirmed by data provider (the original source data are unavailable)

410 – The value previously detected as an outlier has been corrected (is not an outlier anymore)

491 – Outlier has been confirmed by country as correct value

492 – Outlier has been confirmed by (ETC) content expert as correct value

493 – Measurement has been confirmed by country to be taken from a highly polluted area

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

It is recommended that the records where QA outlier field contains codes 401-409 and eventually also code 493, should not be used in a further analysis or only after a careful consideration.

5.10. Pollutant concentration data – drinking water

The organizations handling the measurements on environmental parameters of drinking water in various European countries and regions, are all of them certified, using the Standard Operational Procedure for analysis of drinking water.

5.11. Food concentration data

A) Food consumption and contamination of the EFSA database

Method validity


Data collection was performed following the Directive 2003/99/EC

Imported data checks

All scientific units complete a self-review form to guarantee a set of key steps is consistently followed in developing scientific opinions and other EFSA outputs. This includes ensuring:

- A shared understanding with the requester (usually the European Commission) of what is expected by the Terms of Reference and a realistic time frame for delivery.
- Approved Declarations of Interests are provided by all involved experts before work begins and harmonised procedures applied where potential conflicts are identified.
- All sources of information and scientific data are clearly described and referenced.
- Key studies are properly identified and reasons given for their selection, while implications of potentially contradictory data are discussed and any diverging views with other expert bodies are explained.
- Significant data gaps and underlying assumptions, limitations and uncertainties in data are identified and discussed.
- Summaries are adequately informative and reflect the conclusions/recommendations of the opinion.

A selection of the self-review process is double-checked by an internal review team which makes comments and recommendations for further improvement. In addition, a working group of independent experts has been set up in order to carry out an external review. Their recommendations are being discussed with the appropriate parties and fed into the annual report of EFSA's Quality Manager which is published on the EFSA

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

website and is used by EFSA to strengthen and improve the quality of its working processes.

5.12. Population data (age/gender/etc)

A) Data from EUROSTAT website - weblink

Imported data checks

Comparability - geographical


Eurostat aims to collect data from EU Member States in relation to populations as of 1 January each year. The recommended definition is the 'usual resident population' and represents the number of inhabitants of a given area on 1 January of the year in question (or, in some cases, on 31 December of the previous year). In accordance with the United Nations international recommendations, the definition of the 'usual residence' is based on a 12 month reference period, in other words, those included should have lived in their place of usual residence for a continuous period of at least 12 months before the reference date, or arrived in their place of usual residence during the 12 months before the reference date with the intention of staying there for at least one year. However, countries may report to Eurostat population figures based on data from their most recent census, adjusted by the components of population change that have been produced since the last census, or alternatively population figures that are based on the registered/legal population.

Comparability - over time

For most of the countries, the population data for the year 2011 and after take into account the results of the latest population census (held in 2011). The time series of populations between the previous census and 2011 will be revised by end 2013 by some of the countries, taking into account Eurostat's recommendation. The comparison of populations between a pre-census and a post-census year may result in differences partially explained by the changes in the population structure and partially explained by the lack of revision of the pre-census population data.

The following countries have transmitted to Eurostat post-2011 Census population revisions by autumn of 2013, which are reflected in the tables 'Demographic balance and crude rates (demo_gind)', 'Population on 1 January by age and sex (demo_pjan)', 'Population on 1 January by five years age groups and sex (demo_pjangroup)' and 'Population on 1 January by broad age group and sex (demo_pjanbroad)':

BG 2007-2011; CZ 2001-2011; EE 2000-2011; IE 2007-2011; EL 2011; ES 2002-2011; CY 2003-2011; LV 2001-2011; LT 2001-2011; MT 2006-2011; AT 2008-2011; PT 1992-2011;

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

RO 2002-2011; SK 2002-2011; UK 2002-2011 (not including post-2011 Census data for Scotland).

The breaks in population series due to the succession of a pre and post census reference year or due to any methodological change in population counts reported by the countries are documented in Eurostat's database with the flag b (break in series).

In addition basic controls are carried out on the absolute figures reported by the NSIs to check if the total of a variable is consistent with the breakdown by sex, by year and by other characteristics (e.g., live births are the sum of births inside marriage plus births outside marriage; total population is equal to the sum of marital statuses, etc.). Cross validations are carried out to check consistency between the different breakdowns of the same variable (e.g., the total of births must be equal to the sum of births by rank, by mother's educational attainment, etc.).

Based on the detailed collected data, Eurostat is computing a series of demographic indicators. The regular calculated demographic indicators are further submitted to several validations concerning the most updated annual value but also the available time series. These validations include check on the plausibility of the most updated annual value, checks on two consecutive annual values, checks by using standard deviation and checks by using weighted average as for example for breakdowns of the life expectancy by educational attainment.

Full details on data collection/validation are available at:
http://epp.eurostat.ec.europa.eu/cache/ITY_SDDS/EN/demo_pop_esms.htm#unit_measure1405581193055


5.13. Noise data

A) Noise data outdoor-EIONET noise data

Method validity

Values L_{den} and L_{night} can be determined by computation or by measurement according to the European Directive 2002/49/EC ANNEX II, where recommended methods are presented for industrial, aircraft, road and railway noise sources.

Measurements of each state member are in accordance with the principles governing long-term average measurements stated in ISO 1996-2: 1987 and ISO 1996-1: 1982

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

Imported data checks

European Directive 2002/49/EC describes in ANNEX VI the procedure for organizing the data for agglomerations, major roads, major railways and major airports.

The European Environment Agency (EEA) and the European Topic Centre on Spatial Information and Analysis (ETC-SIA) has published a handbook for delivery of data in accordance with Directive 2002/49/EC.

The full guide is uploaded in the following link:
<http://www.eea.europa.eu/publications/noise-handbook> .

Briefly this guide covers:

The data submission process

Checklist for the data reporters to be sure that the data reported accomplish the minimum requirements

Statistical information

Spatial information

Supplementary information

Metadata

Naming conventions

Quality check process

B) Noise data indoor/outdoor- Noise Navigator data


Method validity

The data are compiled from references in the literature and from measurements. For each source the reference is listed, and as available additional notes are provided. When the primary reference cites sources for its data, those too are listed.

5.14. Building characteristic data

A) Data concerning the SINPHONIE project

SINPHONIE questionnaires on school building characteristics, and on the classroom characteristics, are developed and completed for every participating school and classroom. Questions on the presence of potential typical indoor sources of air contaminants are

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

included in these questionnaires as well. Questionnaires and building check list are introduced and presented during the SINHONIE training sessions to the SINPHONIE field workers, in order to avoid any ambiguity or misinterpretation.


6. CONCLUSIONS/RECOMMENDATIONS

A large amount of environmental stressors data, have been identified and assessed in the light of HEALS project targets and especially for the WP8 objectives. In addition the Quality Assurance and the Quality Control procedures adopted for the collected data have also reviewed and presented in this document. Data of environmental stressors variables have found for all the media/categories. As shown in table 1 ANNEX II there is found measurements data for almost all the priority environmental stressors proposed by HEALS project team. This was aided by national and European databases, consumer's protection organizations and literature review as well as the existing data from other EU projects (EPHECT, OFFICAIR, SINPHONIE etc.). The literature review was conducted continuously during the work package duration, from scientific publications on environmental stressors that potentially affecting the human health (directly and indirectly). **The main focus of the review was on recent campaigns/data -studies (from 1995 and on for journals) in order to have the ability to evaluate more reliable data.**

- **The literature and databases review limited to peer-reviewed journals and concerning the year of 1985 and on in order to have newly and more reliable data.**
- **Only – emission- data based on standardized measurement methods described in ISO/ASTM/CEN guidelines were taken into consideration or when an alternative method is used this method should have at least been published in a peer reviewed journal.**
- **Data on various environmental stressors were assessed only from authorized organizations (European, international or global) and national public agents.**
- **Data were illustrated the description of sampling and analytical techniques will be the first priority to include them in the database. In that context the calibration procedure for the instruments/monitors used are preferably to be referred.**

Some recommendation based on the experienced gained through the QA/QC procedures assessed for the data collected in the EDMS database:

- 1) In order to obtain for the EXHES study, presenting comparable data, collected in a uniform way, with a measurement uncertainty reduced to a minimum, a training course to all the involvement partners participating in the corresponding study (for sampling location, instrumentation, calibration of the equipment, sampling duration, sample handling etc)


 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

- 2) The analytical procedure of the measurement samples must be based on ISO/ASTM/CEN guidelines. Thus an analytical EXHES method protocol might be created (from sampling to analytical methodologies of analyze the samples to results processing).
- 3) The analysis will be performed by laboratories that are accredited according to ISO 17025:2005 or that are working according to ISO 17025 quality criteria. In any other case the analysis might done by a central laboratory with the above criteria.
- 4) Another quality assurance is an internal quality control of the participating laboratories in order to evaluate the accuracy of the analysis procedure, applied by each laboratory. It involves the analysis of (certified) reference material (RM).
- 5) Additional quality assurance/quality control is an internal quality control of the participating laboratory which involves duplicate analysis.
- 6) A transparent and uniform sample numbering system for all the participants in the EXHES study is essential for further assessment of the samples.

7. CONTRIBUTION


This document has been written under the contribution of the WP8 partners and other data provider partners:

UOWM (I. Sakellaris, N. Pitaraki, M.Siarga)
NCSR (Th. Maggos, M. Stamatellopoulou, L. Leodiadis)
USTUTT (R. Friedrich , Ch. Schieberle, J. Roos)
CNR (G. Viegi)
TNO (A. Pronk)
IOM (J. Cherrie, M. Loh, H. Cowie)
IDMEC (E.Oliveira Fernades, J. Madureira)
CERETOX (J. de Lapuente)
VTT (S. Nousiainen), OIKON (Z. Spiric) NIOM (K. Polanska).
UNIVBRIS (C. Sabel, L. Smith)

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

8. REFERENCES

- [1] DIRECTIVE 2007/2/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE)
- [2] <http://www.eea.europa.eu/themes/water/dc>
- [3] <http://www.eea.europa.eu/data-and-maps/data/waterbase-groundwater-10>
- [4] EEA Technical Report No. 97
- [5] EU 2009/90/EC of 31 July 2009 laying down, pursuant to Directive 2000/60/EC of the European Parliament and of the Council, technical specifications for chemical analysis and monitoring of water status
- [6] Italian aerobiology monitoring network Pollnet (<http://www.pollnet.it>)
- [7] Directive 2003/99/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 17 November 2003 on the monitoring of zoonoses and zoonotic agents, amending Council Decision 90/424/EEC and repealing Council Directive 92/117/EEC
- [8] <https://www.ifc.cnr.it/index.php/it/le-aree-di-competenza/l-epidemiologia/epidemiologia-polmonare>
- [9] DIRECTIVE 2002/49/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 25 June 2002 relating to the assessment and management of environmental noise.
- [10] Salminen, R., Tarvainen, T., Demetriades, A., Duris, M., Fordyce, F. M., Gregorauskiene, V., Kahelin, H., Kivisilla, J., Klaver, G., Klein, H., Larson, J. O., Lis, J., Locutura, J., Marsina, K., Mjartanova, H., Mouvet, C., O'Connor, P., Odor, L., Ottonello, G., Paukola, T., Plant, J.A., Reimann, C., Schermann, O., Siewers, U., Steenfelt, A., Van der Sluys, J., Vivo, B. de, & Williams, L., 1998. FOREGS geochemical mapping field manual. Geologian tutkimuskeskus, Espoo. Opas 47, 36 p. + 1 app
- [11] FOREGS-EuroGeoSurveys Geochemical Baseline Database:
<http://www.gtk.fi/publ/foregsatlas/>
- [12] ISO 16000-1 (2004)
- [13] Geochemical Atlas of Europe; copyright © 2005 the Association of the Geological Surveys of The European Union (EuroGeoSurveys)/ the Geological Survey of Finland.
- [14] DIRECTIVE 2000/60/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 23 October 2000 establishing a framework for Community action in the field of water policy.
- [15] COUNCIL DIRECTIVE 98/83/EC of 3 November 1998 on the quality of water intended for human consumption.

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

[16] Mol, W.J.A. (2012): Quality checks on air quality data in AirBase and the Eol data in 2012. ETC/ACM Working Paper December 2012. Available at http://acm.eionet.europa.eu/databases/airbase/ETCACM_WP_2012_Quality_checks_Eol2012_AirBase.pdf

[17] Waterbase – Groundwater Version 14, Quality control documentation. Available at <http://www.eea.europa.eu/data-and-maps/data/waterbase-groundwater-10#tab-additional-information>.

9. ANNEXES

ANNEX I

Manual for WP8 Data template form

A. Data Class


Select one of the given options from the dropdown arrow

Field type: Obligatory, single selection.

B. Environmental Stressors

Select one or more environmental stressor from the given options

Field type: Obligatory, multiple selection. In case there is no stressor relevant to the data class selected please choose N/A

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

C. Location

1. Country

Select one of the given options from the dropdown arrow.

Field type: Obligatory, single selection.

The data should be given per country or per Europe.

If the data refer to more than one country and less than Europe then the user should submit one template per country.

2. Region(s)

Write the region(s) that the data refer to (text). Use the full names as given in the attached file.

Field type: Optional.

It is desirable the data to be given per region - if practical

If the data cover more than one or a few regions of the country state the regions.


If the data cover all regions of the country write "ALL"

Otherwise (i.e. too many or unspecified or irrelevant) leave this entry blank.

3. Municipality(ties)/ City(ties)

Write the Municipality(ties) that the data refer to (text).

Field type: Optional.

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

Only if the data cover more than one or a few Municipality(ties)/City(ies) of the country state the Municipality(ties)/City(ies). Otherwise (i.e. too many or unspecified or irrelevant) leave this entry blank.

4. Longitude Range

Provide values if applicable

Field type: Optional.

5. Latitude Range

Provide values if applicable

Field type: Optional.

6. Altitude Range


Provide values if applicable

Field type: Optional.

7. Above Ground Level Range

Provide values if applicable

Field type: Optional.

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

8. *Spatial Resolution*

Provide the space dimensions that the data represent

Provide single value if applicable

Field type: Obligatory (use N/A if it is irrelevant)

D. Medium/Medium Characteristics

Select one or more types from the given options (air, water, soil, dust, food, other)

a. If the medium is air


Ambient environment type

Select one or more types from the given options (N/A, urban, suburban, rural, industrial, other, non specified)

Field type: Optional.

Microenvironment Class

Select one or more of the given options (Outdoors, Indoors, In vehicle, other, non specified)

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

Select one or more microenvironment(s) from the given list from the selected above option

Field type: Optional


b. If the medium is not air

Specify the type (text)

E. Data Information

Provide information per stressor – If applicable

1. Time period
Insert starting and ending date (dd/mm/yyyy) –if applicable
2. Time resolution range
Insert min and max values with units – if applicable
3. Information/table(s) with quantified parameters and units
Provide list/ table (obligatory)
4. Additional or/and explanatory information on the stressor(s) (if applicable)
Provide text if necessary
5. Additional or/and explanatory information concerning time
Provide text if necessary
6. Additional or/and explanatory information concerning space
Provide text if necessary

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

7. Support information (including links if necessary) to read the data file(s)

Provide text if needed

8. Other useful information

Provide any other information (text) concerning the data that it is not covered elsewhere

Field type: Obligatory

F. DataSource / Weblink

9. References / Website

Write down the literature reference(s) or the website that describes the data.

Field type: Obligatory

10. Data weblinks


Provide all the necessary data links so that the full picture of the data is given– If applicable

Field type: Optional (obligatory only if the option below “upload the data file” is not selected)

11. Upload the data file

Upload the single data file only for data that cannot be found in a database- If applicable

Field type: Optional (obligatory only if the option above “data weblinks” is not selected)

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

G. Measurement methodology

Provide information per stressor – If applicable

1. Analytical principle / method

Provide the measurement methodology(ies)

Field type: Obligatory if the information exists

2. Equipment

Provide information on measurement equipment

Field type: Obligatory if the information exists

3. Sampling characteristics


Provide relevant information that includes sample configuration, sampling time etc.

Field type: Obligatory if the information exists

4. Additional Comments

Provide any other missing information useful to the user

Field type: Optional

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

H. QA/QC Procedure

Provide information per stressor – If applicable

1. Method validity

Provide any available information on compatibility with the reference methods, standards and directives

Field type: Obligatory if the information exists

2. Equipment validity

Provide any available information on equipment accreditation – If applicable

Field type: Obligatory if the information exists

3. Calibration

Provide any available information on calibration schedule and methodology– If applicable

Field type: Obligatory if the information exists


4. Imported data checks

Provide any available information on data numbers validity (e.g. uncertainties, quality grading, missing data, outliers, etc)

Field type: Obligatory if the information exists

5. Additional comments

Provide any other missing information useful to the user on the QA/QC procedure


 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

Field type: Optional

I. SUBMITTING PARTNER

Provide institution, name and e-mail address


Field type: Obligatory

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100


ANNEX II

Table 1. Environmental stressors and environmental media (categories) that data identified and populate the EDMS database.

STRESSOR	Air emission data	Air emission data – consumer products	Pollutant concentration data – air	Pollutant concentration data – soil	Pollutant concentration data – water	Pollutant concentration data – drinking water	Food concentration data
PM10	x	x	x				
PM2.5	x	x	x				
VOCs	x	x	x				
Phthalates		x					
Toxic metals	x		x	x	x	x	x
Organic	x	x	x	x	x	x	x
Organoalogenes		x		x	x	x	x
Dust mites							
Pesticides				x	x		x
Endocrine					x		x
POPs	x	x		x	x		x
NO ₂	x	x	x				
O ₃		x	x				
Mold							
Pollen			x				
Noise							
other (SO ₂)	x		x				
other (NH ₃)	x	x					
other (CO)	x		x				

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

other (CO2)	x						
other (CH4)	x						
other (N2O)	x						
other	x	x					
other (H2S)			x				
other (PAHs)			x				
other (multiple,					x	x	





 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

ANNEX III


EDMS template for part A. Data Class

[←](#)
[→](#)
[↺](#)

heals.uowm.gr/index.php

HEALS Environmental Data Management System



HEALS

[Home](#)
Get Started

[WP8 Data](#)
Template Form

[Logout](#)
partner

WP8 Template Form

Examples (templates already stored): [Single Data Set](#) [Multiple Data Sets](#)

Choose from the options below.

A. Data Class

SELECT AN OPTION


SELECT AN OPTION

Land Use / Land Cover
 Meteorological Data
 Comfort data indoors (temperature - humidity)
 Air Emission Data
 Air Emission Data – consumer products
 Pollutant Concentration data – air
 Pollutant Concentration data – soil
 Pollutant Concentration data – dust
 Pollutant Concentration data – water
 Pollutant Concentration data – drinking water
 Food Concentration Data
 Population data (age/gender/etc)
 Noise data
 Buildings characteristic data

Country

Region(s)

Municipality(ties)

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

ANNEX IV

EDMS template for parts **B. Environmental Stressors** and **C. Location (Country)**

←

→

↺

heals.uowm.gr/index.php

🔍

🔑

☆

🔴

B. Environmental Stressors

☐ N/A
 ☐ PM₁₀
☐ PM_{2.5}
☐ VOCs
 ☐ Phthalates

☐ Toxic Metals
 ☐ Organic compounds
 ☐ Organoalogenes
 ☐ Dust mites
 ☐ Mold

☐ Pollen
 ☐ Pesticides
 ☐ Endocrine disruptors
 ☐ POPs
 ☐ NO₂

☐ O₃
☐ Noise
 ☐ Other (please specify)

C. Location

Country

SELECT AN OPTION ▼

SELECT AN OPTION

Europe

Albania

Andorra

Austria

Belgium

Bosnia and Herzegovina

Bulgaria

Croatia

Cyprus

Denmark

Estonia

Finland

France

Germany

Greece

Hungary

Iceland

Ireland

Italy

Region(s)


Municipality(ties)

To (Longitude)

To (Latitude)

To (Altitude)

Above Ground Level Range (m)

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

ANNEX V

EDMS template for part C. Location (Regions)

← → ↺

heals.uowm.gr/index.php

🔍

🔑

☆

🔴

☰

B. Environmental Stressors

☐ N/A
 ☐ PM₁₀
☐ PM_{2.5}
☐ VOCs
 ☐ Phthalates

☐ Toxic Metals
 ☐ Organic compounds
 ☐ Organoalogenes
 ☐ Dust mites
 ☐ Mold

☐ Pollen
 ☐ Pesticides
 ☐ Endocrine disruptors
 ☐ POPs
 ☐ NO₂

☐ O₃
☐ Noise
 ☐ Other (please specify)

C. Location

Country

Austria ▼

Region(s)

SELECT ONE OR MORE REGIONS

AT11 - Burgenland (AT)

AT12 - Niederösterreich

AT13 - Wien

AT21 - Kärnten

AT22 - Steiermark

AT31 - Oberösterreich

AT32 - Salzburg

Municipality(ties)

Longitude Range

From (Longitude)

Latitude Range


From (Latitude)

Altitude Range (m)

From (Altitude)



To (Altitude)

Above Ground Level Range (m)

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

ANNEX VI

EDMS template for parts D. Medium/Medium Characteristics and E. Data Information


heals.uowm.gr/index.php


SELECT AN OPTION ▼

D. Medium / Medium Characteristics

Medium

SELECT AN OPTION ▼

SELECT AN OPTION

N/A

Air

Drinking Water

Water (other)

Soil

Dust

Food

Other

E. Data Information

Time Range (Dates)

From (Date: dd/mm/yyyy)


To (Date: dd/mm/yyyy)

Time Resolution

☐ Hour
☐ Day
☐ Month
☐ Year
☐ Other(please specify)









Information/table(s) with quantified parameters and units

Give details

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

ANNEX VII

EDMS template for parts **F.** Data Source/Weblinks, upload data file and **G.** Measurement methodology




heals.uowm.gr/index.php






F. Data Source / Weblinks

1. References / Website

2. Data web link

3. Upload the data file:

Select a file to upload

G. Measurement methodology

1. Analytical principle / method

Give details

2. Equipment


Give details

3. Sampling characteristics

Give details

4. Additional comments

Give details

 HEALS FP7-ENV-2013-603946	D8.1 - Report on quality assessment and quality control of environmental data collected		
	WP8: Environmental data mining	Security:	
	Author(s): E. Tolis, P. Karagiannis, J. Bartzis	Version: 1	4/100

ANNEX VIII

EDMS template for parts **H. QA/QC Procedures** and **I. Submitting Partner**

←

→

↺

heals.uowm.gr/index.php

🔍

🔑

☆

🔴

H. QA/QC Procedure

1. Method validity

Give details

2. Equipment validity

Give details

3. Calibration

Give details

4. Imported data checks

Give details

5. Additional comments

Give details

I. SUBMITTING PARTNER

Submit