Application of Sensor Web Technology for Analysing Correlations between Health and Environmental Data

Simon Jirka[1], Johannes Brauner[2], Arne Bröring[3], Siegbert Kunz[4], Ingo Simonis[5], Kym Watson[6]

Abstract

This paper introduces the European FP7 project EO2HEAVEN which aims at investigating the complex relationships between environment and health. To support such research a Spatial Information Infrastructure (SII) is introduced which allows combining different environmental data sources with health data sets and analysing correlations between these data sets. The paper presents the general challenges of combining environmental and health data, the EO2HEAVEN SII and its underlying improved Sensor Web concepts.

1. Introduction

Often, environmental conditions have an influence on human health. For example, air quality may have an impact on respiratory diseases or weather conditions may increase the spread of certain infectious diseases. Research is necessary to discover and understand these impacts and to provide people with guidance on how to minimize the health impact of environmental factors. For The European project EO2HEAVEN (Earth Observation and Environmental Modelling for the Mitigation of Health Risks, FP7 GA 244100) was started to provide experts from the health and environment domain with the necessary tools and data for their work (i.e. health and environmental data and services).

Several challenges need to be tackled to achieve the goal of better supporting health experts. These comprise on the one hand policy and security concerns which have to be considered when dealing with highly sensitive information such as health data sets. On the other hand, the integration of relevant data sources into analysis and processing tools of experts is still a cumbersome task (i.e. there are different data models, data types, and data formats that need to be integrated). To overcome these limitations, this paper introduces the EO2HEAVEN Spatial Information Infrastructure (SII) as an approach based on a service oriented, interoperable and distributed architecture.

In the next section the case studies underlying the EO2HEAVEN project are introduced. These case studies, covering cardiovascular and respiratory diseases as well as cholera, are a core driving factor of the EO2HEAVEN project. They deliver the requirements of domain researchers as input to the architecture and technology development of EO2HEAVEN. Further requirements result from the specific challenges that arise when integrating environmental and health data. These challenges are discussed in section 3.

[1] 52°North GmbH, Martin-Luther-King-Weg 24, 48155 Münster, Germany, email: jirka@52north.org
[2] Technische Universität Dresden, Helmholtzstr. 10, 01069 Dresden, Germany, email: johannes.brauner@tu-dresden.de
[3] 52°North GmbH, Martin-Luther-King-Weg 24, 48155 Münster, Germany, email: broering@52north.org
[4] Fraunhofer IOSB, Fraunhoferstr. 1, 76131 Karlsruhe, Germany, email: siegbert.kunz@iosb.fraunhofer.de
[5] OGC Europe, United Kingdom, email: isimonis@opengeospatial.org
[6] Fraunhofer IOSB, Fraunhoferstr. 1, 76131 Karlsruhe, Germany, email: kym.watson@iosb.fraunhofer.de

Subsequently, in section 4, the EO2HEAVEN SII is introduced which is the architectural foundation for fulfilling the needs of domain researchers. Central features of the SII are improved concepts for the OGC Sensor Web Enablement architecture as explained in section 5. Finally, the paper closes within an outlook on future work (section 6) and conclusions (section 7).

2. The EO2HEAVEN Case Studies

The research carried out in EO2HEAVEN is organized around three distinct case studies (CS) each dealing with particular problems located in Europe and southern Africa. By aligning the research work to case studies, the EO2HEAVEN methodologies and infrastructure are validated from a user perspective.

Case Study 1 – Environmental Effects on Respiratory and Cardiovascular Diseases in Dresden and Saxony, Germany

Various health and environment studies show regional proof that air pollution endangers human health, especially the respiratory and cardiovascular system (e.g. Strickland et al., 2010, McConnell et al., 2010). This link has so far only been poorly established for Germany's federal state Saxony. Thus, Saxony was chosen to be a case study region for EO2HEAVEN; in addition, it represents a typical central European region and could thus be of general interest. Combining human health data (from a public health insurer and official morbidity and mortality statistics or emergency hospital admissions) with air pollution and meteorological data (both measured by the regional governmental in-situ sensor network) is essential to derive health risks from exposure to air pollutants. Various geostatistical and health statistical models including visualization concepts have been developed to detect and analyse the correlation between the two data groups. The implementation is based on Web services deployed in a Spatial Information Infrastructure (SII) in the context of the regional INSPIRE (Infrastructure for Spatial Information in Europe) implementation.

Case Study 2 – Environmental Challenges to Health in the South Durban Industrial Basin, South Africa

The CS in Durban deals with air pollution and its influence on human health, as well. In contrast to the European study in Saxony, Durban (especially the South Durban basin) is exposed to major industrial air pollution. The percentage of childhood asthma is considerably higher than in any other report found worldwide. Over 50 per cent of school children living and learning in the direct neighbourhood of industrial emission sources suffer from acute or chronic asthma (Robins et al, 2002). Additionally, the reduced data availability (as compared to Saxony) in terms of health statistics and the lack of area-wide in-situ sensor observations pose challenges, although, in general, the same models and analyses as in CS1 are carried out in Durban. The usage of remote sensing imagery to support the health-environment analysis is currently being explored. EO2HEAVEN is developing methodologies and software aiming to support the work of local healthcare authorities and non-governmental organisations to warn citizens of risky environmental conditions. Additionally, an alert system based on mobile technology is being developed presenting an easy to understand user interface.

Case Study 3 – Cholera in Kasese, Uganda

CS3 investigates the dynamics of cholera outbreaks and the responsible pathogen Vibrio cholera (toxigenic) in the Kasese district in Uganda. It relates the outbreaks to potentially influencing environmental conditions and factors. To facilitate the design of re-active and pro-active counter measures and intervention strategies, the findings are of interest to researchers, decision makers in governmental organisations and non-governmental organisations active in cholera prevention and cholera outbreak emergency situations. EO2HEAVEN provides decision makers with spatiotemporal enabled visualizations of cholera outbreaks. Embedding the environmental findings into a social context is especially important to analyse the potential vulnerability of an area where the pathogen is found or known to exist. It is crucial to better understand when environmental conditions become favourable for an outbreak or intensify an outbreak once it has begun. The software tools and components currently being developed in EO2HEAVEN for CS3 integrate datasets from remote sensing and in-situ sensors and include datasets from microbiological field sampling campaigns and laboratory results. Scientists from different field should find, access, extract, visualize, analyse and model the provided data and publish their own findings in an easy manner.

3. Challenges when Integrating Health and Environmental Data

"In recent years a number of emerging and re-emerging diseases have been seen around the world […]. This is largely due to a continued neglect of basic public health practices in general and environmental health services in particular. Neglecting [..] [those services] often resulted in an increase in diseases associated with environmental factors." (South African National Department of Health, 2011). In a number of developing countries, this process was accompanied by a politically motivated change from a preventive to a curative health system, which further ignored research on correlations between environmental factors and epidemiological effects. In other countries, the development of policies for preventive health systems has not accelerated to full speed till date. Although some voices call for a salutogenic model as a theory to guide health promotion (e.g. Antonovsky, 1996), economists usually identify the preventive health care system as the most cost-efficient solution (e.g. Secker-Walker 1997, Kelly 2000, Coco 2005). The base requirement for any preventive system is the identification of the causes and the proper understanding of their correlation with (potential) effects. In turn, this understanding requires the systematic exploitation of environmental and public health data. With its beginnings in the development of social statistics in the 1800's when governments and individuals started to collect social data and information in a systematic way and the first use of 'social' data and information for epidemiological purposes, there are roughly 200 years of research and experience that could be used today (Simonis 2011). The poor understanding of the correlations between environmental factors and diseases is still an important impedimental factor in epidemiology.

Research in the domain of environmental health is often driven by a rather narrow focus. If the hydrocarbon-induced stress in earthworms or the evaluation of surface water quality in aquatic bodies under the influence of uranium mining, research is often not tackling the full line of correlations from base environmental factors to disease outbreaks and distribution patterns among humans. The reasons for those limitations are manifold, but often include the heavily restricted access to health data. It is easier to stop the research at the level of identification of potential adverse effectors rather than to proof and quantify the possible effect on humans in a complex environment of stress factors, environmental parameters, and individual susceptibility.

The privacy of medical information has always been a tenet of responsible medical care, but the electronic distribution and analysis techniques combined with developments of new medical tests and examination techniques have resulted in a new quality of health data and interrelated risk of illegal use. This data contains highly sensitive information and could, if misused, negatively influence the future of individuals, e.g. if the results of HIV or genetic susceptibility tests become public. In many cases, data privacy is now reflected in legal frameworks to prevent unscrupulous and self-interested exploitation of health data. Unfortunately, those legal frameworks make life harder for the research community as well. To ensure that data of an individual person cannot be recovered from an anonymised dataset, high-resolution input data is aggregated multiple times to the level of obfuscation of individuals. This deliberate aggregation is in the opposite direction of general research practice, where the highest quality of results is strived for. In addition, the trace back to individuals often includes additional potential for health care and health research. There is no pure technological solution for this situation. In a world that continuously commercializes research, no isolated technology can help. In contrast, it requires a highly interwoven framework of ethical, technical, and legal rules and regulations to address the issue of data privacy. Those frameworks have to be ruled out on an international level to allow cross-border analysis and comparability.

The next section introduces the EO2HEAVEN Spatial Information Infrastructure which is the technological basis for achieving the integration of environmental and health data.

4. The EO2HEAVEN Spatial Information Infrastructure

The specification of the implementation architecture of the EO2HEAVEN Spatial Information Infrastructure (SII) is the basis for all EO2HEAVEN developments. Particular emphasis is put on the support of environmental and health monitoring as required by the three EO2HEAVEN case studies.

The architectural work in EO2HEAVEN is carried out iteratively in order to proceed step-by-step in the architectural specifications. To ensure a coherent methodological approach across all development cycles, each cycle results in one version of the SII Implementation Architecture. Currently work on the 3rd iteration is ongoing. The EO2HEAVEN architecture incorporates further developments of advanced concepts for Sensor Web Enablement (SWE), distributed Geo-Processing and Spatial Decision Support. This includes event-based interactions across all functional domains and the inclusion of models from the environmental and health domains as virtual sensors. The EO2HEAVEN architecture specification is structured as a six part (Part I to Part VI) document with a coherent and redundancy-free set of architectural specifications including concept developments.

The structure of the SII architecture is also oriented at an interpretation of the five viewpoints of the ISO Reference Model for Open Distributed Processing (RM-ODP). These viewpoints (partly combined) are specified in the six document parts of the SII architectural specification. It continues the series of architecture specifications of previous European FP6 projects which resulted in OGC documents as follows:

- The foundation was set by the ORCHESTRA project (http://www.eu-orchestra.org) and its Reference Model for the ORCHESTRA Architecture (RM-OA), approved as an OGC best-practices document (OGC 07-097)
- The RM-OA was extended by the SANY project (http://sany-ip.eu) in its Sensor Service Architecture (SensorSA), accessible as an OGC discussion paper (OGC 09-132r1).

Referring to the original copyrights of these documents and in agreement with their editors, the EO2HEAVEN project inherits and extends them in form of the present SII implementation architecture specification.

The published EO2HEAVEN SII architecture of the 2nd iteration focuses both on the sensor architecture and on the EO and health data aspects. It investigates the consequences of special health data privacy and security requirements on the architecture. Relevant security aspects include security setting descriptions, mechanisms to establish security and how to propagate security settings through processing chains. The SII provides further extensions and refinements, e.g. taking into account the requirements to share and process huge amounts of datasets provided by earth observation agencies and health institutions in order to investigate and assess correlated risks.

Further major topics of the EO2HEAVEN SII are processing and fusion services, processing of quality information, handling sensor data uncertainty, its encoding and visualization, etc. The uncertainty of sensor measurements (e.g. the relative or absolute accuracy of the manufacturer's device) is encoded in OGC Sensor Model Language (SensorML) documents for each sensor. An interpolation process can take account of the data accuracy by acquiring it automatically from the SensorML descriptions accessible via an OGC Sensor Observation Service. The result of such an interpolation process is a coverage, i.e. a set of estimated property values for the sampling points together with a quantified description of their uncertainty. Figure 1 illustrates as an example the special situation of a two dimensional wind values interpolation in a selected area as a coverage layer (heat map) of the resulting wind speed in combination with an arrow field indicating the wind direction (white arrows). It is processed using the measured mean wind values (red arrows) at three measurement stations (red dots) via an OGC based fusion SPS/SOS Web service. This includes also the adequate visualization of the uncertainty in diverse illustrations such as a heat map using in parallel both a colour scale (in areas with results of a low uncertainty) encodings and grey scale encodings (in areas with uncertain results above a given threshold due to the long distance from the measurement stations).

Other challenges are advanced services e.g. to adjust the different data resolutions, to handle intermittent data (e.g. due to sensor unavailability or communication interruptions) or to identify and handle data outliers (e.g. due to sensor failures). However, this architectural specification still reflects a work in progress so that further enhancements are expected by the end of the project.

All versions of the EO2HEAVEN SII architecture will be publically available, whereas at least the last one shall be disseminated by means of the OGC channels. The next section focuses on a specific part of the SII architecture, the OGC Sensor Web Enablement (SWE) framework and the advanced SWE concepts that were developed within EO2HEAVEN.

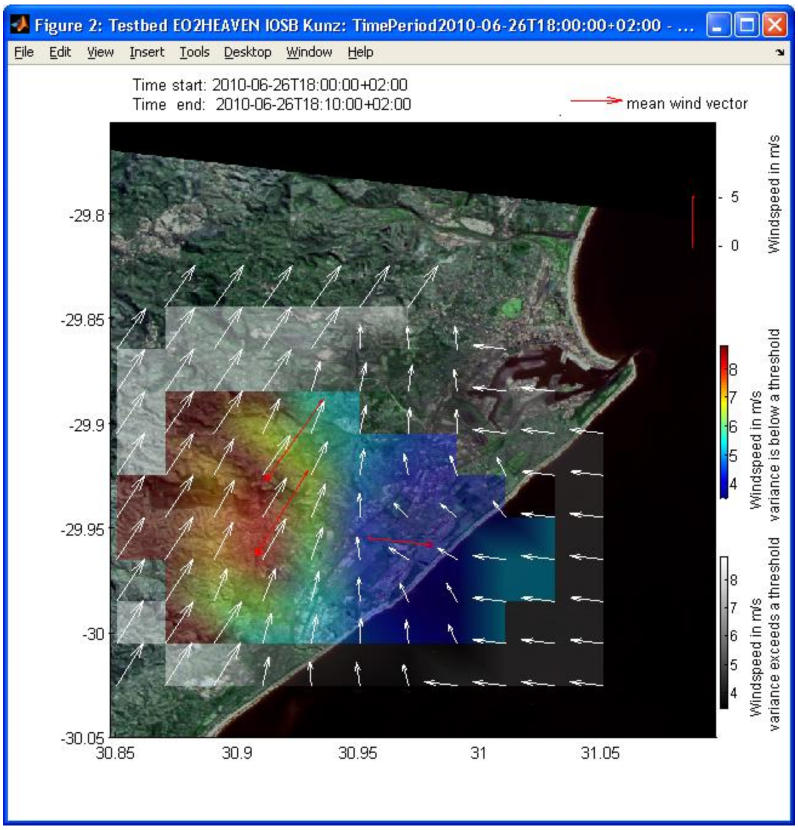

Figure 1

Visualisation of the Uncertainty of Interpolated Wind Values by Combining Colour and Grey scale Encodings Depending on the Local Variance Values of the Interpolation (Durban in South Africa)

5. Advanced Sensor Web Concepts

When investigating relationships between environmental influences and health, sensor data play a central role. This comprises on the one hand historical observation data to analyse changes that occurred during the course of time and on the other hand real-time data for building alerting mechanisms in case of critical situations. The Sensor Web Enablement (SWE) technology of the Open Geospatial Consortium (OGC) is used within EO2HEAVEN to integrate heterogeneous sensor data sets with health data. To make SWE fit for the requirements of the EO2HEAVEN Case Studies, several challenges had to be addressed. This section introduces these challenges and describes how they were tackled.

An important feature of the OGC SWE standards is their domain independent character so that a very broad range of usage scenarios and applications can be covered. This leads at the same time to a need for flexibility in the different SWE specifications which might hinder interoperability (e.g. the same functionality can sometimes be achieved through different ways). Profiles of the SWE standards, that take into account the specific needs of certain domains, are a good approach to improve interoperability within EO2HEAVEN and to reduce the necessary implementation effort. As a first step in this direction, a lightweight Sensor Observation Service profile for stationary in-situ sensors was developed. This profile can be seen as a least common denominator of domain specific profiles that ensures a common approach between these profiles. For example, this profile offers clear guidance on the minimum set of metadata to be provided to describe sensor data sets, and on which filtering mechanisms need to be supported when querying for sensor data.

Besides accessing sensor data through the EO2HEAVEN infrastructure, it is equally relevant to provide researchers an easy way to integrate and publish their observation data in the EO2HEAVEN Sensor Web architecture. In the past, the publication process required a general understanding of the SWE architecture and its different standards. However, such IT specific knowledge cannot be expected from a domain expert who is primarily interested in working with observation data. Thus the structure of the underlying databases as well as the Web service requests necessary for publishing observation data needs to be hidden from the user. To achieve this goal, EO2HEAVEN has developed an import tool for publishing sensor data on the Sensor Web. This tool provides the user with a graphical user interface to describe the structure and content of existing sensor data files (i.e. CSV files). Through this process the user is able to define how information such as the measured value, the time stamp of measurements, etc. is encoded in a data file. Using this information, the import tool is able to automatically convert the observation data to the corresponding SWE data models and encodings (i.e. according to the Observations and Measurements standard) and to insert the data into Sensor Web components.

After the sensor and health data sets have been published on the EO2HEAVEN SII, the next step is to make use of the data. Thus, it is necessary to link the sensor data sets provided through OGC SWE services to interoperable processing services which perform the processing required by the domain researcher's analyses. For this purpose the OGC Web Processing Service and Sensor Planning Service specifications have been chosen. Using the OGC Web Processing Service, researchers receive the opportunity to link any kind of input data provided by Sensor Web services to their processing and analysis models. By relying on interoperable standards, one can ensure a high degree of freedom in selecting data sets and accompanying processing algorithms: data sources and processing components can be flexibly exchanged as long as they adhere to the SWE and Web Processing standards of the EO2HEAVEN SII.

Further challenges addressed by EO2HEAVEN to improve the applicability of the SWE framework comprise the uncertainty of sensor data, the integration of remote sensing data (i.e. raster data sets such as satellite images) with in-situ data (e.g. local air quality measurements), and the discovery of sensors, sensor data, and Sensor Web services.

To sum it up, EO2HEAVEN relies with its SII on advanced Sensor Web concepts that are driven by the needs of the environment and health domains. Special focus has been put on increased interoperability and user friendliness, especially for non-IT domain experts. Thus, the Sensor Web technology is a core element for realising the Case Studies introduced in section 3.

6. Outlook and Future Work

The Sensor Web concepts used in EO2HEAVEN are mainly intended for a pull based access to sensor data. For example, if a researcher wants to perform an analysis using sensor data, he is able to request these data from a Sensor Observation Service (SOS) and subsequently receives the matching observation document. However, such a pull-based mechanism is not always sufficient. For example, an alerting application needs to receive new sensor measurements as soon as they are available in order to quickly dispatch any alerts. Thus, in the future further work should be done to extend the currently available publish/subscribe mechanisms of the SWE architecture. Especially the currently on-going activities of the OGC Publish/Subscribe Standards Working Group are likely to be an important input for future versions of the SII.

On the one hand, the case studies in Saxony and Durban either lack a fine-granular or an area-wide air pollution sensor network necessary to monitor the current local conditions properly. On the other hand, enlarging the number of regular sensors would be too expensive. To tackle both issues, extending the official sensor net by a larger number of cheaper sensors (e.g. http://www.kickstarter.com/projects/edborden/air-quality-egg) could be promising. Even if they provide less precise measurements, they are suitable to show a rise or fall of certain parameters. Additionally, a larger number of sensors is expected to even out measurement errors (assuming there are no systematic errors of a device class) and raise the quality of the sensor network as a whole.

An important aim of EO2HEAVEN is to establish persistent data accessibility of the various environmental data sources to the environmental science community via OGC SWE services beyond the project end. The EO2HEAVEN components developed for the Case Studies will be implemented in a long term and stable way, so that local staff are able to maintain their functionality in the future. The results of the EO2HEAVEN project will be published in a set of public deliverables and in a separate EO2HEAVEN book on best practices, which continues the book series started in the previous projects ORCHESTRA and SANY.

7. Conclusion

This paper has introduced the EO2HEAVEN project which provides a framework and tools for analysing the complex correlations and interdependencies between environmental factors and human health. Special challenges when combining environmental and health data were discussed: due to the highly sensitive character of health data special focus needs to be put on privacy and data protection issues. Furthermore, a common framework is needed that facilitates the integration of heterogeneous health and environment data sources into analysis and processing tools of domain experts.

Based on the three EO2HEAVEN Case Studies, the EO2HEAVEN Spatial Information Infrastructure (SII) was designed to support experts in their research. This ensures that the project results are in line with the requirements of practitioners.

Sensors are of special importance to determine the state of the environment (real-time as well as historic information). Consequently, the OGC Sensor Web Enablement (SWE) framework is a core element of the EO2HEAVEN SII. To support the needs of researchers, several advances to the SWE framework have been developed such as profiles for increased interoperability, user friendly data publication tools, concepts for handling uncertainty, and best practices for combining in-situ as remote sensing data.

In summary, the EO2HEAVEN project builds on results of previous European projects and advances these concepts to better support the integrated analysis of environmental and health data sets. Thus, researchers receive a valuable set of tools that support their work of understanding the influence of the environment on human health.

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 244100.

Bibliography

Antonowsky, A. (1996): The Salutogenic Model as a Theory to Guide Health Promotion, in: Health Promotion International, 11, pp. 11-18

Brauner, J. (et al.) (2011): Specification of Advanced Geo-Processing Services (second issue), Public deliverable D4.9, available online: http://www.eo2heaven.org/sites/default/files/D4.9_Specification_of_Advanced_Geo-processing_Services_Issue_2.pdf

Coco, A. (2005): The Cost-Effectiveness of Expanded Testing for Primary HIV Infection, in: Annals of Family Medicine, 3 (5), pp. 391-399

Jirka, S. (et al.) (2011): Specification of the Advanced SWE Concepts (second issue), Public deliverable D4.8 available online: http://www.eo2heaven.org/sites/default/files/D4.8_Specification_of_the_Advanced_SWE_Concepts_Issue_2.pdf

Kelly, C. S. (et al.) (2000): Outcomes Evaluation of a Comprehensive Intervention Program for Asthmatic Children Enrolled in Medicaid, in: Pediatrics 105, pp. 1029-1035

Kunz, S. (et al.) (2011): Specification of the SII Implementation Architecture (second issue) Parts I-VI; Public deliverable D4.7 available online: http://www.eo2heaven.org/sites/default/files/D4.7%20Specification%20of%20the%20SII%20Implementation%20Architecture.pdf

McConnell, R. (et al.) (2010): Childhood Incident Asthma and Traffic-related Air Pollution at Home and School, in: Environ Health Perspect., 2010 Mar 22 (Epub ahead of print)

Robins T. (et al.) (2002): Air Contaminant Exposures, Acute Symptoms and Disease Aggravation among Students and Teachers at the Settlers' School in South Durban. Interim Report, University of Natal Faculty of Health Sciences/University of Michigan School of Public Health, February 2002

Secker-Walker R. H. (et al.) (1997): A Mass Media Programme to Prevent Smoking among Adolescents: Costs and Cost-Effectiveness, in: Tobacco Control 6, pp. 207-212.

South African National Department of Health (2011): National Environmental Health Policy. National Department of Health

Strickland, M. J. (et al.) (2010): Short-term Association between Ambient Air Pollutants and Pediatric Asthma Emergency Department Visits, in: Am J Respir Crit Care Med, Apr 8 (Epub ahead of print)

Usländer, T. (et al.) (2007): Reference Model for the ORCHESTRA Architecture Version 2. OGC Best Practices Document 07-097, 2007, available online:
http://portal.opengeospatial.org/files/?artifact_id=23286

Usländer, T. (et al.) (2009): Specification of the Sensor Service Architecture Version 3.0 (Rev. 3.1). OGC Discussion Paper 09-132r1, 2009, available online:
http://portal.opengeospatial.org/files/?artifact_id=35888&version=1