EO2HEAVEN

mitigating environmental health risks

by the EO2HEAVEN Consortium

EO2HEAVEN

mitigating environmental health risks

Project Contact: Dr. Kym Watson
Mail: kym.watson@iosb.fraunhofer.de

Based on a Design by
Craig Stevens – www.craig-stevens.co.uk
Printed by Butler, Tanner & Dennis, UK

[1] www.eo2heaven.org/node/4

Editor: Martin Klopfer, OGCE
Contributing Authors and Project Participants:
52N: Bache, Felix; Bredel, Henning; Bröring, Arne; Buurman, Merret; Fiedler, Eric; Hollmann, Carsten; Jirka, Simon; Jürrens, Eike Hinderk; Nüst, Daniel; Osmanov, Artur; Proß, Benjamin; Rieke, Matthes; Schäffer, Bastian; Schmidtchen, Christian; Schulte, Jan; **ATOS:** Barrio, Alejandro; Esbrí, Miguel Angel; López, Jorge; Lorenzo, Jose; Nuñez, Mario; Palomares, Angel; Peiro, Millán; **BRGM:** Boulahya, Faiza; Pasquier, Fabien; Perrier, Pascal; Tertre, François; Tellez-Arenas, Agnès; **CSIR**: le Roux, Wouter; du Preez, Martella; de Klerk, Leanie; Hohls, Derek; McFerren, Graeme; Vahed, Anwar; van der Merwe, Marna; van Zyl, Terence; **IOSB:** Ballin, Wolfgang; Batz, Thomas; Kunz, Siegbert; Mulder, Johannes; Rudolf, Michael; Schmieder, Martin; Stumpp, Jörg; Usländer, Thomas; van der Schaaf, Hylke; Watson, Kym; Wieser, Martin; **ITC:** Boerboom, Luc; Duim, Job; Horn, John; Lemmens, Rob; Maathuis, Ben; Pei, Linlin; Retsios, Bas; Schouwenburg, Martin; **JRC:** David Rickerby; Andreas Skouloudis; **KIT:** Bakker, Mirjam; Maldonado, Fernando; Scheelbeek, Pauline; Rood, Ente; van Beers, Stella; **MAK:** Mpungu, Steven Kiwuwa; Kakumba, Rose; Luswa, Lukwago; Malimbo, Mugaga; Wamala, Joseph; Bwambale; Ericana; Madinna, Mugisha; Yusuf, Baseka; Adriko, Charles; **OGC**: de Lathouwer; Bart, Klopfer; Martin, Percivall, George; Simonis, Ingo; **SPOT:** Caldairou, Valérie; Jacquet, Carole; Cauchy, Arnaud; Mérigot, Philippe; Favre, Sébastien; Eichelbrenner, Fréderic; Escalbert, Bruno; **TUD:** Arloth, Jana; Elsner, Bernhard; Koch, Rainer; Siegert, Joachim; Maas, Hans-Gerd; Pech, Katharina; Richter, Anette; Stelling, Nadine; Bernard, Lars; Brauner, Johannes; Grafe, Bernd; Hutka, Luise; Kadner, Daniel; Karrasch, Pierre; Mendt, Jacob; Mittig, Karsten; Richter, Silke; Wiemann, Stefan; **UKZN:** Ahmed, Fethi; Moodley, Deshen; Naidoo, Rajen; Padayachee, Yerdashin; Ramsay, Lisa; Reid, Caroline; Spangehl, Brett; Naidoo, Shalin; Simpson, Andrew; Chetty, Siva; Maharaj, Sharveen; Narayansamy, Maduray Lennie.

Why EO2HEAVEN?	**1**	**5**
Is this Book for You?	1.1	5
The Need for EO2HEAVEN	1.2	5
Using this Book	1.3	7
EO2HEAVEN in a Nutshell	**2**	**9**
EO2HEAVEN Objectives	2.1	9
Project Approach	2.2	9
Key Results	2.3	10
EO2HEAVEN Case Studies	2.4	11
Current Status Health and Environment	2.5	16
Common Challenges	2.6	19
The EO2HEAVEN Value Proposition	2.7	22
The Bigger Picture	**3**	**24**
GEOSS	3.1	24
EO2HEAVEN Contribution to GEOSS	3.2	29
INSPIRE	3.3	32
COPERNICUS	3.4	34
The EO2HEAVEN Approach	**4**	**35**
Scenario Based Engineering	4.1	35
The Spatial Information Infrastructure	4.2	38
Data We Used	4.3	43
Methods We Applied	4.4	46
Community and Capacity Building	4.5	54
The Impact of Air Quality on Respiratory and Cardiovascular Diseases in Saxony	**5**	**58**
Expected Stakeholders Benefits	5.1	58
Bounding Conditions and Challenges	5.2	60
Air Quality Modelling	5.3	63
Health Risk Analysis	5.4	68
Results for Cardiovascular Diseases	5.5	70
Results for Respiratory Diseases	5.6	72
Web-Based Result Dissemination	5.7	73
Conclusions	5.8	75
Relationship Between Industrial Pollutant Exposure and Adverse Respiratory Outcomes	**6**	**78**
Expected Stakeholders Benefits	6.1	78
Bounding Conditions and Challenges	6.2	80
Air Quality Modelling	6.3	83
Health Risk Modelling	6.4	84
Durban Web Client	6.5	84
Conclusions	6.6	87
Links Between Environmental Variables and Cholera	**7**	**88**
Expected Stakeholders Benefits	7.1	88
Bounding Conditions and Challenges	7.2	89
Cholera Research Outcomes	7.3	89
Integrated Approach to Study Cholera Dynamics	7.4	91
Correlation Analyses Processes	7.5	92
Conclusions	7.6	93
Cholera Case Surveillance and Health System Enhancement	**8**	**96**
Expected Stakeholders Benefits	8.1	97
Bounding Conditions and Challenges	8.2	97
Disease Incidence Reporting App	8.3	98
Conclusions	8.4	100
Disease Mapping and Spatial Analysis for Disease Control	**9**	**102**
Mapping for Public Health	9.1	102
Health Assurance	9.2	108
System Coverage	9.3	109
Access to Health Services	9.4	110
Using Earth Observation Data	**10**	**111**
Remote Sensing Data Modelling	10.1	113
Study on Using UAVs	10.2	118
EO2HEAVEN Software	**11**	**123**
Time2Maps	11.1	123
Generic Advance Client	11.2	127
Mobile App to Register Cholera Cases	11.3	129
Community Collaboration on Models	11.4	130
OGC WMS Wrapper for WPS	11.5	132
Sensor Web Components	11.6	133
References	**12**	**140**


LOADING 100%
MEDIA
BUSINESS
WORLD
-EUROPE
-AMERICA
-ASIA
-AFRICA
Projected sales of main products in 2013
NETWORK SEARCH
Share of market activity
25%
75%
Passive market share
RLD

Why EO2HEAVEN? 1

IS THIS BOOK FOR YOU? 1.1

- Are you dealing with the impact of environmental factors on human health?
- Have you ever wondered about putting medical data in a geographical context?
- Would you like to leverage satellite data to complement environmental field data?
- Do you need to integrate data to issue alerts based on environmental conditions and their potential impact on health?
- Are you looking for options to share existing data across domain boundaries?

If your answer to any of these questions is yes, then this book may provide you with some interesting aspects on how current trends in geospatial technology and global initiatives can help to bridge the divide between environmental studies and health research.

THE NEED FOR EO2HEAVEN 1.2

The project acronym EO2HEAVEN stands for Earth Observation and Environmental Modelling for the Mitigation of Health Risks and is a major collaborative project funded under the 7th Framework Programme of the European Commission.

EO2HEAVEN has the primary objective to contribute to a better understanding of the complex relationships between environmental changes and their impact on human health. To achieve this, the project followed a multidisciplinary and user-driven approach, involving public health stakeholders as well as technology and service providers in both the Earth observation and in-situ environmental monitoring domain.

As a result of this collaboration, EO2HEAVEN designed and developed methods and tools to correlate environmental data with exposure and health data, to support the collection and integration of data and to visualise results in their geographical context. The overall aim is to support research of human exposure and early detection of potential health endangerments.

EO2HEAVEN builds on the results of two preceding major projects funded by the European Union:

- The ORCHESTRA project, which was completed in 2006, developed a standards based reference architecture for risk management, with a focus on natural disasters, such as earthquakes or forest fires.
- The SANY project was completed in 2009 and focussed on the development of Sensor Web Enablement standards, complementing the work of ORCHESTRA.

Both, ORCHESTRA and SANY, supported the development of internationally recognised interoperability standards, that enabled exchange of information from heterogenous IT systems across technological and administrational boundaries.

The underlying concepts and approaches, as well as the lessons learned from both of these projects, were leveraged by EO2HEAVEN to address the challenges of health and environment studies: whilst the former projects dealt primarily with IT aspects of the geospatial domain, we are now building a bridge that allows stakeholders from the health community to utilise the huge potential of gathering, mapping and analysing their data in the spatial and socio-geographical context.

The synergies of close cooperation between stakeholders from the health domain and the environmental studies domain and the use of common technology approaches are a major stepping stone to better informed decisions for a safe environment and good public health.

In the beginning there may be merely raw observation data, intelligible only to the scientists: a value for ozone or particulate matter concentration, observed as part of an environmental monitoring activity; a value for personal vulnerability of a patient, recorded as part of a medical study. The final result of processing this data should be information tailored to the specific needs of the targeted audience. This can be used on various levels of decision making: from large scale governance of a good and healthy environment, to the planning of individual preventive measures, such as sms alerts to asthma patients based on real-time air quality indices.

To this regard, what has been stated for SANY, is today even more relevant when data from various areas of expertise needs to be integrated to derive actionable information.

Our biggest challenge is to ensure that observations can be turned from data into information; information leads to knowledge, to understanding and, ultimately, understanding may even lead us to the wisdom to act accordingly.

USING THIS BOOK 1.3

This book is split into three mostly self contained parts. Depending on your primary interests we suggest you start with the part you feel most comfortable with:

The Project Perspective 1.3.1

The initial chapters of this book summarise the relevant facts of EO2HEAVEN, put them into the context of major ongoing initiatives and outline the approach taken. Reading this section will provide you with:

- A concise overview on the EO2HEAVEN project and case studies
- An introduction to the common challenges
- A review of the current initiatives that influenced our work
- An introduction to the user driven project approach

The Scenario Perspective 1.3.2

The middle part of this book presents in detail the case studies that were carried out by the project consortium in close cooperation with local community stakeholders:

- Impact of Air Quality on Respiratory and Cardiovascular Diseases
- Relationship between Industrial Pollutant Exposure and Adverse Respiratory Outcomes
- Links between Environmental Variables and Cholera
- Cholera Case Surveillance and Health System Enhancement

For each study the descriptions follow a similar pattern. After introducing the scenario setting, the following questions are addressed in detail:

- What are the expected stakeholders benefits?
- What are the bounding conditions and challenges?
- What workflow was chosen to process the data?
- What are the results?
- What are the conclusions?

Each part concludes with a dedicated chapter that summarises our lessons learnt and provides recommendations, should you wish to implement your own project following the EO2HEAVEN examples.

1.3.3 The Software Perspective

In the last part of the book you can find a detailed description of the software components, services and concepts that were developed by the EO2HEAVEN partners and used to implement the workflows described for each of the case studies.

EO2HEAVEN in a Nutshell 2

EO2HEAVEN OBJECTIVES 2.1

The objective of EO2HEAVEN is to contribute to a better understanding of the complex relationships between environmental changes and their impact on human health, with an emphasis on three particular case study areas.

The main result of the project is the design and development of a toolbox based upon an open and standards-based Spatial Information Infrastructure (SII), which can support research of human exposure and early detection of potential health threats. For this reason the project developed models to relate environmental data with exposure and health data.

In addition EO2HEAVEN examined different Earth Observation products, especially those resources available free of charge for the research community. In order to study the impact of human activity on health the project took advantage of the availability of this data to evaluate its potential for detecting and mapping environmental variables. For this purpose EO2HEAVEN also worked on the integration of remotely sensed and in-situ environmental measurements.

PROJECT APPROACH 2.2

Projects like EO2HEAVEN only provide a sustainable value as long as they are carried out in close collaboration with local stakeholders. The significant amount of research work that was undertaken, may not directly impact today's IT solutions, but the strong interest and commitment of the consortium partners to engage with potential system users and stakeholders at an early stage and involve them in the *design – build – validate* cycles was of tremendous value to understand the IT needs of the health domain from a geospatial perspective and likewise raised awareness of the huge potential of geospatial approaches to health research topics.

A recommended way to formalise the collection of user requirements is described in the ISO Reference Model for Open Distributed Processing (RM-ODP). This is an international standard for architecting open, distributed processing systems that provides an overall conceptual framework for building distributed systems in an incremental manner, and offers a good starting point to identify user requirements and translate them into generalised specifications. Once the real world requirements for a certain scenario are documented, these can be broken down into requirements for the system environment and software development based on suitable Open Standards.

To ensure that developments meet exploitation requirements, the project followed an iterative approach of three cycles of the following steps, in which the results of each completed cycle were used to further refine the requirements for the following phase:

- Identification of user requirements,
- Development of system and architecture specifications,
- Implementation of prototypes,
- Validation by end-users.

The SII and toolbox based approach therefore facilitates the set-up of observation and decision support systems that rely upon the correlation and fusion of Earth observation, in-situ and human health data.

2.3 KEY RESULTS

The already mentioned SII and collection of models, processing chains and software tools are described in this book, with more details being available on the EO2HEAVEN website (eo2heaven.org). The iterative approach and close cooperation with the project partners allowed a steep learning curve to be mastered. Concepts which are technically state-of-the-art or beyond may be a feasible approach to a requirement, but can still turn out to be not suitable on an inter-organisational or cross-domain context. In particular the handling of health data and the dissemination of analysis results to the public is a sensitive issue and has to be approached with extreme care and respect to privacy and ethical issues. The relationships between environmental factors and human health are manifold and very complex. Adding to this, every human being carries his or her personal biography and thus has an individual vulnerability to any given exposure. Hence, a risk map can only provide an indication of a generalised risk level, but with a limited certainty for the individual person.

Yet, research in this field is vastly progressing. For your own projects, the lessons learnt in EO2HEAVEN may help you to avoid potential obstacles in the planning stage and to utilise the concepts and tools as a starting point. In this regard, one or more of the following public EO2HEAVEN documents might be of interest to you, helping to define a realistic project plan and overcome initial hurdles:

1. The collection of requirements for the EO2HEAVEN scenarios and their validation at the end of the implementation cycles.
2. An overview of current Earth observation data sources, environmental information products and relevant organisations.
3. A description of the EO2HEAVEN processing chains and used models.
4. A description of the EO2HEAVEN architecture concept and software tools.
5. A collection of educational material for decision makers and technicians interested in developing their own health applications.

We have compiled the most relevant experiences and results in this book, which is complemented by online training materials and the public deliverables. However, do not hesitate to contact the EO2HEAVEN Consortium (eo2heaven.org) for further information or support when needed.

Once again, it has also been recognised that the discussion not only amongst project partners, but in the community of like-minded experts across domain boundaries often helps to find the missing building block for new approaches and solutions. The networking of the stakeholders should therefore start as early as possible in each project.

To this regard, the EO2HEAVEN project was a major stepping stone to bring experts from the very different areas of expertise together. The experience of the joint engagement in developing standards-based tools for the scenario requirements has raised a high level of awareness for each others requirements and capabilities as well as domain specific boundaries.

EO2HEAVEN CASE STUDIES 2.4

Impact of Air Quality on Respiratory and Cardiovascular Diseases 2.4.1

The first case study focuses on respiratory and cardiovascular diseases in relation to the environmental parameters ground-level ozone, particulate matter and a number of meteorological conditions for the Federal State of Saxony, Germany.

The purpose of this Case Study is to develop an online information system so that preventive measures can be taken to avoid adverse

health effects resulting from environmental air pollution and extreme weather events.

An analysis of mortality and morbidity rates in Germany revealed that in the adult population cardiovascular diseases and respiratory diseases are among the most common causes of death. Both are known to be driven to some extent by environmental aspects such as exposition to air pollution and extreme weather events. Cardiovascular diseases and respiratory diseases such as asthma have been chosen as an EO2HEAVEN scenario, because they allow direct correlations between environmental factors, such as pollen, ozone and particulate matter exposition and resulting symptoms on a comparatively short time-scale (Meyer et al. 2007).

The percentage of people affected by allergies (i.e. bronchial asthma, contact allergies and others) has increased steadily in recent years. This is particularly true for younger age groups (25 to 29 year olds). The increase has been stronger in the male than the female population (7.7 % versus 3.1 %). About 13 % of the children and about 20 % of the adult population have been diagnosed with allergies. These rates are very much in accordance with the average prevalence rates in other European countries. Regarding cardiovascular diseases Saxony has one of the highest burdens of disease of all German federal states (211 cases of death per 100.000 inhabitants for women and 316 per 100.000 for men compared to the German mean of cases of death due to cardiovascular diseases: 192 per 100.000 inhabitants and 275 per 100.000, respectively) (Robert-Koch-Institut 2011).

Allergic asthma is one of the most common chronic diseases in childhood in Germany and hence in the investigated area. Several European and international studies have verified the association between health and environmental conditions. According to the German KIGGS study approximately 4.7 % of the children aged under 17 suffer from allergic asthma. There is a significant difference between genders, with 5.5 % of the boys and 3.9 % of the girls being diagnosed with asthma (Schlaud et al. 2007). The German Allergy and Asthma Alliance even state that 10 % of the children suffer from asthma (Deutscher Allergie- und Asthmabund 2010). About 5 % of the German adult population are diagnosed with allergic asthma. The results of the national health survey suggest that the prevalence rate in the investigated area was lower than in the Western part of Germany. During the past decade, however, there has been a steady increase in the East of Germany and thus in Saxony as well. Each year about 5.000 deaths are attributed to asthma in Germany and a literature review suggests that this figure is representative of European levels (Heinrich et al. 2002).

So far, there has been little sound evidence on how environmental hazards may influence human health in Saxony. A recent retrospective ecological study has investigated the association between particulate matter (PM10) and ultrafine particles (UFP) in the region of Leipzig, Saxony (Franck et al. 2011). The results revealed that there was an association between the diagnosis of a hypertensive crisis and exposition to UFP with a lag of two days. No associations were found regarding PM10 or PM2.5.

Relationship Between Industrial Pollutant Exposure and Adverse Respiratory Outcomes 2.4.2

The second case study focuses on the relationship between industrial pollutant exposure and adverse respiratory outcomes. Much debate exists around causality of effects, role of specific pollutants and the populations particularly vulnerable to elevated pollution. Although certain pollutants, such as oxides of nitrogen, particulate matter and sulphur dioxide are known to result in adverse outcomes, the ability to use this information in developing interventions to improve the life of affected and vulnerable sub-populations is limited.

The purpose of the case study is to develop a system, which permits ready access to pollutant and meteorological data that allows for the prediction of exposure of vulnerable populations in areas that are at risk for elevated excess ambient pollutant exposure.

The geographical focus is on the southern sections of Durban, known as the Durban South Industrial Basin (DSIB). This area is a densely populated area, consisting of approximately 250.000 people, is a key industrial hub for the city and the country as a whole. This residential-industrial complex arose during the era of apartheid governmental planning, and today represents a highly conflictual situation between stakeholders. This area has been reported to have a high unemployment rate, with approximately 52% of the population not economically active.

The Durban South Industrial Basin (DSIB) is at particularly high risk for exposure to significant levels of ambient air pollution because of its geographic relationship with certain stationary sources of air pollutants. Specifically, two major petroleum refineries are within the community, together with a pulp and paper manufacturer, a waste water treatment plant and several small to medium industries. Up to a few years ago, each of these refineries has emitted, on average, in the range of 35.000 to 40.000 kg of SO_2 per day. Through rigorous monitoring and enforcement, the total industrial SO_2 emissions were reduced from 107 tons per day in 2000, to 61 tons per day in 2005. Owing to a combination of its geographical relationship to the refineries, land contours, prevailing meteorological conditions, the use of a relatively short emissions stacks at these facilities (50 – 100 meters), the lack of or relative ineffectiveness of emission control devices on refinery stacks, the many sources of so-called fugitive air emissions at refineries, emissions from industrial and passenger vehicles, as well as the proximity of other industries and, until recently, the Durban International airport, the community is believed to be at risk for intermittent substantial exposure to ambient air pollutants. Available data on sulphur dioxide indicate that average and/or maximum exposures at sites in south Durban have frequently exceeded World Health organization (WHO) and the South African Department of Environmental Affairs (DEA) standards.

Health studies in the area have indicated elevated risk for respiratory outcomes among those exposed. In a study among students and teachers at the Settlers' Primary School, unusually high prevalence rates were reported for asthma, with ranges of any type of non or probable asthma (symptoms assessed) from 53.5% to moderate to severe persistent asthma of 16.8%. In addition, approximately 20% of the study sample had marked airways hyperresponsiveness as diagnosed by methacholine challenge testing, the prevalence higher than any other population based reports in the scientific literature (Kistnasamy et al, 2008). This study found statistically significant associations between prior day and prior 48 hour PM10, SO_2, and NO_2 levels (continuously measured at the school) and increased respiratory symptoms and diminished pulmonary function measures (measured by digital recording peak flow meters) among students with persistent asthma. These effects were observed during a time period when all ambient pollutant measures were well within national and international standards.

In a more comprehensive study, ambient and indoor exposure monitoring was conducted in south and north communities during the study period, May 2004 to February 2005. Additional sources of pollutant and meteorological data covering January 2004 to October 2005 were accessed. Pollutants monitored included sulphur dioxide (SO_2), carbon monoxide (CO), volatile organic compounds (VOCs), particulate matter in two size classes (PM10 and PM2.5), ozone (O_3), metals, and semi-volatile organic compounds. The health data from that study suggested high prevalence of respiratory disease. Caregivers reported that 14.7% of children had been diagnosed with asthma by a doctor, 10.9% with hayfever, and 4.0% with chronic bronchitis. Based on symptoms described by the caregiver, 31.5% of children had some grade of asthma, with 7.9% having mild persistent asthma, and 4.1% having moderate to severe persistent asthma. Objective lung function assessments using methocholine challenge testing were well correlated with these reported assessments, with 7.6 % having marked bronchial hyperreactivity (BHR) (PC20 < 2 mg/ml) and a further 17% having either probable or possible BHR. The covariate-adjusted prevalence of marked BHR varied substantially between the south (8.0%) and north (2.8%). Logistic regression models adjusting for age, gender, race, caregiver education, household income and the presence of smokers in the household, found that attending a school in the south was statistically significantly associated with increased risk for persistent asthma (Naidoo et al., 2007).

The relatively moderate ambient concentrations of NO_2, NO, PM10, and SO_2 were strongly and significantly associated with decrements in lung function among children with persistent asthma and/or genetic polymorphisms associated with reduced ability to respond to oxidative stress. Moreover, attending primary school in south Durban, as compared to the north, was significantly associated with increased risk for persistent asthma and for marked airway hyper reactivity in covariate-adjusted regression models.

Links Between Environmental Variables and Cholera 2.4.3

The third case study of the EO2HEAVEN project focuses on links between certain environmental, including climatic, variables and the outbreak of cholera in Africa. The Case Study includes the use of remotely sensed and in-situ data together with ex-situ data, i.e. data collected as part of a field sampling campaign and health data, e.g. cholera case data. Health data availability and accessibility and the costs and logistics associated with a field sampling campaign determined the final selection of a study area on the south western parts of Uganda, specifically the Kasese area close to Lake George and Lake Edward; the Kazinga Channel; the Lubila Tako Chaka River that forms the border between Uganda and Democratic Republic of the Congo and surrounding areas.

The relevance of scale, i.e. the aggregation of data at different temporal and spatial scales when investigating the association between different variables and the disease, is an important factor that has been addressed in this case study. Environmental factors and processes operate at different spatial and temporal scales, which necessitate the extraction and analysis of environmental data at different scales together with the health data.

Environmental factors such as the accumulation of rainfall water over a period of time and a large spatial extent correlate more closely with case data whereas daily sunlight and water temperature values are associated with the pathogen dynamics.

The purpose of Case Study 3 is to gain insights into the potential environmental and climatic drivers of cholera outbreaks in the Ugandan and central African context and the dynamics of the

***Vibrio cholerae* pathogen in the natural environment.**

The environmental risk of potential outbreaks in cholera endemic areas, i.e. areas which report regular outbreaks or areas where the outbreaks generally start, has been estimated using satellite and in situ data together with field and laboratory results and, where possible, taking into account demographic and socio-economic factors.

Weather data has been used to track changes in rainfall and air temperature over time, whereas predicted climate data, generated by the downscaling of global circulation models to a regional level in order to incorporate local conditions, has been used to determine areas potentially under threat of cholera outbreaks under changing climate conditions.

The level of vulnerability of communities in different areas due to the environmental conditions and the general socio-economic situation in these areas has been analysed and is highlighted in interactive online maps, which are based on current conditions and provide an indication near real time. The environmental risk estimates can be used as the basis for further developments towards an early warning system.

2.5 CURRENT STATUS OF HEALTH AND ENVIRONMENT

2.5.1 Air Quality and Health

The multi-dimensional relationship between health and environment is a matter of concern all over the world. One aspect of this relationship is the connection between air pollutants and meteorological parameters, the environmental status and human health. This link has been well documented in international and European studies. Research has shown that acute as well as chronic respiratory and cardiovascular diseases are highly influenced by ambient air pollution concentration (Strickland et al. 2010; McConnel et al. 2010).

Although much environmental as well as health data proves this connection, detailed understanding of the complex relationship is still rather limited due to, for instance, the difficulty of collecting and linking relevant data at the right geographical and temporal scales. What makes an analysis and understanding of this data even more complicated is that it is extremely difficult to demarcate environmental effects on human health from other influencing factors (e.g. nutrition, living and working conditions, genetic make-up). The consideration of environmental effects on human health is in addition complex and requires careful thought regarding the mutual dependence of influencing factors and resulting cumulative effects. Different population groups face different risk and exposure levels. The chronically ill, babies, children, the elderly and economically deprived populations living under low housing conditions are at a higher risk of suffering from environmentally related diseases compared to the general population. It is therefore indispensable to take account of individual vulnerability levels to undertake sound preventive measures.

To assess the health risks resulting from exposure to air pollutants, it is essential to consider air pollution and meteorological data together with human health data, and, if possible, socio-economic data and confounding factors. Consequently there is a need for constant monitoring of air pollution as well as inpatient and outpatient data. Besides other factors, the risk of mortality and morbidity increases with the number of air pollutants and the duration of exposure. Efforts at European and international level take into consideration at least two air pollutants, which is essential for the development of an online information system. Against this background, an EO2HEAVEN aim was to contribute to increased understanding of the complex relationship between environmental factors and human health.

Environmental Triggers for Cholera Outbreaks 2.5.2

Cholera is a complex disease driven by a non-linear combination of environmental and human related factors, which interact on the human, human-environment, and aquatic ecosystem levels. Research into diseases in general and cholera in particular is constrained by a combination of factors ranging from the lack of knowledge of how to access and use satellite data, the availability of accurate environmental and health data, the accessibility of disparate datasets from different sources, to the availability and accessibility of tools to find, extract, reformat, analyse, model and visualise data and results.

This case study addresses the data issues experienced by health researchers when studying a complex environmental related disease such as cholera and by decision makers in making timely decisions regarding intervention options and strategies.

Tools to find, extract, reformat, analyse, model and visualise data and results are desired to support users from different groups such as researchers and government.

In developed and developing countries, people are affected by emergent and re-emergent diseases. Changes in human behaviour and activities, and/or the adaptive and recombinant properties of responsible pathogens, drive these changing disease patterns. Environmental, including climatic conditions, play an important role in the emergence and re-emergence of some of these diseases. Cholera is a re-emerging disease that affects an increasing number of countries. Fifty-six countries reported cholera outbreaks to the WHO in 2008 (WHO WER, 2009). About 64 % of these cholera reporting countries were from Africa. Traditionally cholera outbreaks originate in coastal areas before spreading inland. This is not always the case and in central and south eastern Africa, outbreaks often start in areas located in close proximity to inland lakes.

Although cholera is associated with poverty and the lack of clean water, sanitation and good hygienic practices, the pathogen responsible for outbreaks has been detected in water sources in both developing and developed countries.

The relative contribution of human practices versus environmental and climatic conditions to cholera outbreaks and the severity and spread of outbreaks is often difficult to determine. Environmental and climatic factors and conditions seem to play a major role in the dynamics of the disease in poor countries located in the tropical and subtropical regions. Environmental factors such as phytoplankton abundance and biomass production, water temperature, salinity, amount of sunlight penetrating water bodies and the availability of nutrients have been shown to play a significant role in the presence and persistence of the bacterium in the aquatic system and the transmission of the disease in areas where people depend on untreated water sources. Changes in these factors and conditions are driven to a large extent by rainfall and temperature patterns, which in turn are affected by climate variability and change.

COMMON CHALLENGES 2.6

Cross Domain Communication 2.6.1

In a classical communication pattern, sender and receiver need to share the same understanding of a topic. However, this is not always true for cross domain communication because of different working methods, background knowledge or organisational structures of the participants. This needs to be bridged to avoid information loss or, even worse, misinformation.

EO2HEAVEN project scientists from the environment, health and the IT domains join forces. The challenge of building a common knowledge base for both groups needs to be addressed in order to get health and technical requirements across to all participants. Therefore, continuous coordination and collaboration effort is required.

Spatial and Temporal Data Fit 2.6.2

To receive meaningful results from the comparison and correlation of information from different sources, a common spatial and temporal reference is a key requirement. This applies to environmental data, such as in situ observations, remote sensing data or laboratory field data, as well as for health data on prescriptions, mortality or morbidity.

Having the same spatial reference usually means that information covers the same area of interest on the Earth's surface, like administrative units or postal code areas. On the other hand, the temporal reference describes a point or duration in time. The target scale for data analysis for both, spatial and temporal reference, is determined by the lowest resolution of all input datasets. Thus, data needs to be interpolated or aggregated, if at least one input for data analysis is only available on a lower resolution.

In general, it is advisable, that a combined analysis of information from different sources should be based on the same spatial and temporal reference, to ensure that bounding conditions, which are not included in the analysis, are comparable.

Accuracy of Analysis Results 2.6.3

The accuracy of a scientific analysis depends on three main aspects:

- The accuracy of input data
- The analysis or model used
- The interpretation of results

Unfortunately, uncertainty measurements and their meaningful documentation are often neglected at all of those stages, although such accuracy information is important to prevent misinterpretations of analysis results and improper decision making.

Uncertainty of input data usually depends on the accuracy of measurements and can be determined by conducting validation or descriptive statistical analysis on the data. It becomes more difficult when assigning uncertainty to models, as it requires extensive model validation, taking into account the uncertainty of input data and modelling workflows. Yet, the greatest challenge is to enable a correct interpretation of results, which heavily relies on personal background knowledge, expertise and subjective perception. Thus, decision making must be supported by a detailed documentation of the analysis and mechanisms for uncertainty propagation.

2.6.4 Data Privacy & Ethical Issues

Concerning the use of data two distinct conditions have to be distinguished:

Copyright and intellectual property rights apply in the same manner as in every other case citing data of extrinsic origin. These questions are always persistent in the evaluation of secondary data.

Furthermore data protection legislation may limit the access to and the publishing of health data. According to the professional codes for physicians in the different German states (based on state data protection laws, that have a guiding basis in the German federal data protection law, which itself has a legal basis in the European data protection directive [Directive 95/46/EC]) individual-related data and fully anonymised data have to be distinguished.

To process, use or publish individual-related data the corresponding individuals have to each give personally – his/her legally binding consent (for health data usually in written form). Fully anonymised data, which has been processed in a way that re-individualization is impossible with “normal” effort, may be used openly only with the limitations of the above mentioned copyright and intellectual property restrictions. However, this may adversely impact the spatial and temporal resolution (cf. Section 2.6.2).

2.6.5 Availability of Health Data

Although in most European countries various health data sources are available, data availability in countries such as South Africa or Uganda is much more limited. The absence of health insurance data for the majority of the population, lack of district level HMIS or the presence of representative samples of health studies result in an inadequate characterisation of the impacts of exposure on health. This in turn impacts on health policy development, implementation of interventions or resource allocations.

The case studies within EO2HEAVEN on both cholera and respiratory diseases attempted to circumvent some of these challenges and provide alternative data approaches to address the health concerns.

IT Challenges 2.6.6

While working with researchers and stakeholders from different scientific disciplines, it became apparent that before the newest trends and developments in mainstream IT can be utilised for software development, the following three key aspects need to be addressed first:

- To gain user acceptance, the software usage needs to be as simple as possible. This means to reduce the number of available features to a minimum and focus on the important ones instead.
- Slow and unreliable Internet connections are a common problem, especially in developing countries. Software needs to be usable offline and be able to synchronize its data when an Internet connection is available.
- Landline Internet is often not readily available in developing countries, whereas accessing the Internet via mobile phones and smartphones is common and far more reliable. Accordingly, the development of mobile applications, which are usable on a tablet pc or smart phone, should be considered to provide access to traditional pc based software applications.

Researchers often face diverse challenges depending on many different factors and therefore want to stay in full control of their workflows. Accordingly, this user group usually does not look for a black box one-stop solution that solves them. To them, offline access is very important, because they want to use 'their' own software and just require access to the datasets. Providing this user group with help on finding suitable datasets, tools to import and export these datasets into and out of their own software, and meaningful information (metadata) about what has already happened with these datasets are far more important factors for acceptance than initially anticipated.

2.7 THE EO2HEAVEN VALUE PROPOSITION

Monitoring environmental aspects and public health and correlating both starts with the collection, processing, and interpretation of data on hazards, exposures and health outcomes. The results can then be disseminated in various formats, ranging from processed data that may be used as input for further processing to reports or risk maps on specific topics. This process is illustrated below:

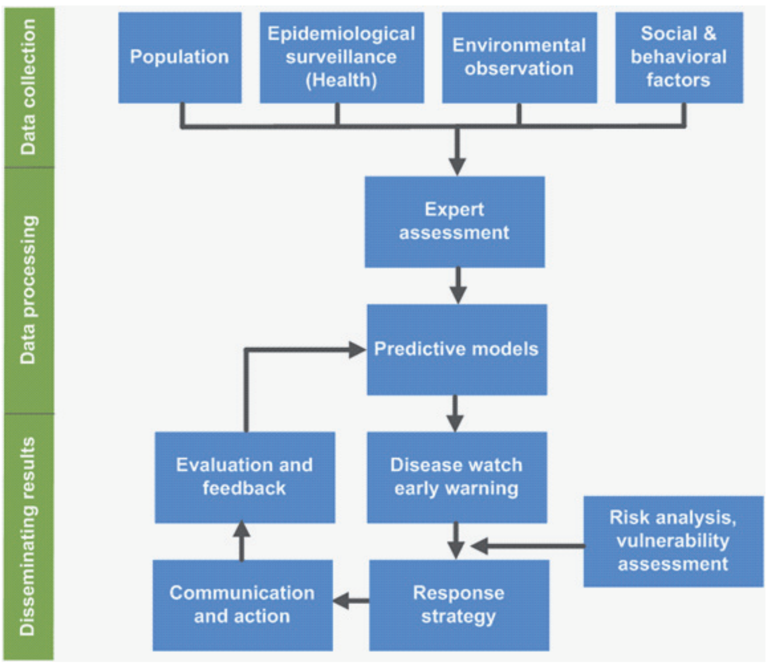


Diagram adapted from Trtanj 2009

Direct actors in the management cycle of environment and health threats include environmental and health managers, health practitioners and the research community. But also boundary institutions such as operators of sensor networks, environmental advocacy groups, etc. need to be addressed in any project that intends to make a lasting impact. Several challenges complicate environmental public health monitoring:

- The ability to link specific environmental causes to adverse outcomes is often limited by our poor understanding of disease processes, long lead times, inadequate measures of exposure, and multiple potential causes of a disease.

- Data collected for other purposes, e.g. Health Insurance Data, rarely includes sufficient information to meet a case definition for a condition caused by an environmental agent.
- Issuing a public alarm is often out of proportion to the hazard of concern, and sentiment will often influence public policy disproportionately to scientific information (Peter Nsubuga et al, 2006).

EO2HEAVEN tackles these challenges by providing methodologies, correlation models, spatial data services (OGC Standards) and applications (SOA paradigm) supporting the main activities involved in environmental health:

- discovery and acquisition of data sets
- integration of heterogeneous Earth observations (satellite, insitu and field campaign data)
- extraction of time series and visualization of graphs and maps
- development of models of health effects
- development of risk maps
- development of predictions forearly warning systems

Those aspects help to overcome the expected complexity of integrating health and Earth observation data. Key issues addressed include the general availability of health data as well as related data privacy aspects that require substantial aggregation of health data, nonmatching spatial and temporal resolutions, different approaches during the early data cleaning phases, and data availability and accessibility in general.

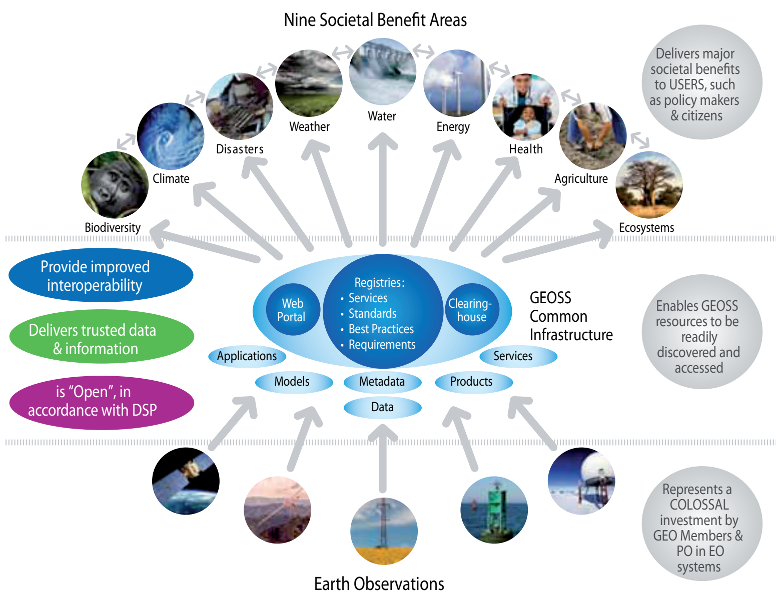
EO2HEAVEN components and workflow systems provide a huge improvement on cross-domain research as it is addressed here. Scientists get new tools at hand that help to shorten traditionally labor-intensive tasks, such as data discovery, pre-processing and integration. Being compliant to international standards helps to ensure sustainable developments and builds a solid base for future research.

3 The Bigger Picture

3.1 GEOSS - A GLOBAL EARTH OBSERVATION SYSTEM OF SYSTEMS

The Group on Earth Observations (GEO) is coordinating efforts to build a Global Earth Observation System of Systems (GEOSS). GEO was established in February 2005 by the Third Earth Observation Summit meeting in Brussels. This followed calls for action by the 2002 World Summit on Sustainable Development and the Group of Eight (G8) leading industrialised countries.

GEO is a voluntary partnership of governments and international organizations. It provides a framework within which these partners can develop new projects and coordinate their strategies and investments.



(c) www.earthobservations.org

The vision for GEOSS is to realize a future wherein decisions and actions for the benefit of humankind are informed by coordinated, comprehensive and sustained Earth observations and information.

GEOSS aims to meet the need for timely, quality, long-term global information as a basis for sound decision-making and will enhance the delivery of benefits to society. GEOSS aspires to encompass all areas of the world, and to cover in-situ, airborne and space-based observations. Many efforts are oriented towards solving the issues of standardisation and interoperability within GEO.

GEOSS addresses nine Societal Benefits Areas (SBA) of critical importance to people and society. It aims to empower the international community to protect itself against natural and human-induced disasters, understand the environmental sources of health hazards, manage energy resources, respond to climate change and its impacts, safeguard water resources, improve weather forecasts, manage ecosystems, promote sustainable agriculture and conserve biodiversity. GEOSS coordinates a multitude of complex and interrelated issues simultaneously. This cross-cutting approach avoids unnecessary duplication, encourages synergies between systems and ensures substantial economic, societal and environmental benefits.

At the GEO Ministerial Summit in November 2010, the GEO Data Sharing Action Plan was adopted. The Data Sharing Action Plan builds upon the concept of full and open exchange and on the Implementation Guidelines accepted by the Plenary which state that data, metadata and products made available through the GEOSS are made accessible with minimal time delay and with as few restrictions as possible on a non discriminatory basis, at minimal cost for no more than the cost of reproduction and distribution.

For the most current information on GEOSS visit the GEO website at: www.earthobservations.org

The Challenge 3.1.1

Relationships between health and environment are a matter of concern all over the world, as changes in the natural environment can compromise human health. Climate change and extreme weather events are associated with a wide range of health risks. Exposure to persistent organic pollutants, mercury, and other environmental contaminants has the potential to harm human health. Changes in biodiversity can affect the transmission of diseases.GEOSS addresses these entire issues trough the development of observation systems that contribute to improving our understanding of how the environment affects human health and well-being.

Key variables include airborne, marine, and water pollutants; stratospheric ozone depletion; land-use change; persistent organic pollutants; food security and nutrition; noise levels; weather-related stresses and disease vectors; and many others.

3.1.2 Europe's Capability to Respond

Contributions of health datasets are expected to increase with emerging technology (sensors, computing) which will allow more accurate and near real-time autonomous monitoring / products of health indicators with increasing complexity, e.g. in situ genetic analyses. Like its predecessor GMES, COPERNICUS provides contribution to this SBA through its Atmosphere, Marine and Land Core Services, and several research projects are contributing to their development.

3.1.3 The GEO Health Societal Benefit Area in Context

The 10-year implementation plan for GEOSS outlines health issues with Earth-observation needs: airborne, marine, and water pollution; stratospheric ozone depletion; persistent organic pollutants; nutrition; and monitoring weather-related disease vectors.

GEOSS will improve the flow of appropriate environmental data and health statistics to the health community, promoting a focus on prevention and contributing to continued improvements in human health worldwide. Among health-related potential products one can list:

- Early warning system for vector-borne diseases including malaria, dengue fever, Rift Valley fever
- Early warning system for water-borne diseases and risk including cholera and leptospirosis, and algae bloom risk

- Early warning system for airborne diseases
  including meningitis and impacts of air quality and sand & dust storms
- Early warning system for influenza
- Tracking systems for pollutants
  including Earth monitoring systems for mercury and persistent organic pollutants
- Disease transmission dynamics – linkages with other Societal Benefit Areas
  e.g. disasters and vulnerable areas for vector and waterborne diseases; biodiversity, ecosystems and vector-borne diseases

Target 2015 3.1.4

The GEOSS Strategic Target for Health aims to substantially expand the availability, use, and application of environmental information for public health decision-making in areas of health that include allergens, toxins, infectious diseases, food-borne diseases, and chronic diseases, particularly with regard to the impact of climate and ecosystem changes.

This will be achieved through working with the World Health Organization (WHO) and the global community of human health and environment experts in order to develop and implement health and environment projects, which will:

- advance the application of observation, monitoring and forecasting systems to health decision-making processes;
- foster the use of established and emerging observation systems in operational health related applications for air and water quality, infectious diseases, and vector-borne diseases, and develop associated products such as forecasts and alerts compliant with the Common Alerting Protocol (CAP);
- include efforts to examine terrestrial, freshwater, and marine (ocean) ecosystems and their services, to establish causality between changes in flora, fauna and other factors affecting the emergence and transmission of disease;
- document links between water and communicable diseases, as part of the life cycle of vectors or as a medium infecting populations;
- facilitate the integration of Earth science databases and emerging information products with public health data, socioeconomic data, and epidemiological information needed in decision support systems for health care planning and delivery.

This will be complemented by the development of a global network of scientists, researchers, practitioners and other operational end users which will:

- provide free access to an expanded inventory of available Earth observation data, metadata and products applicable to public health;
- provide input relating to the technical specification of new major environmental observation capabilities, including in-situ and remotely sensed observations that will allow historical data analysis and early detection of changes that influence health;
- facilitate Earth observation training and capacity building for future scientists,
- researchers, public health policy makers and practitioners, and end users, including contributions of best practices in this domain to the GEOSS Common Infrastruture (GCI) best practices registry.

The achievements of these actions will be demonstrated by access to improved environmental information and tools to support the global community of human health and environment experts. This is expected to lead to a an increased use of environmental information and tools to support decision making in epidemics and/or disease management and planning for well-being. The concept of applying outcomes from other Societal Benefit Areas will also help to overcome the boundaries which currently exist between separate application and research domains (**GEOSS for Health, The GEO Health Societal Benefit Area**).

3.1.5 The GEO 2012-2015 Workplan

The GEO 2012-2015 Work Plan has two main tasks in the health area:

- HE-01 Tools and Information for Health Decision Making
- HE-02 Tracking Pollutants

The task HE-01 is coordinated by the World Health Organisation (WHO) and comprises four so-called components:

- HE-01-C1: Air-borne Diseases, Air Quality and Aeroallergens
- HE-01-C2: Water-borne Diseases, Water Quality and Risk
- HE-01-C3: Vector-borne Diseases
- HE-01-C4: A Holistic Approach to Health: Transmission Dynamics, Urban Health Forecasting, Linkages and New Technologies

The objective of HE-01 is:

> **"Develop tools and information systems for the environment and human health. Advance the integration of Earth observations and forecasts into health decision-making processes. Engage with health users and decision-makers to identify needs. Carry out capacity building and a plan for the promotion and sustainable use of Earth information by the health user-community. Establish linkages with other Societal Benefit Areas such as Ecosystems, Biodiversity, Climate and Disasters (e.g. in connection with floods, earthquakes, volcanic eruptions, cyclones, and tsunami events)."**

In addition, the following complementary GEO tasks are also highly relevant to work in the health SBA:

- IN-05: GEOSS Design and Interoperability with one component of the same name. The Architecture Implementation Pilots (AIP; see below) are important activities in this task.
- ID-02: Developing Institutional and Individual Capacity

EO2HEAVEN CONTRIBUTION TO GEOSS 3.2

The main engagement of EO2HEAVEN was facilitated through active participation in the GEOSS Architecture Implementation Pilot (AIP) and Community od Practice (CoP) activities. EO2HEAVEN has contributed results to HE-01-C1 Air-borne Diseases, Air Quality and Aeroallergens and HE-01-C2 Water-borne Diseases, Water Quality and Risk from its respective case studies. EO2HEAVEN placed a special focus on community and capacity building as a contribution to Task ID-02 (cf section 3.2.2).

The GEOSS Architecture Implementation Pilot 3.2.1

A key contribution was made by EO2HEAVEN to the GEOSS Architecture Implementation Pilot (AIP) phases 3 to 5 of GEO Task AR-09-01b in the GEO 2009-2011 Work Plan and GEO Task IN-05 in the subsequent GEO 2012-2015 Work Plan.

The GEOSS Architecture Implementation Pilot (AIP) develops and deploys new process and infrastructure components for the GEOSS Common Infrastructure (GCI) and the broader GEOSS architecture.

EO2HEAVEN started its involvement with the AIP development at the very early stages of the project with AIP phase 3. The primary contribution focussed on the participation of our case study experts in the Societal Benefit Area Health Thread, discussing our project approach with expert teams who participated in the previous GEOSS AIPs to identify the lines of collaboration. This led to a generalization and extension of our three case studies to the health Societal Benefit Areas identified by the Pilot, i.e. Air Quality and Early Warning of Malaria scenarios. With this response EO2HEAVEN primarily strengthened its visibility within the GEOSS community and paved the way for a continous working relationsship throught the lifetime of EO2HEAVEN and beyond.

In AIP phase four, EO2HEAVEN took a leading role in activity #1, 'Access to Priority EO Data Sources'. The goal of this activity was to bring datasets online, which are listed as priority EO data sources in the Critical Earth Observation Priorities report published by GEO task US-09-01a. Various existing lists of datasets and corresponding services have been merged and were processed by the AIP-4 consortium. EO2HEAVEN also supported the development of a number of tutorials as well as open source reference implementations, which intend to help both data providers as well as GEOSS users to work with GEOSS. The tutorials were developed in close cooperation with the Standards and Interoperability Forum (SIF), which has been established "to facilitate the interchange of information and the development of recommendations for standards and interoperability in GEOSS" (IEEE 2011).

In AIP phase five, EO2HEAVEN introduced three different scenarios that have been implemented to test the current GEOSS architecture, services, information models and knowledge representations in the context of health related scenarios.

3.2.2 The GEOSS Community of Practise

The GEO Health and Environment Community of Practice (CoP) is an informal group of international experts in the health and environment field. It meets annually to exchange experience on investigating relationships between health and the environment and approaches to mitigating the burden on health. This includes assessment of relevant health and environment data, model- and data-driven analysis, application of GEOSS components and proposals to GEOSS for future work. The CoP members report on on-going activities and results of various projects and initiatives around the world. EO2HEAVEN has participated in the GEO Health and Environment CoP since 2011 and hosted the 2012 meeting in Karlsruhe. The GEO Health and Environment CoP was instrumental in formulating the current GEO Work Plan 2012-2015 in the health tasks HE01 and HE-02. Its members maintain the online GEO component sheets reporting on progress towards the task objectives. The CoP is also a forum for initiating collaborative proposals to funding agencies, e.g. on early warning systems.

The Health and Environment CoP liaises closely with the GEO Integrated Global Water Cycle Observations CoP since water and sanitation (WATSAN) are critical to health. The operation of both CoPs is supported by the GEO Secretariat.

Conclusions and Future Work Items 3.2.3

Based on the experiences made in the three EO2HEAVEN case studies, a number of isues and challenges emerged that are also relevant to the SBA Health in general:

- Data availability is often an issue with regard to a matching temporal and spatial resolution.
- Data discovery and accessibility is an issue in particular where health data is concerned. There is a lack of standardized interfaces for online exchange of data and information.
- Spatial variability in urban environments requires extremely high spatial data resolution.
- Satellite data to ground level conversion/mapping is very challenging. Though this is simple in theory, specific situations such as areas with with high cloud coverage, or regions with unsafe travelling conditions etc. can pose a major obstacle in validating satellite data. The lack of in-situ observation stations is general an issue, as satellite data calibration and evaluation depends on the availability and quality of reference data.
- Data cleaning tasks remain very time consuming: Earth observation as well as health data is often provided in formats and maturity levels that require intensive manual data cleaning, gap filling, interpolations, error handling etc.
- The automated mapping of data-to-tools remains a challenge: Existing Web services often make use of complex hierarchical, often XML-based data structures, whereas the majority of statistical and AI tools requires flat data structures.
- The usability of health data suffers from data privacy policies and regulations, which often make it impossible to access the raw data directly. The required data aggregation often adds artificial uncertainty to the data.

Despite the above mentioned items, which required future attention, the usability of GEOSS has been succesfully tested and confirmed to be functional. The general architecture with distributed services and data supports all requirements set by SBA Health. Only more sophisticated security has been missed in some cases. In particular the applied OGC Web service interfaces work well with both EO and health data.

One surprising observation was, that GEOSS is often unknown to many domain researchers. The activities that EO2HEAVEN carried out in the Health thread provided valuable input to the project and helped to link ongoing research and developments from various areas of expertise. Hence the promotion of GEOSS beyond the Earth observation community should be fostered.

3.3 INSPIRE

3.3.1 The INSPIRE Directive

INSPIRE stands for “Infrastructure for Spatial Information in the European Community” and is the first European Directive that is dedicated to the interoperability for geospatial information and the implementing of a European Spatial Data Infrastructure. The INSPIRE Directive 2007/2/EG came into force on 15 May 2007 and the full implementation is required in 2019 (THE EUROPEAN PARLIAMENT AND COUNCIL OF THE EUROPEAN UNION, 2007). following a step-by-step approach (INSPIRE, 2010a). The INSPIRE Directive covers five main aspects:

1) Interoperability of Spatial Datasets and Services:
   In its Annexes the Directive lists the 34 data topics for which digital public administration data shall be made available in the final infrastructure. The directive mandates provisions to allow for an as far as possible seamless and uncomplicated integration of datasets covering these topics and stemming from different European sources.
2) Metadata:
   To allow the discovery and evaluation of INSPIRE relevant datasets and services in Europe, the INSPIRE directive requires the provision of Metadata about these resources.
3) Network Services:
   The directive mandates the provision of Network Services, to make it possible to discover, transform, view and download spatial data and to invoke spatial data and e-commerce services.
4) Data Sharing:
   Data policy provisions are mandated to allow an as easy as possible data exchange between public bodies and to allow third parties, especially citizens to have an as far as possible free and easy access to spatial information covered by INSPIRE.

5) Coordination and Complementary Measures:
This part of the directive covers the organisational and management aspects of the INSPIRE implementation. Especially measures to allow the monitoring of the INSPIRE implementation process are required.

Providing INSPIRE-compliant data-sets is mandatory for all national public authorities. Moreover the Directive includes a mandate for the European Commission to implement an INSPIRE geo-portal. The European geo-portal shall give access to the Member States' implementation of the INSPIRE Network Services.

Further to the INSPIRE Directive the INSPIRE Implementing Rules are currently being developed to define and govern the technical details of the INSPIRE implementation. These Implementing Rules also become part of the European legislation. The normative Implementing Rules are seconded by informative INSPIRE Guidance Documents (which are not legally binding) to specify detailed technical aspects and thus to allow for shorter updating cycles for this guidance. The official INSPIRE web site (http://inspire.jrc.ec.europa.eu) is an excellent source to find further details on INSPIRE.

INSPIRE Human Health and Safety Data Specification 3.3.2

One of the 34 data themes named in the annex of the INSPIRE Directive is Human Health and Safety. This data theme is defined as:

> "Geographical distribution of dominance of pathologies (allergies, cancers, respiratory diseases, etc.), information indicating the effect on health (biomarkers, decline of fertility, epidemics) or well-being of humans (fatigue, stress, etc.) linked directly (air pollution, chemicals, depletion of the ozone layer, noise, etc.) or indirectly (food, genetically modified organisms, etc.) to the quality of the environment."
> INSPIRE Directive 2007/2/EG

Further details, as well as the data specification, are published by the INSPIRE Thematic Working Group Human Health and Safety in document D2.8.III.5 Data Specification on Human health and safety – Draft Technical Guidelines. This document is available online at: http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_HH_v3.0rc3.pdf.

However, applying these guidelines and using the data models operationally is still work in progress and mostly future work. Here EO2HEAVEN acts as an early adaptor and follows the INSPIRE guidelines to the extent feasible and helps in gaining first experiences in their applicability.

3.4 COPERNICUS

Copernicus, formerly known as GMES, is a European system for monitoring the Earth. It consists of a complex set of systems which collect data from multiple sources: Earth observation satellites and in situ sensors such as ground stations, airborne and sea-borne sensors. It processes these data and provides users with reliable and up-to-date information through a set of services related to environmental and security issues.

The services address six thematic areas:

- Land
- Marine
- Atmosphere
- Climate Change
- Emergency Management
- Security

They support a wide range of applications, including environment protection, management of urban areas, regional and local planning, agriculture, forestry, fisheries, health, transport, climate change, sustainable development, civil protection and tourism.

The main users of Copernicus services are policymakers and public authorities who need the information to develop environmental legislation and policies or to take critical decisions in the event of an emergency, such as a natural disaster or a humanitarian crisis. As such, Copernicus primarily supports reserachers with information to prepare environmental studies, e.g. for guidance in legislation and policies development (with a particular focus on Climate Change) and as well as monitoring of their implementation and assessment of effects.

Based on the Copernicus services, many other value-added services can be tailored to more specific public or commercial needs.

The Copernicus programme is coordinated and managed by the European Commission. The development of the observation infrastructure is performed under the aegis of the European Space Agency for the space component and of the European Environment Agency and the Member States for the in-situ component.

Further information on Copernicus is available at: www.copernicus.eu

The EO2HEAVEN Approach 4

SCENARIO BASED ENGINEERING 4.1

Developing useful software is a constant challenge, because the two main actors: the users and the developers. Both tend to live and think in completely separate worlds. Hence, bringing both together in one room to plainly discuss what this software should be able to do – and what not – is a common point of failure in software development.

Users tend to have a picture of software in mind, which is based on their day to day business and workflow and simply solves their problems by clicking one button. IT people, on the other hand, are usually in love with their technology and have a clear picture how this technology has to be used and how the software could look like. In most cases, both pictures differ considerably, because neither party knows, let alone understands, the others motives.

Crucial as it is that both sides develop a common understanding and share a common language, a way of mediation is required to come up with a result that makes both sides happy. To provide a common ground where both can start to understand each other, an often applied approach in software engineering is to describe typical use case examples and from there derive the related user requirements. An important aspect in doing so, is to take different viewpoints, not only from the engineering side, but also accommodating the different backgrounds of the involved parties. Let us use the following example to illustrate the complexity:

A health practitioner is asking for a map of the number of asthma cases versus air quality that were reported last month in a given district, he might think of a simple solution, where he enters the time period in question, clicks one button and gets a simple map with a chart on top. The scientist from the environmental domain, however, has a vast fault of measurement data - so there should be option buttons to also display those. Hand this to the software developer and he will undoubtably have various ideas on the background map alone: using a satellite backdrop from Google or Bing would be nice add-ons, so let's start there...

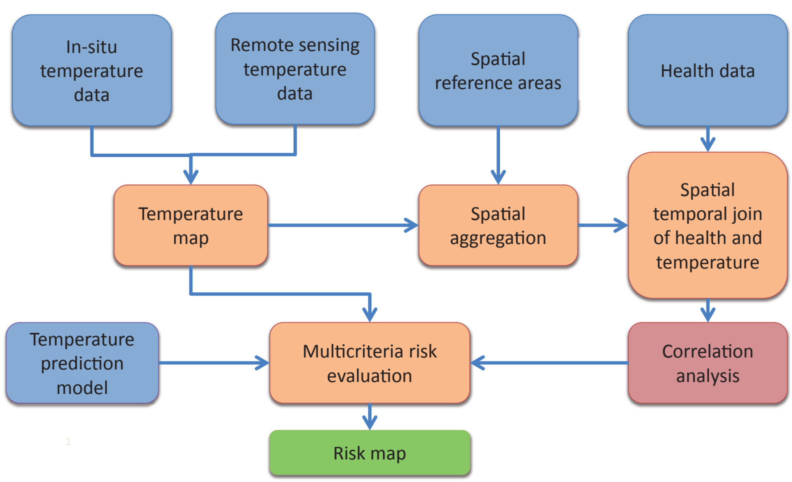
It is easy to see, that the poor health practitioner of our example will not end up with what he had in mind, but will rather have to face an endless array of options.

This may sound far fetched, but actually is one of the challenges EO2HEAVEN dealt with: educating the developers and the two user groups from the health domain and the environmental domain to speak a common language. Accordingly the first rule is to keep it simple and start with small steps in an iterative approach. In the first iteration of the project the software developers and architects defined a rigorous and much formalized approach based on a common template:

Use Case Name		**EO Data Access**
Use Case ID		***CS3-UC03***
Revision		CS3-UC03-01
Status		Active
Goal		Access data selected in UC2
Summary		The user is able to access the data, or a portion thereof, by directly downloading it, interacting with the data via a service or providing pointers to the data for other services to interact with
Category		Primary
Actor		Researchers
Primary Actor		Cholera researchers
Stakeholder		Researchers and research organizations
Requested Information Resources	Data input	EO Datasets
Requested Information Resources	Data access control	Likely to be as determined by data provider
Requested Information Resources	Data format	• Scientific data formats (Geospatial; hierarchical) e.g. GeoTiff, HDF, spatial features • Network compatible data formats • Text files (e.g. CSV files) • Tables and lists (e.g. spreadsheets)
Preconditions		The user has chosen one or more datasets for downloading or linking to a service
Triggers		The user executes a data download or parameterisation of services
Main success scenario		• The user confirms the datasets for access • The user accepts any terms and conditions on the use of the data • The user selects a download location or parameterises a service endpoint • The user may request a subset of the data on some dimension • The user executes the data access
Extensions		• The user accepts terms and conditions of data use, if necessary • The user specifies a download location or, • The user specifies a service which must access the selected data and perform some process on it • The user specifies a subset of the data along some dimension e.g. spatial subset, temporal subset, thematic subset
Alternative paths		Datasets may not be directly downloadable or linkable – may have to engage in further requests to individuals or organizations
Post conditions		The user is in possession of required data or has linked required data to online services that may facilitate the next steps (e.g. a Web Processing Service
Notes		Each dataset may provide a different mechanism for downloading, e.g. MODIS L2 data requires a http download from a directory tree; others may be found at OpeNDAP services, OGC W*S services
Author and date		CSIR, 2010-05-12

The formal purpose is to achieve harmonization and traceability – which is good for the subsequent evaluation of the developments. Fine granular use cases and requirements were requested from both user groups. Though they were not familiar with this rather abstract concept at all, it provided a base for a first common understanding and allowed the IT developers to get started on the basic building blocks for each scenario. This initial iteration basically helped to bridge the existing expertise boundaries and build a better understanding of each others requirements and capabilities.

However, to understand process requirements in the context of their actual workflow, the initial approach is far to abstract. In the second iteration the focus therefore moved from functional, fine granular use cases requirements to the development of selected scenarios, which represent one complete sequence or method to reach a concrete aim in the project, for example "display a map via a website indicating the risk of suffering from cardiovascular disease due to heat". The detailed description of scenarios includes the specific workflows and processing steps as well as all necessary input data, the processing steps and the expected overall result. In most cases the processing steps are similar to the more abstract use cases from the first iteration, but put the functional requirements in a more user oriented context. To achieve this, additional end user interviews were carried out, so that this modified approach was based mostly on natural language descriptions and basic diagrams:



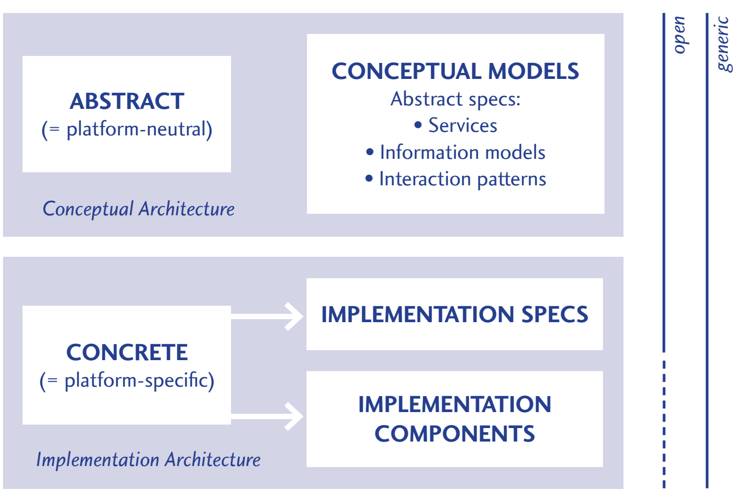
A major benefit of this more holistic approach is that it also helps to improve the understanding of the tasks and challenges among all project partners and across the boundaries of their respective area of expertise. The results of the second implementation iteration were now focussed on the coverage of process workflows, rather than the initial development of tools and building blocks. At the end of this process the pilot implementation were presented in stakeholder workshops and discussed with end users, to collect their feedback and ideas for improvement the final iteration phase.

4.2 THE SPATIAL INFORMATION INFRASTRUCTURE

The specification of the implementation architecture of the EO2HEAVEN Spatial Information Infrastructure (SII) is the basis and starting point for all EO2HEAVEN developments.

The objective of the SII implementation architecture is to motivate and specify the basic design decisions derived from user requirements and generic architectural principles. Its focus is on a platform-neutral specification, i.e. it provides the basic concepts and their relationships as conceptual models and abstract specifications.

By abstract it is meant that the specification is independent of the specifics of a particular service platform. Such an abstract specification comprises service specifications, information models and interaction patterns between the major architectural components, as illustrated below:



The structure of the SII architecture is aligned with an interpretation of the five viewpoints of the ISO Reference Model for Open Distributed Processing (RM-ODP) (ISO 10746-1, 1998). The RM-ODP explicitly foresees an engineering step that maps solution types, such as information models, services and interfaces specified in information and service viewpoints, respectively, to distributed system technologies. We describe this mapping step in terms of engineering policies. These policies constitute architectural blueprints that enable a system engineer to specify implementation architectures according to given user requirements, as outlined in the lower part of this graphic.

These viewpoints were partly combined and specified in the six document parts of the SII architectural specification. As such, it continues the series of architecture specifications of previous European FP6 projects, which resulted in the following OGC documents:

- The Reference Model for the ORCHESTRA Architecture (RM-OA), was developed by the ORCHESTRA project and approved as an OGC best-practices document (OGC 07-097).
- The RM-OA was extended by the SANY project in its Sensor Service Architecture (SensorSA), which is available 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 inherited and extended this work in form of the present SII implementation architecture specification. Particular emphasis is put on the support of distributed environmental and health monitoring. It also addresses the consequences of special health privacy and security requirements on the architecture, recognizing the fact that some of these requirements can and will only be tackled on organisational level, e.g. by aggregating and anonyffmizing data in non distributed, offline computer centre environments, before this data can be offered to other applications via distributed online accessible service infrastructures.

The architectural work in EO2HEAVEN also followed the overall iterative approach. To ensure a coherent methodological approach across all development cycles, each cycle resulted in a revised version of the SII implementation architecture.

The EO2HEAVEN architecture incorporates 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. e.g. 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 EO2HEAVEN SII architecture documentation takes into account emerging technologies, especially for the Sensor Web Enablement (SWE) architecture and the geo-processing architecture and provides extensions and refinements, e.g. for the requirement to share and process huge amounts of datasets provided by Earth observation agencies and health institutions in order to investigate and assess correlated risks.

An important project aspect was to improve the applicability of the Sensor Web technology to the case studies of EO2HEAVEN and to facilitate the practical use of the OGC Sensor Web Enablement (SWE) framework. This included the definition of a lightweight SWE profile, an analysis of and contribution to the specification of the Sensor Observation Service (SOS) 2.0 as well as an approach how the data used within EO2HEAVEN could be integrated more easily into Sensor Observation Service instances.

A second focus addressed the handling of different types of Earth Observation (EO) data, i.e. remote and in-situ data, within the SWE framework. Whereas the application of the SWE technology to in-situ data is quite common, there is a lack of best practice guidance how remote sensing data can be handled. Accordingly, EO2HEAVEN provided a list of recommendations for developing such a best practice approach. This also includes potential extensions of the Sensor Model Language (SensorML) for remote sensing data and the linkage of SWE with the GEONETCast technology, a near real time, global network of satellite-based data dissemination systems designed to distribute space-based, air-borne and in-situ data, metadata and products to diverse communities.

Further major topics of the SII are processing and fusion services, processing of quality information, handling of data uncertainty, its encoding and visualization, etc.

4.2.1 Design Principles

A service oriented architecture for an Spatial Information Infrastructure in the environment and health context cannot solely rely on existing design principles that are typically applied in commercial SOA environments (Erl, 2008). EO2HEAVEN therefore followed the refined architecture approach of SANY:

- **Rigorous Definition and Use of Concepts and Standards**
  The SII should make rigorous use of proven concepts and standards in order to decrease dependence on vendor-specific solutions. This helps to ensure the openness of an information network and support the evolutionary development process.
- **Loosely Coupled Components**
  The SII should allow the components involved in a service network to be loosely coupled, in which case loose coupling implies the use of mediation to permit existing components to be interconnected without changes.
- **Technology Independence**
  The SII should be independent of technologies, their cycles and their changes, as far as practically feasible. Accordingly, it is possible to accommodate changes in technology (e.g. lifecycle of middleware technology) without changing the architecture itself. This also implies independence of specific implementation technologies (e.g. middleware, programming language, operating system).
- **Evolutionary Development – Design for Change**
  The SII should be designed to evolve, i.e. it shall be possible to develop and deploy the system in an evolutionary way. Hence it is able to cope with changes of user requirements, system requirements, organisational structures, information flows and information types in the source systems.
- **Component Architecture Independence**
  The SII should be designed in a way that service network and source systems (i.e. existing information systems, data sources and sensor networks) are architecturally decoupled. The architecture shall not impose any architectural patterns on source systems for the purpose of having them collaborate in a service network, and no source system shall impose architectural patterns on a SII. Here it is important to point out, that a source system is seen as a black box, i.e. no assumptions about its inner structure are made when designing a service network.
- **Generic Infrastructure**
  The SII services should be independent of the application domain, i.e. they can be used across different thematic domains and in different organisational contexts. Ideally, any update of integrated components (e.g. data sources, applications, systems, ontologies) requires no or only little changes to the users of the SII services.

4.2.2 Reference Model for Open Distributed Processing

As previously mentioned, the conceptual foundation for the SII has been the Reference Model for the ORCHESTRA Architecture (RM-OA).

The RM-OA provides a platform-neutral abstract specification of a geospatial service-oriented architecture that responds to the requirements of environmental risk management applications. It comprises generic architecture services and information models based on and extending existing OGC specifications.

The design of the EO2HEAVEN architecture follows the guidelines and viewpoints of the **ISO Reference Model for Open Distributed Processing** (ISO/IEC 10746-1:1998). However, since the requirements from the EO2HEAVEN scenarios show the characteristic of a loosely-coupled network of systems and services instead of a 'distributed processing system based on interacting objects' as presumed by RM-ODP, the RM-ODP concept is not followed literally. The RM-ODP viewpoints are applied on a higher scale to the structuring of ideas and the documentation of the architecture itself, and on a small scale to the description of the data sources and models:

- The **Enterprise Viewpoint** reflects the analysis phase in terms of the business contexts, related system and the user requirements expressed in use cases as well as the assessment of the current technological foundation for the architecture. It includes rules that govern actors and groups of actors, and their roles.
- The **Information Viewpoint** specifies the modelling approach of all categories of information, with which the architecture deals, including their thematic, spatial, and temporal characteristics, as well as their meta-information.
- The **Service Viewpoint** specifies the interface and service types that aim at improving the syntactic and semantic interoperability between services, source systems and environmental applications.
- The **Technology Viewpoint** specifies the technological choices for the service platform, its characteristics and its operational issues, e.g. the specification of the platform 'Web Services' including a profile of the Sensor Model Language, or the physical characteristics of data sources, sensors and sensor networks.
- Finally, the **Engineering Viewpoint** specifies the mapping of the service specifications and information models to the chosen service platform,

considers the characteristics and principles for service networks, e.g. synchronous or asynchronous interaction patterns, and defines engineering policies, e.g. about access control and resource discovery.

DATA WE USED 4.3

Remote Earth Observation 4.3.1

Earth observations include measurements and monitoring of the Earth, be it under water, on the land surface or beneath, of air and water quality, or of atmospheric conditions. Earth observations are based on the measurement of physical, chemical and biological parameters not only globally, but also locally almost anywhere on the globe.

With the help of Earth observation satellites a massive quantity of diverse information about our planet is continuously being gathered. Various recording instruments measure and photograph the Earth, transmitting data on many different environmental parameters, such as atmospheric particulates. reactive gases and marine pollution, in real time. More information on using Earth Observation Data is provided in chapter 10.

In-situ and Sensor Observations 4.3.2

In contrast to remote sensing, in-situ measurements are taken on site at a specific point in time and space. Thus, they describe the concentration of a certain pollutant in the direct surrounding of the sensor. In comparison to remote sensing data, the advantages are:

- Higher accuracy because of less external (e.g. atmospheric) influences on the measurement
- Reduced time delay between sensor measurement and data provision
- Provision of continuous timeseries for specific locations

However, a considerable drawback is the need for interpolation techniques to obtain a continuous spatial coverage of air quality information. The accuracy of those models depends essentially on the number and distribution of available in-situ sensors as well as bounding conditions, such as the meteorological situation and chemical dispersion of the considerred pollutant.

In many countries, in-situ air quality measurements are required by legislation and are frequently used for air pollution control and prevention measures. In Europe the EU-Directive 2008/50/EC requires the member states to implement a network of measurement sampling points for air quality, covering all substances in ambient air that are likely to have a negative impact on the human health. In the context of EO2HEAVEN, Particulate Matter (PM10/PM2.5), Ozone (O_3), Nitrogen Dioxide (NO_2) and Sulphur Dioxide (SO_2) are considered as most relevant. Usually, in-situ data is continuously measured and provided in short term intervals (e.g. 10min). Therefore, they are ideally suited to detect short term changes in ambient air quality.

One important aspect to be considered with some types of in-situ measurements is, that they strongly influenced by their original specific purpose and hence their particular location. If, for example, air quality measurement are taken to identify the impact of traffic on a main road or industrial emission control in the downwind area, the observed values will not necessarily represent trends or larger areas.

4.3.3 Ex-situ and Laboratory Observations

Whilst many observations on environmental parameters can be collected in or near real-time with sensors, which monitor a particular location, EO2HEAVEN also had to incorporate data that resulted from health statistics, patient records or the analysis of field samples in laboratories. An example for the latter are water samples that where taken for cholera detection.

To integrate such data into the overall EO2HEAVEN architecture concept, these data source were treated as virtual sensors. Conceptually, there is no difference whether an observation is made by a physical sensor device, reporting for example a temperature value for a given location, or a laboratory result, which confirms the presence of *Vibrio cholerae* in a water sample. In both cases, we have an observation value, a location to which it relates and meta-information on the observing entity.

Following this approach allows us to turn data which might only exist on paper records into discoverable data source in the Spatial Information Infrastructure.

4.3.4 Health Data

The scope of EO2HEAVEN was to use health data, which is already available and did not need to be newly gathered or collected. In general, such health data may be extracted from a variety of data sources:

- Population statistics contain data on birth and death which allow extracting temporal and spatial information concerning the events. Furthermore causes of death and additional information are recorded (e.g. age at death).
- Health insurances have vast datasets concerning their payments to healthcare providers and insured persons. In principle this contains detailed information for whom (patient) what (medication, laboratory, physical

therapy, etc.) was paid why (diagnosis) when (temporal information) and to whom (spatial information).

- Health data collected to address specific scientific questions on the effects of environment on public health provides useful information to quantify and calibrate environmental health risk models. If available, data obtained from clinical trials or community based studies provides a strong evidence base to quantify health hazard resulting from environmental exposures.
- Health data routinely collected from national disease programs or national health surveillance managed in Health Management Information Systems (HMIS) provide a vast body of longitudinal data monitoring disease burden at sub national scales. Although such observational data records might be subjected to error, the general availability of large quantities of health data over extended time periods make these valuable information sources to assess environmental health impacts.

Whilst health data for individual persons are sensitive and protected, general health data are important information concerning public health and public services (e.g. emergency departments) and economic planning for the health sector. According to applicable data protection laws, such as the professional code for physicians, state and federal German law, as well as European directives, such sensitive data may either be used in an anonymised manner or with the written consent of the person concerned. As it is very difficult to obtain written consent for the whole population of a larger city or a region or state, such datasets can only be used provided that anonymisation is warranted in a way that re-identification of the individuals is impossible with justifiable effort.

Beyond population statistics and health insurance data, a variety of other organizations own statistics, that might hold the clue for a more complete picture on a given research topic:

- Sales statistics of pharmaceutical manufacturers
- Sales statistics of pharmacies
- Hospital statistics
- Import/export statistics
- Health questionnaires and surveys
- Employee data of employers
- Public health statistics
- Emergency service statistics

Depending on your research requirements, such data might be additionally used to analyse health environmental relationships.

4.4 METHODS WE APPLIED

New developments in epidemiological modeling rely on environmental datasets such as land surface temperature, topography, urban density, vegetation distribution and maps of water bodies. This information can be used to build early warning systems to predict outbreaks of diseases, such as malaria, and also to investigate the unknown environmental factors, which determine the spread of less well understood diseases. This type of information can also be used to help target eradication programs for disease vectors.

But before introducing the methods and concepts that were developed and applied in EO2HEAVEN, let us first take a look the general understanding of the frequently used term health risk:

4.4.1 Definition of Health Risks

The relationship between environment and health is a complex non-linear relationship with many key factors which are sensitive to human interventions.

Environmental health is defined by the World Health Organization as those aspects of the human health and disease that are determined by factors in the environment.

It also refers to the theory and practice of assessing and controlling factors in the environment that can potentially affect health. Environmental health as used by the WHO Regional Office for Europe, includes both the direct pathological effects of chemicals, radiation and some biological agents, and the effects (often indirect) on health and well being of the broad physical, psychological, social and cultural environment, which includes housing, urban development, land use and transport. (Novick, Robert (ed.), 1999).

A Definition of Risk

Risk is the probability that a risk event, such as disease resulting from exposure to a pollutant, will occur during or over a specified time period. This probability is basically defined by the individual's vulnerability and the exposure to particular pollutants. The level of risk can be viewed as a function of probability and severity of impact. Variables associated with an increased or decreased risk are the risk factors or protective factors, respectively. The risk factors and protective factors are any characteristics, such as personal behaviour, inheritance, or environmental conditions that are considered to be associated with occurrence of the risk event. "Health risk" is defined as "a factor that raises the probability of adverse health outcomes" (WHO, 2009)

Risk Assessment

Risk assessment has been defined as the characterization of the potential adverse health effects of human exposures to environmental hazards. The process of risk assessment can be split into four steps (Gordis, 2009):

- Hazard identification: determination of whether a particular pollutant is causally linked to particular health effects.
- Dose-response assessment: determination of the relationship between magnitude of exposure and the probability of occurrence of the health effects in question.
- Exposure assessment: determination of the extent of human exposure before or after application of regulatory controls.
- Risk characterization: description of the nature and often the magnitude – of human risk, including attendant uncertainty.

Vulnerability

Vulnerability is the degree to which individuals and systems are susceptible to, or unable to cope with, adverse effects of climate variability and extremes. Vulnerability is a function of the character, magnitude and rate of climate variation to which a system or individual is exposed, its sensitivity and its adaptive capacity (McCarthy et al., 2001). Individual vulnerability is largely determined by genetic factors, which are very complex by nature. The effects of exposure to a certain pollutant are related to its toxic effects, intensity of exposure and individual vulnerability (Kirch, 2008).

Data Discovery 4.4.2

To be able to analyse health risks and their links to environmental factors, experts rely on suitable health data as well as environmental and geospatial data. To facilitate the discovery of according datasets, EO2HEAVEN also addressed the provision of metadata, that is necessary to discover which relevant information sources might be available for a given research aspect.

To enable the discovery of datasets, it is essential to provide sufficient descriptions that can be used by catalogues and registries. For geospatial data it is possible to rely on existing standard such as ISO 19115 “Geographic Information – Metadata” and ISO 19139 “Geographic Information – Metadata XML Schema Implementation”.

Since a dataset which contains observations with a time and space reference can conceptually be seen as a sensor, EO2HEAVEN evaluated options to apply the existing Sensor Web Enablement (SWE) standards of the Open Geospatial Consortium. The SWE framework SensorML contains the encoding for the provision of sensor metadata. However, since SensorML has been designed to support a very broad range of use cases and requirements, it offers the user much flexibility regarding the content and structure of SensorML documents. In order to support reliable sensor discovery it is therefore necessary to define a meaningful minimum set of metadata.

To support automatic harvesting and processing of sensor metadata it is furthermore necessary to define the structure in which these metadata elements have to be encoded. This has been addressed within EO2HEAVEN through the definition of a SensorML Profile for Discovery.

Whilst the described approach works well for most geospatial and environmental datasets, it is a completely different story for health data. Though conceptually the same principles can be applied, data privacy and protection issues are a considerable hurdle in the process of data discovery. For understandable reasons, the conditions of data storage and its usage are handled very restrictive. As an example, the health data required for the EO2HEAVEN scenario located in Germany must not be accessible via the internet and significant efforts for anonymization of patient-specific data was required, before the data could be used beyond the boundaries of the data owning organisations. This significantly reduced the quality of results in terms of informative value and granularity.

4.4.3 Data Fusion

Data fusion generally refers to activities that determine the existence of a link, and its relative importance, between environmental factors and disease outbreaks and outbreak patterns over time and space. This is linked to the definition of data fusion presented by the Joint Directors of Laboratories (JDL), who formed the Data Fusion Subpanel, that later became known as the Data Fusion Group:

> **"A process dealing with the association, correlation, and combination of data and information from single and multiple sources to achieve:**
> - **Refined position and identify estimates, and**
> - **Complete and timely assessments of situations and their significance".**

Within EO2HEAVEN data fusion includes:

- creating associations between environmental and health data to identify potential correlations
- linking different data sources for validation
- linking different data sources to enrich information on a specific topic.

One challenge for EO2HEAVEN in preparation for health-risk analysis is the provision of environmental data and derived air quality information as coverage for the whole area of interest. Available in-situ observations and remote sensing data need to be processed, validated and prepared for health data fusion. This includes appropriate methods for spatial and temporal interpolation, data aggregation as well as pollution dispersion modelling to get most reliable air quality information for the further health-risk analysis. In addition, data fusion methods need to be applied in order to map and transform different environmental observations on the spatial reference (e.g. read network) required for correlation with certain health datasets.

Data and Information Sharing 4.4.4

The idea of sharing data and information, such as environmental data or analysis results, is gaining more and more attention. There are many examples for open data initiatives running in many domains. Especially for scientists the exchange of data is a core element in support of their research activities.

EO2HEAVEN relies on its Spatial Information Infrastructure (SII) to facilitate data sharing. This infrastructure comprises several web services interfaces allowing interoperable access to different types of data. An example are datasets, which are conceptually treated as virtual sensor or actual sensor observations: in order to make this type of data accessible to other users within the SII, the OGC Sensor Observation Service is used. EO2HEAVEN has developed a wizard tool to support the import of such sensor data into SOS instances. Data owners are thus enabled to describe the structure of their existing data using a graphical user interface.

Whilst our stakeholders realised the potential, there was also an initial reluctance to invest the perceived efforts for adoption, since it does require some IT skills and, most importantly, education about the limited real efforts and major benefits.

Applying the EO2HEAVEN data sharing approach on a broad scale can open up a wealth of data for the research community, which is currently locked up on local computers. However, it requires a paradigm shift to motivate organisations and individuals to follow this approach and tap the potential of data sharing.

Sharing information on current observations or analysis results with end users is mostly facilitate through thematic mapping, which is described in detail in chapter 9

4.4.5 Air Quality Modelling

The modelling of continuous air pollution information from in-situ or remote sensor observations is of great importance for environmentally based health risk analysis.

However, a model can only be an abstraction of the real world, containing a multitude of simplifications and relying on the quality of the input data.

To estimate air quality information, there are basically three commonly used approaches:

- Emission modelling from pollution sources including dispersion modelling relies on direct emission information for pollutant sources like industrial complexes, traffic or combustion heating. However, exact emission rates and the proportion between natural and anthropogenic sources are most uncertain.
- Interpolation techniques or geostatistical approaches to derive coverage information on air quality based on discrete or continuous measurements from in situ or remote sensor systems. Here, the accuracy and reliability of results is based on the availability and distribution of sensor measurements.
- An estimate of air quality is based on the natural and anthropogenic characteristics and assumes that air quality mainly depends on local factors such as land cover, surface structure or traffic density. It asserts that similar regions will most likely share similar air quality characteristics and relies on the availability of representative stations for each region.

Which approach or combination is best suited for an application always depends on available input data and target requirements. For the EO2HEAVEN project the following bounding conditions were defined:

- Input data should be already available and preferably openly accessible.
- Pollutant concentrations should be modelled close to the ground, where people are actually exposed to it.
- Modelling tools should be available at low additional cost or preferably open source.
- The modelling should be robust and transferrable to other regions using comparable input data.
- The performance of the model should allow for ad hoc modelling using web service technology.

Due to different input data, context assumptions and modelling strategies, different air quality models tend to produce different results and thus, need to be validated. Uncertainty of modelling results is inherent and needs to be considered during subsequent data exploitation and/or analysis.

Validation is performed by either leaving out and comparing data subsets or comparing results to external datasets. However, while the former strongly depends on the size and distribution of the selected subset, the latter is hindered by the missing 'golden standard' for air quality modelling. Detailed information on the approaches taken in the EO2HEAVEN scenarios are provided in sections 5.3 and 6.3.

Health Risk Indicators for Air Pollution 4.4.6

EO2HEAVEN has invested efforts to develop or adapt an index that informs and provides advice to health professionals and general population about the health risks linked to the exposure to air pollutants measured routinely. To develop and implement a robust and sustainable index to assess public health risks the index must provide timely information about the health risks faced by those exposed to a certain level of pollution. Hence, such an index should meet in the following requirements:

- The index must take into consideration health risks arising from acute and chronic exposure to air pollutants.
- The index should allow to use locally collected (health) data for the calculation of risk estimates.
- The index must be easy to read/understand by health professionals and general population. A visual representation of the index is desired.

Existing air quality indexes are currently in use across the world (AQI [USA]; CAQUI [EU]; API [Hong Kong]). In spite of some local adaptations and variations, some similarities are visible. Most systems monitor the levels of the 6 major air pollutants (NO_x, PMx, SO_2, O_3) using a graphical color scale to show health risks linked to different levels of pollution. The levels of the major pollutants are usually taken into account for defining the value of the index in a given point of time. However, it is often the highest individually calculated value, i.e. the individual index for each pollutant, that determines the final value for the 'global' index. Therefore, similar index values in 2 geographical points or in 2 points in time do not necessarily represent the same level of pollution.

Cutoffs for health risks linked to levels of air pollution are mostly based on WHO recommendations. These cutoffs are purely based on the health impact resulting from the exposure to a certain level of air pollution. However, national standards for air pollution tend to differ from these WHO recommendations. The differences are explained as governments tend to take into consideration not only health effects but also other factors like:

- Background pollution
- Technical capacity to monitor air pollution
- Social, cultural and economic factors
- Results from a cost-benefit exercise analyzing the different scenarios where different standards are enforced

As most air pollution health indexes are focused on health risks resulting from acute exposures to pollutants, health risks resulting from chronic (and lower) exposure are not clearly represented in many of the currently used indexes. Hence, most current indexes do not fulfill all characteristics necessary for a quality index.

In 2011 Wei-Zhen Lu et al. proposed a revision of the Air Pollution Index (API). The API with some local adaptations is a widely used system all over the world that was originally developed by the EPA from USA. The proposed Revised API (RAPI) deviates from the traditional API by including into the calculation the combined effects of all pollutants. Therefore the new proposed index does not depend on the highest individual value calculated but on the combined results of all pollutants being monitored. The resulting index provides therefore a more accurate representation of the situation of air pollution in a given place and moment taking into consideration the combined effects of all major pollutants and not only the one pollutant with the highest individual index.

Another recent paper from Tze Wai Wong et al. (2012) proposes a risk-based air quality health index. This new approach uses as a base the Canadian AQHI system which in certain aspects resembles the API previously described. The main characteristic of the proposed new methodology is to use relative risks of hospital admissions for cardiovascular and respiratory conditions linked to the major air pollutants to calculate a risk-based, multi pollutant air quality health index. As a result this method provides an index that reflects the risk based on local health statistics.

EO2HEAVEN has adopted a similar approach where the relative risks of respiratory morbidity amongst children with a high predisposition of asthma are calculated though a time series analysis using historical data (2004-2006). The model takes into account potential random confounding and fixed environmental factors, over-dispersion (non-asthmatics), temporal auto-correlation as well as various lags of 1 to 4 days. Risks for each of the pollutants is calculated individually and aggregated to calculate the individual risks are aggregated to result in a number that represents the % Excess Risk (%ER). Different %ER are calculated for specific sub-groups of the population according to their susceptibility to air pollution (e.g. children, elderly, etc.).

The %ERs are finally grouped into categories according to the level of risk. These categories were set up using as a reference based on expected morbidity at minimal exposure (values for short-term or acute exposures). Categories were then defined by multiplying the lower level (for %ER) or by using adjustment factors. Mostly the categories (or risk bands) were created using pragmatic criteria and not health related ones. The resulting index is finally contrasted with a grid that provides tailored behavioral advice to specific sub-groups depending on their susceptibility to air pollution (e.g. children).

Digital Report Handling for Cholera Cases 4.4.7

Cholera surveillance in Uganda is currently managed by the National Ministry of Health surveillance division and is described in the Ugandan Health Management Information System (HMIS) manual (Uganda Ministry of Health, 2010). Following administrative needs of the Ministry of Health in Uganda, the peripheral health facilities and district hospitals have to create surveillance reports of the current cholera cases. Currently, the recording of new cholera cases is an error prone and inefficient process. Data is copied from one handwritten paper form to another, sometimes several times digitized and aggregated, and reported in weekly or monthly intervals only. A major challenge is to overcome duplication and inherent risk of corruption of analog data collection methods and enhance the existing reporting lines with the capability to share and view currently reported data at all stakeholder levels in a timely fashion.

Based on discussions with stakeholders in the field - the hospitals, the district office, and the ministry of health - EO2HEAVEN agreed that a tablet computer-based client app using cell phone communication and Internet technologies would be a major improvement. EO2HEAVEN subsequently developed an Android based App to record cholera cases on a mobile tablet computer to facilitate the digital data acquisition process during the registration and further reporting of cholera patients. The software is described in section 8.3.

Particular challenges that had to be considered in the development phase were posed by health data security requirements, unreliable Internet access and limited bandwidth.

4.4.8 Automated Alert Services

If sufficient reliable information on environmental conditions as well as suitable risk indicators are available, it is possible to implement automated alerting services. Based on user defined alerting criteria these services can alert subscribers via e-mail, text message etc. and ensure a timely notification on critical conditions.

Since EO2HEAVEN had to deal primarily with aggregated anonymized datasets, the implementation of a specific altering service was neither feasible nor within the scope of the project. However, the conceptual requirement is accommodated in the EO2HEAVEN architecture and the usability of event handling methodologies was evaluated. This included the assessment of existing approaches for describing the events using the OGC Event Pattern Markup Language, as well as event subscription services, such as the OGC Sensor Event Service. A central aim of this activity was to support the discussion on event handling within the OGC and to advance the specification development. Whilst basic basic rules are already well supported in existing specifications, future work should address topics such as easier integration of new data sources into alerting application and more robust processing technologies in case of irregular data streams.

In addition to the event processing technology itself, it became apparent, that easy-to-use interfaces are needed, which enable the general public to subscribe to alert services that are relevant for them. Intuitive user interfaces are needed to provide users with an overview of available altering rules, but also for defining individual personal rules, for example focussing on selected locations or specific health conditions. For experts the ability to combine multiple rules is an important feature.

4.5 COMMUNITY AND CAPACITY BUILDING

The major aspect in any project is the acceptance of the results by stakeholders from all levels. As indicated in the introduction to the EO2HEAVEN approach, it is important to ensure that all levels of stakeholders can be included in the iterative approach from an early stage. Involving potential end users to understand their requirements is quite self-evident and a common step to ensure, that software under development meets their actual requirements.

Yet, especially in a research and development project there are other levels to be considered to achieve sustainable results, that are adopted and further developed after the project ends. Champions are needed at various levels, who understand the benefits and are prepared to act as early adopters:

- Policy makers and initiatives will value a feedback on how existing policies comply with real world requirements and should be presented with recommendation on areas where room for improvements has been identified.
- Organisational decision makers need to be educated about potential benefits and value that the introduction of new technologies and approaches can bring to their line of business.
- Local end users don't tend to be interested in IT specifics, but prefer answers to their daily challenges in their own language, i.e. formulated in a way that shows a decent understanding of their domain specific issues.
- Local IT staff may not like to be distracted from their existing chores. However, their support is crucial for the introduction and support of new technologies.

To a degree, integrated projects like EO2HEAVEN take the role of System integrators and ought to be aware of the fact, that each of the above groups has particular requirements, which need to be addressed in an individual targeted manner - but not without putting the individual message in the overall context.

Training and Stakeholder Workshops 4.5.1

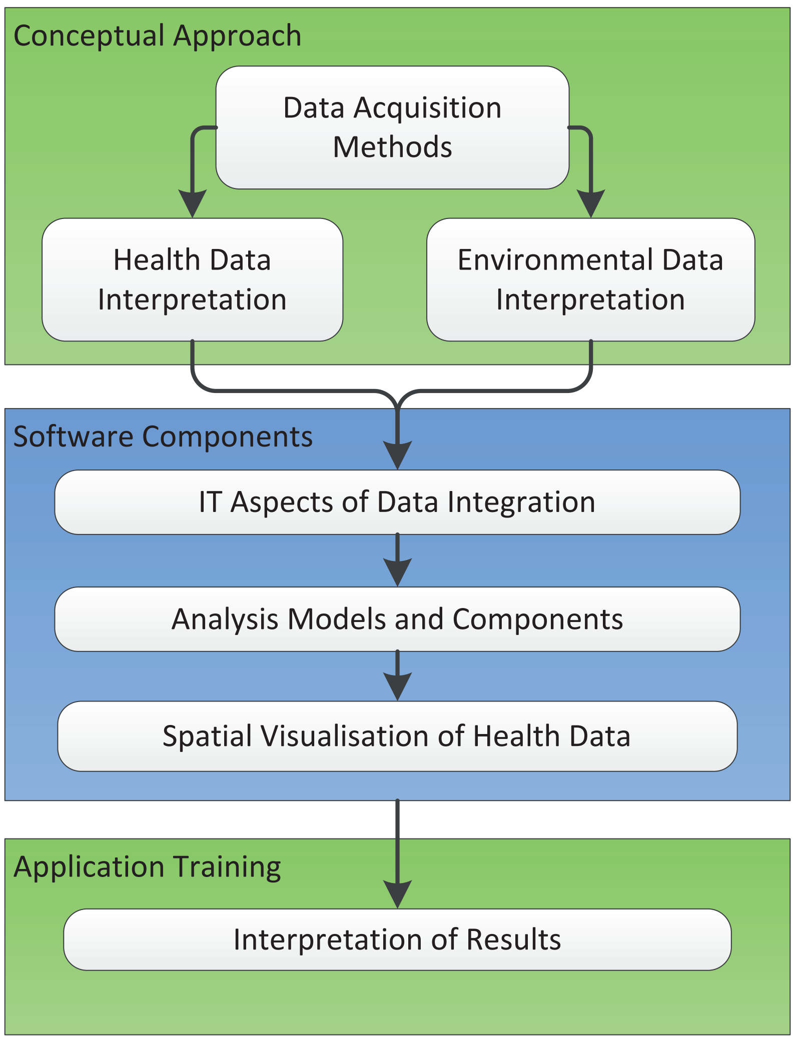
The approach of EO2HEAVEN was to educate stakeholders from the above listed groups through a series of local workshops, which focussed on the particular requirement of each EO2HEAVEN scenario. The objective embraced the "train the trainer" approach, i.e. building up sufficient local knowlegde to enable participants to use and champion the basic concepts in their own organisational context in the long run.

Training on using the mobile cholera client for health inspectors and nurses in Kasese

In total, nine workshops were organised, which combined the principles of showing project results, inviting feedback from the audience and offering training on tools developed within the project. Uptake of project methods and system software on the longer term was supported by demonstrating the implemented methods in the local environment.

A good example is the cholera case study, where a semi-operational system was installed at the Ministry of Health in Kampala. This comprised android tablets running the cholera client registration app, a fully-functional server with data services running in virtual machine environment and a printer with wireless printing functionality for the purpose of paper archiving. This local setup served as an intermediate testing phase towards an operational system at the Ministry of Health and the regional health centers in Uganda.

For the organisation of the workshops, the high-level EO2HEAVEN topics were mapped to learning objectives for potential trainees according to the concept illustrated below:



This approach ensured that within each of the workshops specific sessions could be offered to the relevant stakeholder groups, including field workers, such as nurses and health inspectors, local IT staff, scientists or decision makers. Each training session was then specifically designed to address the previously identified requirements of the targeted audience.

Further information on the developed training material, as well as interactive training courses, is available on the EO2HEAVEN project website.

Standardisation and Dissemination 4.5.2

In addition to the direct work with the scenario stakeholders, EO2HEAVEN put a great emphasis on liaison activities with ongoing global initiatives, like the Open Geospatial Consortium and GEO. Both provided an excellent platform to discuss the project approach and results with external experts, which helped tremendously to align project activities with overall technology and research trends and to raise significant awareness and interest in the EO2HEAVEN work.

EO2HEAVEN has proposed the specification OGC 13-015 "OGC Best Practices for Sensor Web Enablement: Provision of Observations through an OGC Sensor Observation Service (SOS)" including the annex OGC 11-169r1 "OGC Lightweight SOS Profile for Stationary In-Situ Sensors " for an OGC Best Practices Document. The objective of OGC 13-015 is to improve the applicability of the OGC Sensor Web technology to application domains such as health and the environment. The specification provides a lightweight profile of the OGC Sensor Observation Service V2.0 to limit the number of data formats that have to be supported and to have less complex filters for requesting sensor data. A subset of the Observation & Measurement standard is taken as the single response format. This simplifies the implementation of SOS servers and clients, but still covers common use cases. The OGC 13-015 specification also defines how to publish tabular information in a CSV file into a SOS server with the help of an import wizard tool, which is described in detail insection 11.6.3.

The OGC Web Processing Service (WPS) 2.0 Standards Working Group (SWG) focuses on creating the next version of the WPS standard - WPS 2.0. This SWG currently simplifies and modularizes the existing 1.0.0 standard. In addition, WPS profiles and several extensions in order to enhance interoperability are discussed. The EO2HEAVEN final results were directly communicated with this SWG. On the one hand, this is expected to improve the standard and base it on a more solid ground and on the other hand it created a sustainable manifestation of research outputs of this project in further applications.

OGC Specifications and Best Practise Papers are freely available on the Open Geospatial Consortium website: www.opengeospatial.org

5 The Impact of Air Quality on Respiratory and Cardiovascular Diseases in Saxony

The Saxon Case Study focuses on the environmental effects of ozone, particulate matter (about 10 micrometres or less), sulphur dioxide and nitrogen dioxide on respiratory and cardiovascular diseases in the German federal state Saxony. Although air quality can be considered good on average, the thresholds according to the EU-Directive 2008/50/EC on ambient air quality and cleaner air for Europe for ground-level ozone, particulate matter and nitrogen dioxide are regularly exceeded at the measuring stations.

There are two main goals in the Case Study:

- To study the potential impact of air quality to the human health.
- To develop an information system providing the means for analysing and visualizing environmental and health information.

As health authorities are responsible for health promotion, health reporting as well as social and health advisory services, they are considered the main users of the information system. Additional users comprise environmental agencies (e.g. the Saxon State Office for the Environment, Agriculture and Geology), medical and environmental scientists and enterprises related to health or environment.

5.1 EXPECTED STAKEHOLDERS BENEFITS

To identify needs and expectations towards a common system for health risk prediction, a number of interviews and a training/stakeholder workshop have been conducted; mostly with local health and environment authorities, but also health insurance employees, scientists and physicians. In general, the authorities are interested in investigating current health risk situations to inform the public accordingly. From the interviews and workshop, the following conclusions can be drawn:

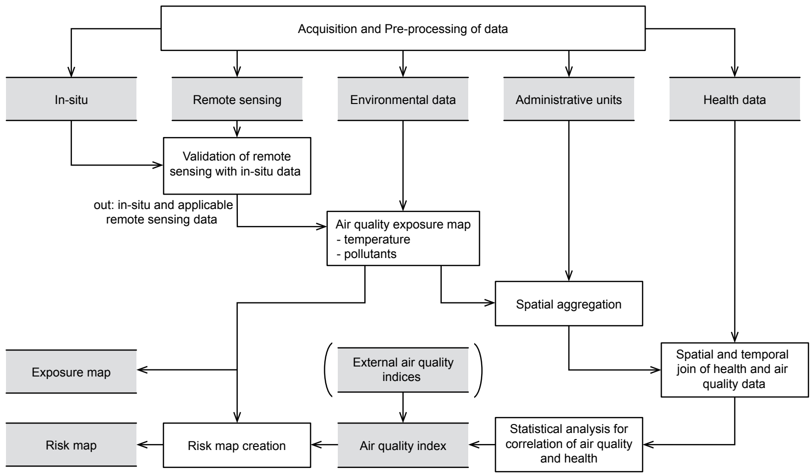
- Almost all stakeholders consider it useful to analyse the correlation between health and environmental data, and specifically the influence of air pollution on respiratory and cardiovascular diseases.
- The majority of stakeholders feel insufficiently informed about data from other application fields. Thus, detailed information on available datasets from different domains is requested. Corresponding datasets shall be identified by EO2HEAVEN.
- Most of the stakeholders experience deficits in current data analysis. They wish for a clear instruction on how to derive a certain health risk from environmental information. Therefore, a kind of best practice guidance for environmental health-risk analysis is requested.
- Risk and exposure maps were identified as most useful in terms of information provision. As most of the stakeholders already use the internet as information medium, an online system is preferred.
- Concerning the spatial resolution, administrative areas are seen as most applicable. Furthermore, seasonal information is most commonly demanded for temporal resolution. However, it is also stated that the spatial and temporal fluctuation of the air quality parameter should be considered.
- In addition to maps, further information on descriptive statistics, correlation and trend analysis is requested. Animated visualizations are preferred, to show spatial and temporal changes.
- An exposure and risk prediction system is seen as highly valuable, especially for a long-term period.

In the long run, the main application for the project and its result, as seen by the stakeholders, are the provision of reliable information on air quality and health issues, an improvement of public relations as well as the possibility of a better prevention of air pollution.

Following these views, the information system could be used by authorities to receive more in-depth information on the impact of air quality on human health, monitor air pollution and corresponding health risks and give recommendations to the public. In conjunction with best practice recommendation for the adoption of the EO2HEAVEN approach, institutions are supported to run their own analysis. The authorities would thus be equipped with the necessary information and tools to identify adverse health effects from air pollution, supporting initiatives for a well-directed and thus, more efficient reduction of air pollution.

5.2 BOUNDING CONDITIONS AND CHALLENGES

The schema below illustrates consecutive tasks as part of the information system identified for this Case Study, followed by the respective requirements:



5.2.1 Data Acquisition and Preparation Requirements

Statistical Data

In this Case Study, medical diagnosis and prescription data from a health insurance as well as morbidity and mortality statistics data were used. The medical diagnosis and prescription data was received from a German Statutory Health Insurance company, the AOK Plus, for the time period 2005 to 2007. This contains day specific information on the person insured (e.g. insurance status, gender, date of birth, place of residence etc.), the time and kind of treatment, the diagnoses and the prescription. For the purpose of our project, treatment data for the following medical conditions was extracted:

- Angina Pectoris
- Ischemic Heart Disease (IHD)
- Bronchitis/Chronic Obstructive Pulmonary Disease (COPD)
- Pneumonia
- Asthma.

Morbidity and mortality statistics data has been captured in Germany since 1993 for all patients who were either discharged from hospitals or died. The data was acquired from the Research Data Centre (RDC) in Dresden and the diagnoses of interest were classified according to the International Statistical Classification of Diseases and Related Health Problems, 10th revision (ICD-10).

Remote Sensing Data

For the determination of PM10, Aerosol Optical Depth (AOD) products gained from the National Aeronautics and Space Administration (NASA) instruments Terra MODIS (Moderate Resolution Imaging Spectroradiometer) and Aqua MODIS and in situ air pollution data are required. Furthermore, to eliminate irrelevant atmospheric conditions, meteorological parameters in the near surface atmosphere, like Boundary Layer Height (BLH) information, are required, since remote sensing data typically represents observed measurements for the whole distance between the satellite and the Earth's surface. More detailed information about this analysis can be found in Chapter 10.

In-Situ Data

In-line with the EU-Directive 2008/50/EC, pollutant concentrations and additional meteorological parameters are measured half-hourly by an in situ station network operated by the Saxon State Office for Environment, Agriculture and Geology (LfULG). The data can be obtained from the LfULG Website or the European Air quality database (AirBase) as tabular data.

Data Analysis Requirements 5.2.2

Air Quality Modelling

The modelling of continuous air quality information from in-situ sensor observations is of great interest for environmental and health risk analysis. However, the accuracy and reliability strongly relies on the number and distribution of in situ sensor stations. Because of the small number of in situ sensors for the state of Saxony, standard interpolation techniques are difficult to apply, especially in sparsely covered regions. Therefore, interpolation techniques need to be developed taking into account the sensor distribution, certain pollutant characteristics as well as uncertainty of measurements and models.

Health Risk Analysis

To derive health risk from air quality data an analysis covering respiratory and cardiovascular diseases was conducted. This was based on prescription data from the AOK public insurance company as well as mortality and morbidity statistics from the German Research Data Centre. The outcome of this analysis (e.g. a mathematical model) can be used to create risk maps and to facilitate the development of an Information System for responsible health authorities.

5.2.3 User Interface Requirements

Besides providing tools and means to address the above described requirements for the information system, the requirements for its user interface are important. The main user requirement focuses on the animated visualization of spatio-temporal data, including time series presentation:

- The change of temperature and air pollutants, including air quality indices, over time and space in a map based on administrative units.
- The change of health information over space and time in a map based on administrative units.
- The change of correlated environmental and health data showing the development of possible risks over space and time in a map based on administrative units.
- Time series of environmental and health parameters as a diagram, showing the variation over time for a single administrative unit.

The results shall be web-accessible in an Internet browser and comprise the following functional elements:

- A risk and exposure map visualization, including an overview map, a map selection and a legend with corresponding explanations.
- Interaction functionality to gain detailed information on administrative units, like current risk classification, descriptive statistics or time series.
- An information button leading to more detailed information on the impact of air quality on the human health, the used risk classification, the methods used for air quality modelling and prediction as well as starting points and recommendations for own analysis and information dissemination.

5.2.4 Data Sharing Requirements in Internet-Based Infrastructures

In addition to the previously described requirements for a web client that primarily supports visualisation capabilities via an Internet browser, there is also a requirement from the research community to facilitate online data exchange. Especially researchers and users from the authorities often want to further analyse the data and visualize the maps in their own software tools. Thus, on-demand Internet based access is required for raw and processed data as well as the computed maps.

Legally mandated information infrastructures such as INSPIRE (cf. section 3.3) already provide the framework for this via standardized data exchange interfaces. By using the same standardized interfaces which are also used in commercial and scientific information infrastructures and software products, a quite universal application potential is created. This allows users to e.g. visualize the computed maps seamlessly in external tools such as Google Earth or desktop GIS software. By registering the computed maps and raw data sets in web accessible product catalogues, they can be offered and published for various external users.

In Saxony, the GDI Sachsen (Spatial Data Infrastructure Saxony) is the regional implementation of the INSPIRE directive. An integration into the GDI Sachsen is required to exploit the project results not only for authorities but also for scientific, commercial and public usage.

AIR QUALITY MODELLING 5.3

Affinity Area Calculations 5.3.1

To model air pollution, affinity area calculations based on in situ measurements are applied. This approach has been successfully applied to air pollution modelling before by McGregor (1996) or the APMoSPHERE project (APMoSPHERE 2005). The underlying cross-correlation concept is that "everything is related to everything else, but near things are more related than distant things" (1st law of geography, Tobler 1970). Corresponding distance measurements are based on attributes which are likely to influence the air quality, such as land cover, elevation, traffic or background emissions.

The model is implemented by a multidimensional Inverse Distance Weighting (IDW) algorithm, using different areal attributes to calculate distances between an area of interest and the in situ stations. A detailed description on how the model works can be found in Wiemann et al. (2012).

The robustness and performance of the model makes it suitable for use in a web service environment, where real time air quality information shall be offered. The workflow of the applied model comprises the following steps:

- Selection of input attributes and corresponding classification of in-situ stations and target areas.
- Determination of optimal attribute weights using the correlation between historic in situ measurement data and corresponding attribute distances.
- Determination of affinity areas based on the in-situ station classifications

and attribute distances to each target area.

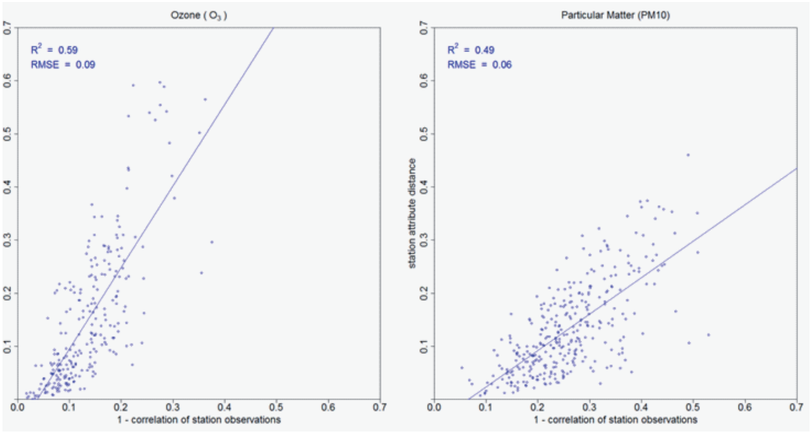
- Air quality modelling within the target areas using the previously defined weights and in situ measurements.

5.3.2 Application of the Air Pollution Model

Within the Saxony case study the previously described model has been applied to calculate air quality information for the years 2003-2007 as the basis for further health risk analysis. As input measurements for the study area the Saxon State Office for Environment, Agriculture and Geology (LfULG) provided measurements for PM10 (26 stations), Ozone (23 stations), NO_2 (26 stations), SO_2 (16 stations) and Temperature (33 stations). As target areas both, the 1km INSPIRE reference raster and postal code areas were used. For attribute distance calculation, the following attributes were included in the modelling:

- Spatial distance (SD) between the in situ stations,
- Land cover information (LC) taken from the Corine Land Cover (CLC) dataset from 2006 at a 100m resolution. To reduce the number of attributes, a land use indicator was calculated for each pollutant based on the methodology described by Janssen et al. (2008),
- Elevation data (EL) from the Shuttle Radar Topography Mission (SRTM) at a 90m resolution,
- Population density (PD) data obtained from the European Environment Agency (EEA) with adjustments using official statistics on the municipality level,
- Road density information (RD) from OpenStreetMap (OSM) calculated from the main street feature classes in OSM in road km per km^2,
- Traffic census data (TC) from LISt GmbH for the major roads in Saxony in couted vehicles per day per square kilometre,
- Emmission data (EM) for PM10 and NO2 provided by the LfULG on street level, aggregated on the corresponding target area.

Based on those input data, the optimal attribute weights were calculated by maximizing the correlation between the attribute distance and the correlation of measurements between each of the in-situ stations. The results for PM10 and Ozone are shown on the following page and indicate a good relationship between the attribute distance and corresponding in situ measurements.

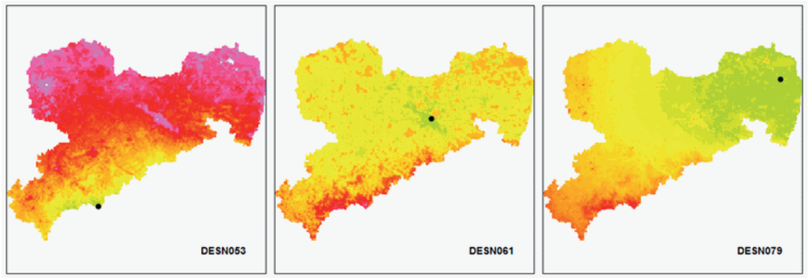


Correlation between the attribute distance
and the corresponding correlation of in-situ measurements

The optimal weights for the attributes within the correlation analysis are:

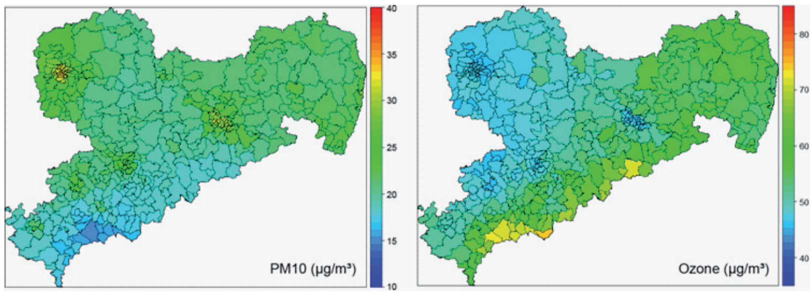
	LC	EL	PD	RD	TC	EM	SD	R^2
O_3	34	38	0	0	8	0	20	0,59
PM10	18	30	0	0	22	10	20	0,49

Based on the previously calculated attribute weights, affinity areas for each station were determined, indicating whether and to which extent an in-situ station can represent a certain target area:



Affinity areas for selected in-situ stations;
green = well represented, red = not represented by the selected station

As the result, each target area is described by a weighted list of in-situ stations. By weighting the corresponding in situ measurements, an air quality map is generated using an IDW algorithm. The following illustration shows an example for the calculation of PM10 and Ozone coverages based on the 5 year average in-situ observations:

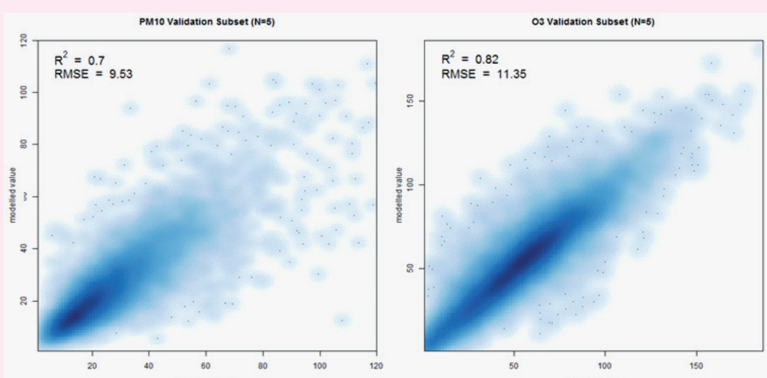


Modelled average concentrations for PM10 (left) and Ozone (right) for the years 2003 – 2007, aggregated on the postal code level in Saxony

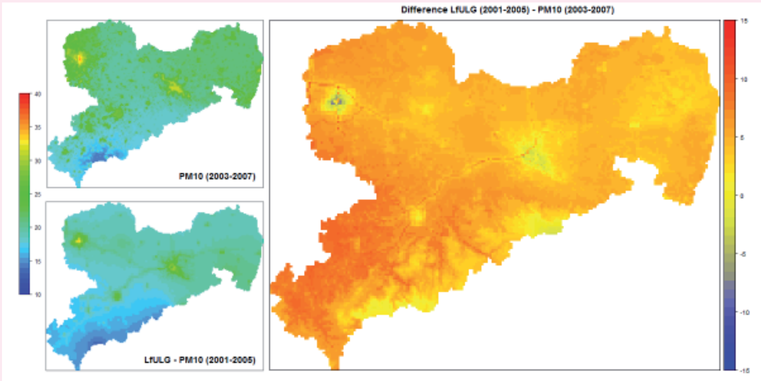
5.3.3 Model Validation

To get estimates on the model accuracy and uncertainty, the following internal and external validation methods have been applied on the model results:

- Internal cross validation with 5 stations randomly removed from the model and taken as a validation subset. This was performed for around 1800 model runs and resulted in an R^2 (coefficient of determination) of 0,7 for PM10 and 0,82 for Ozone. The corresponding scatterplots of the validation are shown below. As stations vary in their significance to the model, the number and type of removed stations considerably influence the model result. Nevertheless, it can be considered as relativly stable and reliable, with a slightly better performance in modelling O_3:



- External validation with model results provided by the responsible state ministry, the LfULG, based on the calculation of annual means for both, the original raster and postal code areas. Compared to the raster (R^2 of 0,31 for PM10 and 0,52 for O_3), the results on the postal code level were much better (R^2 of 0,42 for PM10 and 0,71 for O_3). The results for PM10 are illustrated below, indicating similar trends but also differences in certain areas. However, the models are based on different time periods (LfULG: 2001 – 2005, presented model: 2003 – 2007) and both inherit uncertainty. Furthermore, the LfULG model incorporates dispersion modelling whereas the presented model relies on in situ measurements only, which in this specific case limits the informative value.



Transferring the Model 5.3.4

To test the robustness and transferability of the presented model, it was also applied to the Durban case study area, using the following input data:

- SO_2 Observations from 7 in-situ stations in the Durban area
- Land cover information from the GlobCover 2009 dataset provided by the European Space Agency (ESA) at a 300m resolution
- Household density information taken from the population census in 2001
- Road density information based on the road cadastre
- Digital elevation model

Based on the in-situ measurements, the model successfully produced SO_2 maps for Durban. However, the in-situ sensor network in Durban was rated as not being representative for the whole area because of the small number of stations and their spatial distribution. Furthermore, the meteorological influence is much more dominating the air quality situation in Durban. Thus, it was decided to switch to a dispersion model for air quality modelling. This is described in detail in section 6.3.

5.4 HEALTH RISK ANALYSIS

5.4.1 Health Insurance Data

For the case study, medical diagnosis and prescription data from a health insurance and morbidity and mortality statistics data were used.

Whilst insurance data provides a wealth of information, it must be recognised that it is primarily structured for health insurance billing purposes and not for medical research. Depending on applicable legislation or organisational requirements for this process, it may for example not reflect adequately the point in time when a medication is needed by a patient, but rather when it was prescribed. This emerged during the data analysis when frequent quarterly prescription peaks appeared that could not be related to external, i.e. environmental conditions. The explanation for the peaks was found in two bounding conditions, which are specific to the German health insurance system:

- During the period of the analysed data, patients had to pay a one-time practice fee of 10€ per quarter at their first visit to a doctor. This may lead to effects, that people who frequently need medication tend to stock up on drugs at the end of a quarter to get the most out of this fee.
- Physicians have assigned budgets per year, which can lead to a decline in prescriptions, when the budged is diminished towards the end of a year.

Accordingly, a valid and reliable determination of prevalent or incident cases, i.e. a measure of the population's proportion of ill people and a measure of new occurrences of a certain disease, was not directly possible based on the insurance data. As an alternative approach, the 'Load of Disease' (LOD) was chosen as a surrogate parameter for incident/prevalent cases. LOD means the number of primary care physician consultations in a certain region on a certain day. An advantage of LOD in comparison to incidence/prevalence is that in chronic diseases, such as asthma bronchiale or chronic ischemic heart disease, a measure of incidence/prevalence is less meaningful in measuring the "Health related Quality of life" (HrQoL) than LOD.

The analysis of health insurance data has shown that deriving detailed case information from this data is challenging and that the data can be biased by external, seemingly unrelated aspects. Therefore inpatient-data of the official

Saxon Morbidity and Mortality Statistics of the Research data centre (RDC) has been included in the analyses processes, which contains data of all people admitted to hospital in Saxony.

Statistical Analysis 5.4.2

To analyse the relationship between environmental pollution and health effects, inferential statistics were used on the health insurance data.

In a first step, health data of persons living near the in-situ measuring stations have been analysed in order to assess potential effects of environmental variables on health, independent from an environmental statistical pollution model. Based on the insurance data on physician consultations, all diagnoses of every consultation were aggregated. In a next step the developed algorithms for defining cases were applied in order to reduce the effect of billing process induced peaks. The remaining validated cases of every day were then aggregated onto postal codes, creating “region-days” as a measure of exposition.

To evaluate a possible association between environmental variables and health effects, a mathematical model was developed to analyse the spatial distribution of the results. The association is expressed as LOD-Ratio, i. e. a measure of the association of environmental variables and primary care physician consultations. If the LOD-Ratio equals 1, there is no association between environmental and health variables, wheres if it is lower than 1, there is a decreased association between them. An LOD-Ratio bigger than 1 would indicate an increased association.

Many different types of models, e.g. loglinear, linear, autoregressive, ARIMA and ZIP models, were tried, but all of these turned out to be insufficient, highly indifferent and unstable regarding the parameters as they did not reflect high incidence rates at all. The model for estimating and proving the association between pollutant exposure and occurrence of diseases that was finally chosen is a Zero-Inflated Count Negative Binomial Regression-Model (ZINB, Lambert 1992). The zero-inflated negative binomial regression generates two separate models and then combines them. First, a logit model is generated for the “certain zero” cases, predicting whether or not an event would be in this group. Then, a negative binomial model is generated predicting the counts for those events which are not certain zeros. Finally, the two models are combined. When running zero-inflated negative binomial in Stata, you must specify both models: first the count model, then the model predicting the certain zeros. The zero part of the ZINB model is a logistic model for the event of at least one visit at a physician with a relevant diagnosis, independent from the number of such visits. This part is modeled in ZINB model by the exploratory variable “number of visits total” only.

5.4.3 Morbidity and Mortality Statistics

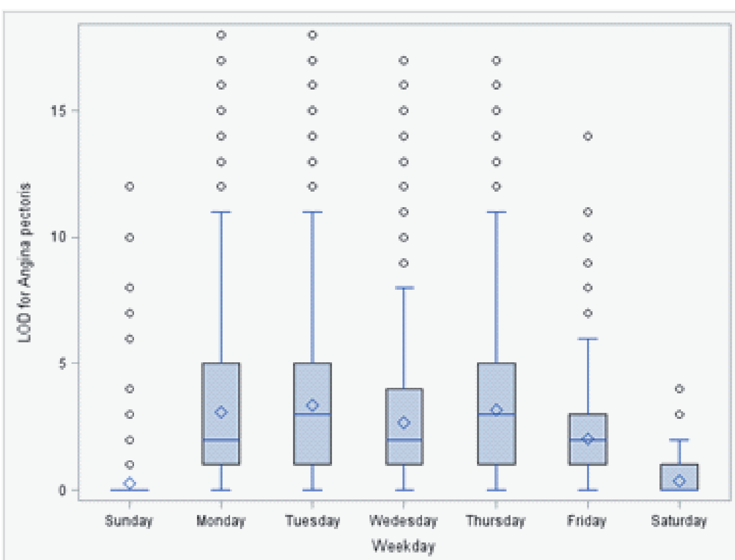
Unfortunately the data protection requirements for detailed level data require a higher than anticipated effort in the analysis processes, so that only descriptive charts have been generated and analysed within the scope of our project.

The descriptive statistics of this data did not show any impact of annual quarters. This is another indication that the health insurance data applicability in this case primarily suffers from its original purpose and the organisational process. Only a slight decline in hospitalization around Christmas was detected, most likely due to the fact that planned surgeries or hospital treatments are avoided by the patients during this time.

5.5 RESULTS FOR CARDIOVASCULAR DISEASES

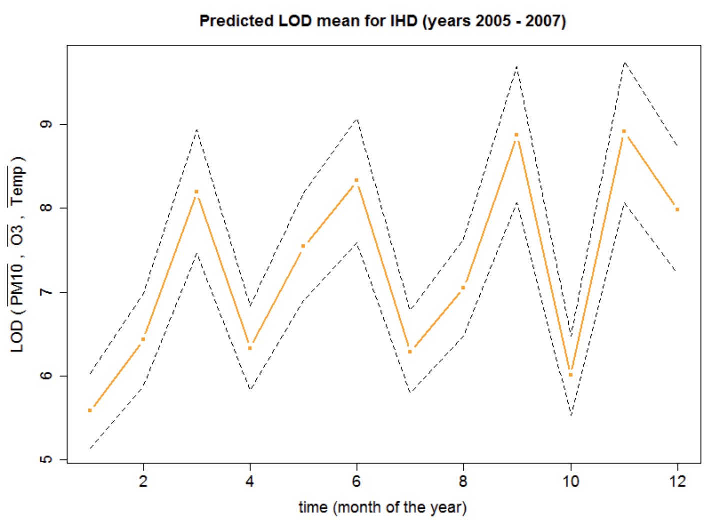
When analyzing the descriptive charts of the relative number of e.g. acute Angina pectoris diagnosis (ICD-10 I20), it became apparent that results could not possibly reflect the real world situation, since most of the given ICD-10 diagnoses per region-day occurred in the last month of each quarter. This occurance could not be explained by descriptive analysis, but is most likely related to the previously discussed effects of the practice fee and billing process.

Moreover, a weekday-effect was discovered, which is most likely influenced by the usual opening regime of German primary care physicians, who usually are open half-day on Wednesdays and Fridays and usually are closed on weekend:



Since these are effects of considerable size (up to 500 %), they are dominating the analysis, and it is therefore difficult to detect any environmental influences on health data, which are estimated to be much smaller. The main reason for this effect probably is that the secondary data, which is being used reflects no real incidences (a measure of the risk of developing some new condition within a specified period of time), but rather cases (i. e. physician consultations). The ICD-10 diagnoses, set up at the consultations, are prone to be biased by medical billing strategies and hence do not reflect reality in a valid way.

When modelling the association between environmental variables and health effects, the effect of quarters again showed up with the highest LOD in the last month of each quarter. No further reliable effects were found:



Predicted mean LOD-Ratio for angina pectoris based on average concentration of O3, PM10 and average temperature for postal code region 01069 (Dresden city)

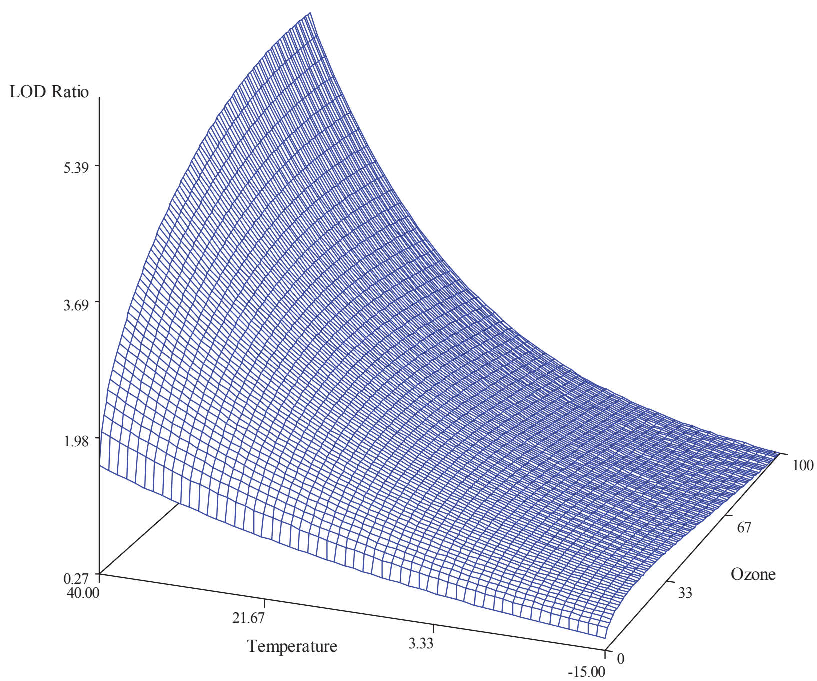
Furthermore, high ambient temperature is associated with an increased LOD-Ratio for Angina pectoris.

Despite the challenges posed by the used datasets, the model itself has good statistical properties, which were evaluated with model diagnostics. Organisations with access to primary data could therefore further test this approach in future research activities.

5.6 RESULTS FOR RESPIRATORY DISEASES

Analysis of the health insurance data revealed a similar effect of quarters and weekdays for the ICD-10 diagnoses of bronchitis and chronic obstructive pulmonary disease (COPD).

However, the analysis of bronchitis/COPD (ICD-10 J20, J40, J41, J42, J44) also revealed an influence of PM10, and in a smaller amount of O_3 too. Increasing values of pollution reach a maximum for the LOD on the same day (for PM10=35 µg/m an increase of LOD of about 50%, for O_3=25 µg/m an increase of LOD of about 80 %). With higher temperatures, the LOD ratio increases, especially in case of higher O_3 concentrations and medium PM10 concentrations at the time of measurement. The following graph illustrates the modelled functional dependence between LOD of bronchitis/COPD and O_3 at the same day (O_3=20):



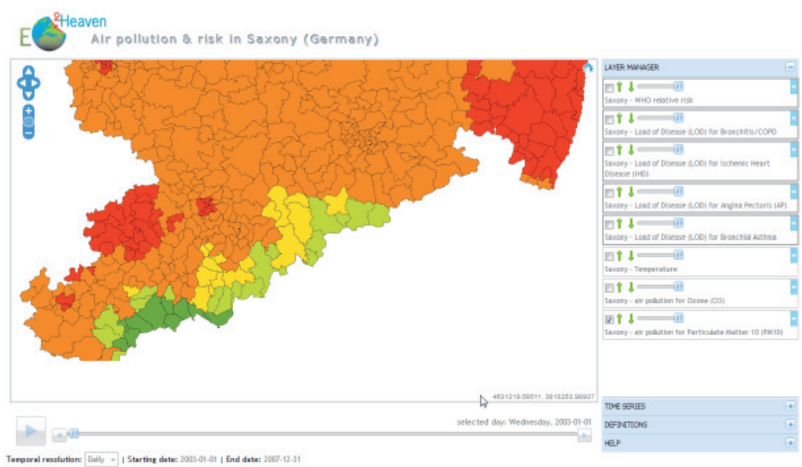
Modelled functional dependence between LOD of bronchitis/COPD and O3 at the same day (PM10=20)

WEB-BASED RESULT DISSEMINATION 5.7

By using Web Service technology and standardized data exchange interfaces, the results of this case study are accessible to user stakeholders via the Internet. The concept following these principles is often called Geodata Infrastructure (GDI) or Spatial Data Infrastructure (SDI).

Time2Maps 5.7.1

For the visualization of available datasets and modelling results, a web application called Time2Maps has been implemented, based on the user interface requirements identified in section 5.2.3. This client combines an interactive display of spatial data (maps) and additional information and allows the user to explore the variation of the provided information on air quality and corresponding health risks over time, displayed as maps as well as timelines:



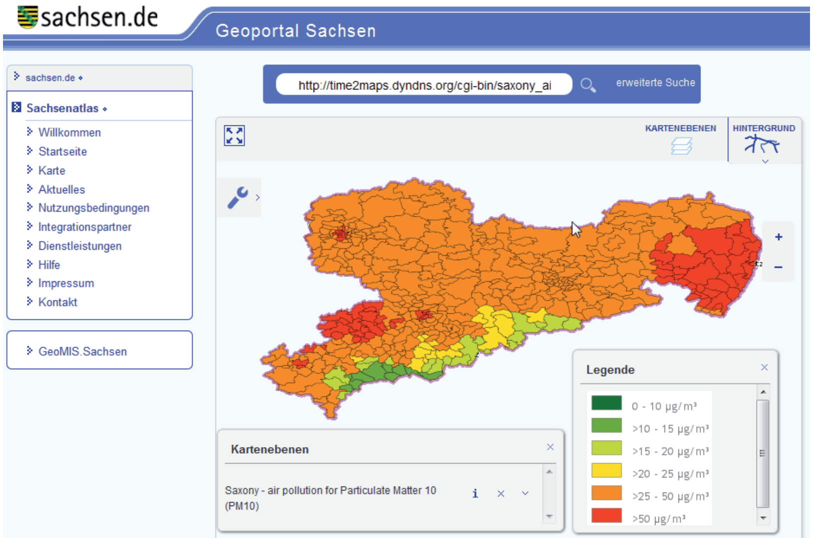
PM10 Map in Time2Maps

Further details about features and the underlying GDI-based architecture can be found in Section 11.1.

Data Sharing in Internet-Based Infrastructures 5.7.2

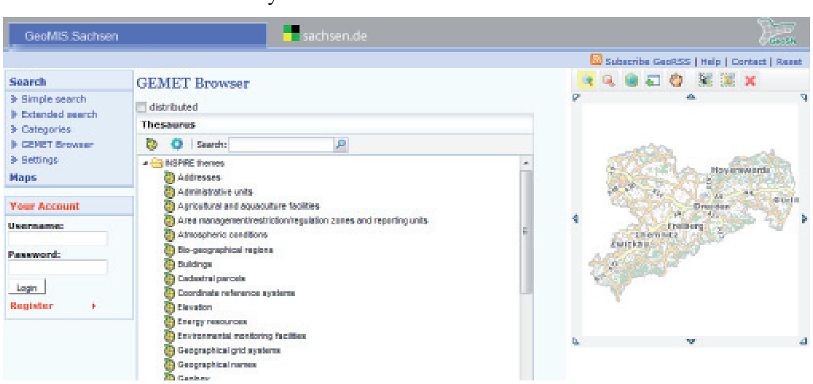
In Saxony, the GDI Sachsen is the regional implementation of the INSPIRE directive. Integration into the GDI Sachsen was achieved to further exploit the project results not only for authorities but also for scientific, commercial and public usage.

The maps can e.g. be visualized in the Geoportal Saxony:



PM10 Air Pollution Map in the Geoportal Sachsen Geoviewer

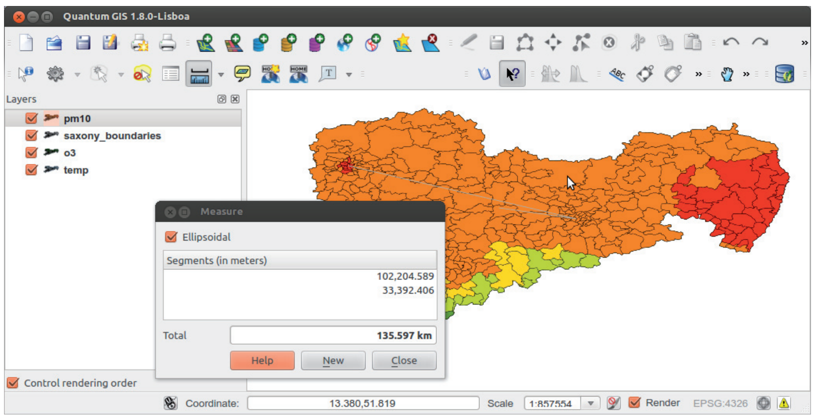
Searching for geodata via usual Internet search engines is often not successful. Finding usable geodatasets and maps requires special search engines called catalogues, allowing for search for datasets over space via published metadata. INSPIRE also suggests to use catalogues to find INSPIRE relevant datasets. The catalogue for the GDI Saxony is called GeoMIS.Saxony (Geo Meta Information System). Consequently, the metadata of the CS maps and datasets are registered and published in this catalogue. A screenshot of the catalog is shown in the Illustartion below. The GeoMIS was used vice versa to search for relevant raw datasets for this case study:



INSPIRE / GDI Saxony Geodata Catalogue (GeoMIS.Sachsen)

5.7.3 Usage in Expert Software Tools

Researchers and users from the authorities often want to import datasets and maps in tools which they usually use for their work. Geoscientists for instance prefer to use desktop-based Geoinformation Systems (GIS) for further analysis. Products, such as GRASS, QuantumGIS, or ArcGIS are often chosen for this purpose. By following the GDI principles, this case study also made their data and maps available for such usage:



Different maps used in QuantumGIS.

CONCLUSIONS 5.8

Health Data Analysis 5.8.1

The analysis of health effects due to environmental variables is associated with some methodological challenges:

- Health data is correlated in two different ways: temporally and spatially. This needs novel approaches, with the best solution being still unclear. Furthermore there are likely non-linear relationships between the level of pollution and health effects, which have to be addressed by using polynomial terms. This in turn makes the model more complex and more difficult to interpret and to communicate.
- Due to the high number of cases which cause nearly every predictor of the mathematical model to be statistically significant, it is hard to apply usual heuristic statistical strategies of model selection (e. g. Aikakes Information Criteria, AIC). It therefore has to be relied on scientific biomedical and physical-chemical knowledge and expertise.

If it were possible to access or collect more suitable primary health data, the statistical modeling would be easier and provide more reliable results. Unfortunately such an approach is beyond the scope of a project like EO2HEAVEN. Budget and time constraints do not allow the collection of primary data through direct surveys. Data protection and privacy requirements prevent the direct use of existing primary data. Though the aim of EO2HEAVEN was to use secondary health data, it turned out that available sources do not necessarily offer sufficiently detailed and reliable data for valid results or predictions. Nonetheless the applied approach and methodologies proved to be working and can be adopted by stakeholders with access to primary data.

5.8.2 Air Quality Modelling

Although air quality modelling has already been studied for decades and is supported by a variety of tools on the market, the challenge of identifying the best solution for a particular use case still remains. The EO2HEAVEN approach demands an air quality model applicable on a large scale, at low costs and with good performance in both, quality and computation time. However, air quality models almost never meet all of those criteria. Thus, compromises have to be made.

As model performance and its complexity usually behave inversely proportional both need to be balanced. As the focus of EO2HEAVEN is on an online system, a slightly higher priority was put on the system performance, leading to the simple but robust and flexible air quality model described in section 5.3.

Concerning the input data a variety of emission and immission data from either in situ or remote sensing data was evaluated regarding its suitability for air quality modelling. Based on the EO2HEAVEN requirements, in situ observations were finally chosen, since they can provide most reliable and near realtime information on air quality. For the interpolation and creation of an air quality coverage, the number and distribution of in situ stations play a major role. In order to achieve meaningful results they should be evenly distributed in space and cover all types of landscape.

Future research activities should evaluate the use of low cost sensors as a solution for areas with either a low coverage of in situ stations or specific local characteristics. They might be a good opportunity to support models by including additional nodes for interpolation and in general increase the validity and reliability of results.

More detailed information on data sources, methods, and implemented process flows are available as public documents on the EO2HEAVEN website.

6 Relationship Between Industrial Pollutant Exposure and Adverse Respiratory Outcomes

The Durban Case Study focuses on the environmental effects of particulate matter (~10 µm or less), sulphur dioxide and nitrogen dioxide on the respiratory health of residents, especially children, in the South Durban Industrial Basin, South Africa. The primary objective was to support the development of a real time spatial-temporal monitoring system that provides information to the health authorities about elevated pollution levels and potential health risks in general, and identify specific areas of elevated pollution within the affected communities.

The Durban South Industrial Basin (DSIB) is at particularly high risk for exposure to significant levels of ambient air pollution because of its geographic relationship with certain stationary sources of air pollutants. Specifically, two major petroleum refineries are within the community, together with a pulp and paper manufacturer, a waste water treatment plant and several small to medium industries. The community is believed to be at risk for intermittent substantial exposure to ambient air pollutants. Available data on sulphur dioxide indicate that average and/or maximum exposures at sites in South Durban have frequently exceeded World Health organization (WHO) and the South African National Ambient Air Quality Standards. Health studies in the area have indicated elevated risk for respiratory outcomes among those exposed.

6.1 EXPECTED STAKEHOLDERS BENEFITS

The health authorities, Environmental Health Services Managers and Environmental Health Practitioners are responsible for health promotion, health reporting as well as social and health advisory services and as such they are considered the main users of the information system. Additional users comprise the Pollution Control and Risk Management Unit and medical and environmental scientists.

Key health authorities, management and health workers, medical and environmental scientists were consulted via workshops and interviews about their needs and environmental challenges they faced in the South Durban

Basin. In general, there was a high interest in receiving a collection of tools that support the integration, correlation and visualisation of real-time air pollution information and corresponding health risk situations:

- Almost all stakeholders consider it useful to analyse the relationship between health and environmental data, and particularly the influence of air pollution on respiratory diseases.
- The majority of stakeholders felt that there was a lack of summarised real-time air quality information from the current data acquisition system, which is presented as easy to interpret information on air quality and related health risks.
- The stakeholders expressed the need for information on air quality and health risk while working out in the field so that they are able to act immediately in the event of an air pollution episode. A web application that is available on a smart phone and tablet is considered a valuable support tool for health workers in the field.
- Spatial resolution at the level of the primary health clinic (PHC) areas were seen as most applicable. However, it the spatial and temporal fluctuation of the air quality pollutants should be considered.
- Visual presentation in the form of maps, graphs and pollution roses were requested, with access to descriptive statistics, correlation and trend analysis.

The common driving force is the desire of the stakeholders to be able to make informed decisions in the event of increased air pollution events.

The Pollution Control and Risk Management (PCRM) Unit of the eThekwini (Durban) municipality are responsible for monitoring pollutant levels and determining the sources of such pollutants, enforcing industrial practice regulations and identifying mechanisms to reduce exposure. Despite having a state of the art in-situ Air Quality Monitoring System (AQMS), resources prevent the implementation of a more extensive system in this major industrial region of Southern Africa. This results in many zones in the city that are not adequately monitored for key pollutants. The EO2HEAVEN aim to incorporate and provide access to data and information from the current system, integrating additional data sources and providing visualisation and mapping capabilities is regarded as a major benefit by the agency.

The Clinical Services Unit in the Health Department of the local government is responsible for the management of the community based clinics throughout the city. Being able to access information on current environmental conditions and trend predictions that result in elevated ambient pollution, will provide this department with information about the likelihood of increased patient load with adverse respiratory outcomes at the local clinics.

In each district within the city, environmental health services managers (EHSMs) of the Health Unit take responsibility for reviewing the daily pollution levels, determining sources and implementing corrective or regulatory action. Receiving support from Pollution Control and Risk Management, EHSMs are required to make management decisions generally with limited data and time. Getting access to more complete data, together with additional meteorological parameters through the EO2HEAVEN tools will support the timely recognition of adverse conditions and identification of pollutant sources.

Scientists at University of KwaZulu Natal see a great potential that improved data access, processing and visualisation, in conjunction with the developed models and process flows of EO2HEAVEN, will stimulate research on environment and health interactions.

6.2 BOUNDING CONDITIONS AND CHALLENGES

6.2.1 Data Acquisition and Preparation Requirements

Statistical Data

In this case study, the medical data used is derived from The South Durban Health Study conducted by the University of KwaZulu Natal (UKZN). This data contains geo-referenced information on the health response of children attending different schools throughout the South Durban Industrial Basin, compared with children from the north communities in the city, and was used in the multivariate risk analysis after undergoing statistical analysis.

Remote Sensing Data

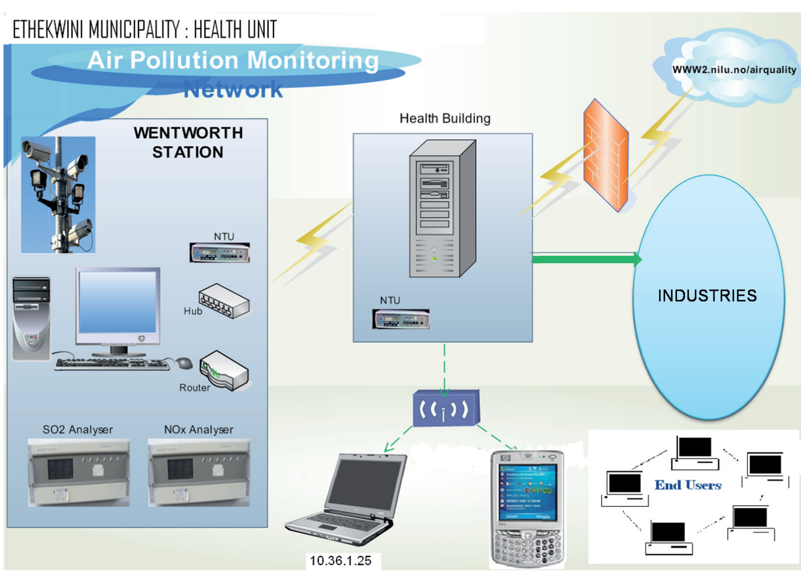
To date, remote sensing environmental data had not been used in any systematic way in Southern Africa for predicting health outcomes. While the data may be available, access is mostly limited to specialist centres. EO2HEAVEN attempted to develop mechanisms to ensure that the relevant data can be obtained by stakeholders, in usable formats, yet data access remained a constant challenge.

In order to ensure that remote sensing data was useful to predict health outcomes, it had to be closely correlated with ground level data. Unfortunately the validation of the available remote sensing data provided strong evidence that a correlation with the in-situ data is not possible. There are a number of reasons for this, particularly relating to local meteorological factors, such as cloud cover and precipitation. These factors influence aerosol optical density, and thus prevent a representation of pollutant concentrations at ground level (cf. section 4.5.2). As a result of this evaluation, remote sensing data was not applied in further modelling and analyses activities.

In-Situ Data

The Pollution Control and Risk Management (PCRM) Unit currently manages the Air Quality Monitoring System (AQMS), which is a real time ground level monitoring system located in fixed "hotspot" locations of the City of Durban:

The meteorological parameters from the in-situ monitoring stations are transferred from the eThekwini Municipality servers every 10 minutes:



6.2.2 Data Analysis Requirements

Air Quality Modelling

The modelling of continuous air quality information from in-situ sensor observations is of great interest for environmental and health risk analysis. However, the accuracy and reliability strongly relies on the number and distribution of in-situ sensor stations. Because of the small number of in-situ sensors for the Durban area, modelling techniques are applied to determine air quality concentrations away from the in-situ stations. Sensor distribution, certain pollutant characteristics as well as uncertainty of measurements and models must be taken into account when modelling.

Health Risk Analysis

Associations between air pollution and increases in respiratory morbidity have been observed in many studies in various parts of the world (Strickland et al. 2010). One of the key objectives specified in the EO2HEAVEN project is the provision of a risk appraisal through the characterization of the potential adverse health effects of human exposures to environmental hazards. The correlation between environmental data and health data controlling for confounding, interaction and seasonality, risk factors for respiratory disease were identified as well as quantified using Generalized Linear Mixed Models (GLMM).

Generalized Linear Mixed Models (GLMM) form a specific group of linear models which can be applied to response variable which are not normally distributed and where the linear predictor function contains both random as well as fixed effects. The model consists of four components: 1) a response, 2) a set of covariates of which we aim to investigate the effect on the response variable, 3) a random component which accounts for random variance in the response variable resulting from the sampling design, and 4) a term describing the residual error not predicted by the model. For this study a GLMM using a log link for count data (assumed Poisson distribution for respiratory morbidity count data) was applied to model the effect of air pollution and meteorological conditions on the outcome of respiratory morbidity among school children in the DSIB.

The fixed component of this model consists of predictors driving the observed trend in a response variable (e.g. respiratory illness) presented here by the different pollutants and meteorological confounders described. The strength and nature of the relation between the response and each of the predictors is predicted by the model. The random component of the model consists of factors which introduce variation in the response variable which could potentially mask or bias the effect of the predictors. The random effects presented here are posed by Schools (7 groups) and Students (repeated measures within schools).

The difference between the observed number of respiratory symptoms observed in a student and the predicted number based on the GLMM-model provides an error term of random residual variation in the data not accounted for by the model, which can be used to assess the accuracy of the model prediction.

User Interface Requirements 6.2.3

The main user requirement focuses on the animated visualization of spatio-temporal data, including time series presentation. The requirements identified for this case study were identical with those from the Saxony case study as described in section 5.3.2 in the previous chapter.

However, a different implementation approach was evaluated for this scenario to support additional requirements of the research community to dynamically include alternative data sources.

AIR QUALITY MODELLING 6.3

Pollutant dispersion modelling is conducted using the Cambridge Environmental Research Consultants' (CERC) Atmospheric Dispersion Modelling System (ADMS Urban) as part of the EO2HEAVEN system for South Durban. ADMS Urban simulates a wide range of buoyant and passive releases to the atmosphere, drawing on the latest plume dispersion mathematics. Output for criteria pollutants has been extensively validated against field data sets in the European Union and the American Standard Test Methods, while consultants have proven its reliability against measured data in South African case studies. For the purposes of this study, an extensive emissions inventory (comprising multiple point, line, area and volume sources), on-site meteorological measurements, and a digital elevation model are inputted for dispersion calculations. The model is run with live meteorological data at ten-minute intervals to produce short-term pollution scenarios for sulphur dioxide and particulate matter. It is likely that this repertoire will be extended to include the oxides of nitrogen and benzene in the near future. The calculated plumes are then mapped with concentration contours for display purposes.

To aid with source identification, a pollution rose is also produced every ten minutes for each areal unit. This diagram indicates the average pollution concentrations from prevailing wind directions. Calculates are based on measured wind direction data and ADMS Urban pollution concentration outputs.

6.4 HEALTH RISK MODELLING

A model written in R language is used to determine the health relative risk pollution based on the concentrations from the ADMS and meteorological data from the eThekwini Municipality server. The outcome of this model is a health risk map and a graph indicating the pollution concentration and relative health risk (cf. 6.5.2).

The mathematical model describing the relation between health outcome and environmental exposure results in an estimate of expected respiratory morbidity. As such a measure is not easily interpreted, this measure is transformed based on the estimated baseline morbidity at when all pollutants were zero using a simple conversion algorithm. The final health indicator therefore represents the risk ratio over the background risk in the DSIB due to exposure to certain levels of air pollution.

6.5 DURBAN WEB CLIENT

The EO2HEAVEN advanced generic client described in section 11.2 has been customised for the Durban case study to provide access to the main results for the Durban environmental health services managers, health practitioners, technical staff of the Pollution Control and Risk Management Unit and academics. The objective is to support the automated computing of pollutant concentration and health risks based on a predefined process workflow.

The web client includes login security. It presents OpenStreetMap as a background, and static layers such as clinics and the different primary health care (PHC) areas.

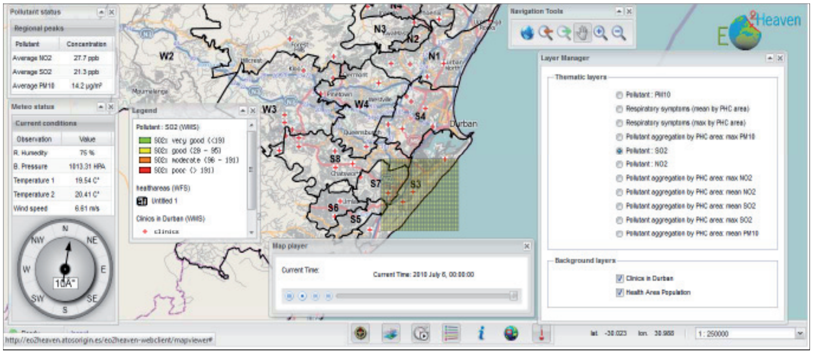
6.5.1 Interactive Maps

Pollution Concentration Map

The user has the option to select a pollution concentration map for SO_2, PM10, or NO_2 per point (250m x 250m grid) from the Layer Manager. The user may also select the mean or maximum concentrations per pollutant per PHC area. The legend for the pollution concentrations assists the users to determine the level of pollution for that period.

Health Risk Map

The user has the option to select a PHC area and determine what the health risks (mean and maximum) are for that area. The legend for the respiratory systems assists the users to determine the level of risk for that period:

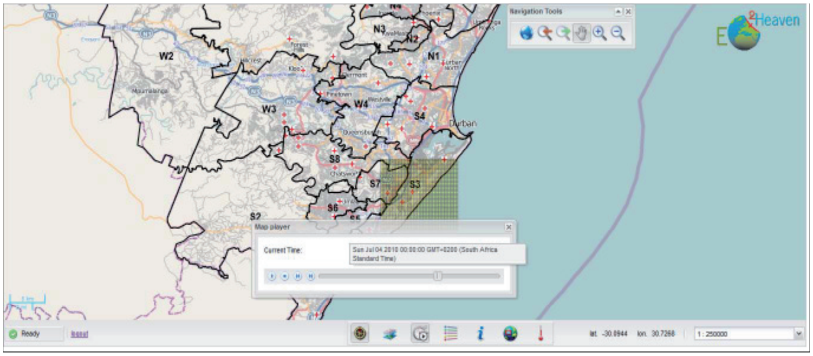


User Interface of the Durban Customised Client

By clicking the 'current status' icon, the user can explore the following information:

- Maximum concentration of pollutants (SO_2, PM10, NO_2) for the entire Durban area
- Meteorological status (pressure, humidity, ambient temperatures, wind speed and wind direction depicted in a 360° chart)

Scientific data based on epidemiological studies suggests, that a lag time from 24 hours to five days from exposure to effect can occur. Therefore a time slider function was implemented, that allows the user to visualise the respiratory symptoms for the last 4 days and the pollution concentration for the last 24 hours for SO_2, PM10, and NO_2 either for a given point in time or as a time lapse animation:



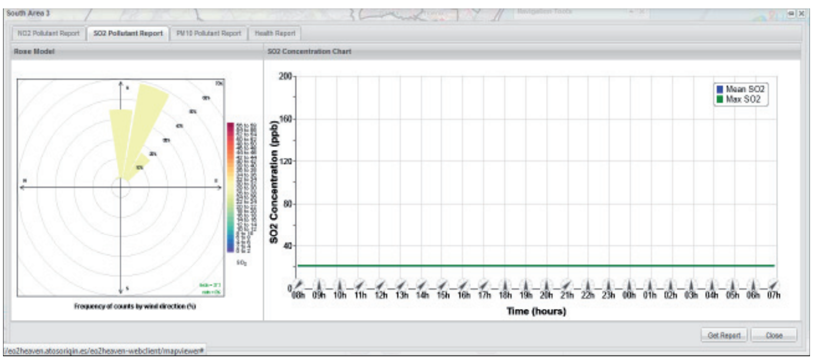
Animated Mapping over a Time Period

6.5.2 Summary Reports

Summary reports for the different pollutants and health per PHC area or grid point (250m x 250m) can be produced and saved by clicking on a point of interest.

Pollutant Reports

Four reports are available for NO_2, SO_2, PM10. The NO_2, SO_2 or PM10 reports reflect a line graph of concentrations plotted against the hourly wind direction for the last 24 hours according to the point selected. A 24 hour pollution rose is also shown for each report. The pollution rose is created by a model using a given point and a given pollutant parameter computed for the last 24 hours from the ADMS model outputs and the meteorological data. Such a report is computed for a point of the grid (output of the ADMS model) or a PHC area. The curent PHC area values can also be displayed in the report:



Health Report

The health report reflects a line graph of the mean pollution concentrations for every hour over four days plotted against the mean and maximum respiratory symptoms over a 24 hour period for last four days for a point or PHC:

CONCLUSIONS 6.6

The relationship between industrial pollutant exposure and adverse respiratory outcomes has been well documented in the scientific literature. Much debate exists around causality of effects, role of specific pollutants and the populations particularly vulnerable to elevated pollution. Although certain pollutants, such as oxides of nitrogen, particulate matter and sulphur dioxide are known to result in adverse out-comes, the ability to use this information to improve the life of affected and vulnerable sub-populations is limited. This case study provides a good example for the difficulties to characterize exposure throughout the city, and hence the challenge to predict adverse health outcomes. To provide a system that alerts users to the risk of childhood respiratory problems associated with air pollution, this needs to be recognised and addressed.

This desire to seek new methods of ambient pollution assessment using latest available technology was based on the knowledge that sustainability of in situ environmental monitoring is resource demanding. In addition, the lack of health data at municipal level and the lack of data on health-pollution relationships meant that new methods and technologies needed to be actively pursued, particularly by local government.

Unfortunately, the results from the validation studies concluded, that readily accessible remotely sensed data cannot be used to predict ground level pollution. Whilst this was a disappointment to researchers and agencies within local government, it emphasizes the need for better understanding the conditions under which the use of advanced technologies can be more widely used. Hence more robust methods, beyond using remote sensing are necessary for ground level air pollution monitoring.

Nonetheless, the approach, methodologies and tools of EO2HEAVEN provided significant benefits to the stakeholders. The core components for an health and environment information system are available with the necessary interfaces to include new data sources, so that this can be extended to an operational system. The major challenge for a widespread deployment is ongoing engagement with end-users and that the necessary resources are available for end users to utilize the systems that have been developed.

The collaborative network that has been established at an international level and the complex nature of setting up systems and working across international boundaries has been a valuable experience. Based on these lessons learned, stakeholders have expressed their interest to continue their collaboration and to get engaged in the planning stages for future projects.

7 Links Between Environmental Variables and Cholera

The main purpose of the Cholera Case Study is to provide insights into the dynamics of cholera at the pathogen level (the organism itself) and as a disease. These insights can then be used for decision-making purposes. In order to accomplish this, the EO2HEAVEN project generated new research outcomes and developed software and analytical tools to:

- support researchers to collect, handle, analyse, integrate and model environmental, human related and health data from different sensors and sources;
- visualize data and results for research purposes as well as decision making at different levels, using historical and current data.

The cholera case study was conducted in the area around Kasese, south-western Uganda, which was selected for its proximity to Lakes George and Edward and the equator. Environmental and human-related data were analysed in conjunction with health data in order to characterize and explain the spatial and temporal patterns in cholera outbreaks. In this case study, remotely sensed data, in-situ data from weather stations and ex-situ data collected as part of field sampling campaigns were used. The results of the Cholera Case Study can be used to seed the development of an Early Warning System.

7.1 EXPECTED STAKEHOLDERS BENEFITS

The following stakeholder groups can benefit from the work carried out in this case study:

- New modelling approaches and tools are provided for researchers, such as epidemiologists and disease modellers.
- An improved understanding of the dynamics of cholera outbreaks will help policy and decision makers, as well as health practitioners and health officials at different levels in government to develop and implement mitigation strategies.

BOUNDING CONDITIONS AND CHALLENGES 7.2

The process of extracting meaningful, actionable information from the combination of health, social and Earth observation/environmental data is complex challenge. Difficulties are magnified in less developed countries where the burden of disease is felt, for access to datasets is limited by technical factors such as Information and Communication Technology connectivity and poor data, as well as human factors like a lack of training in the complexities of working with Earth observation data.

CHOLERA RESEARCH OUTCOMES 7.3

The EO2HEAVEN project produced some new and interesting cholera research outcomes from the Uganda cholera scenario.

Isolation of a New *Vibrio cholerae* Strain 7.3.1

An enterotoxigenic *Vibrio cholerae* strain (named UG01) was successfully isolated from a water body in the study area. The importance of this finding is that the strain carried a tcpA gene that differed from typical El Tor and Classical biotype tcpA genes that are usually isolated. Preliminary study results suggest that the strain is hyper-virulent. However, further studies are required to estimate the risk that this strain may pose in future outbreaks and epidemics. Further work is ongoing to elucidate the phylogenetic relationship of UG01 with other African *Vibrio cholerae* strains, and to determine if UG01 may be related to hybrid El Tor strains as reported by Safa et al. (2008).

New Relationship Between Rainfall and Cholera Outbreak Proposed 7.3.2

The direct and indirect roles of rainfall in cholera outbreaks are generally accepted as being that of

- a spread mechanism, whereby different areas are connected by water ways or the inundation of the landscape,
- a cause of infrastructural damage to water and sewage treatment plants,
- a factor that can dilute the concentrations of *Vibrio cholerae* bacteria, during wet periods or concentrate the bacteria during dry periods in available water sources, and
- a driver of changes in chemical, physical and biological factors which affect the survival and growth of *Vibrio cholerae* in water bodies.

However, the cholera-rainfall association does not always explain cholera outbreak patterns when using the correlation between rainfall amounts and timing of rainfall events and that of cholera cases. It is therefore necessary to identify cause-effect mechanisms which can explain correlation patterns.

A new explanation of the relationship between rainfall season, UV radiation and cholera outbreak is proposed, based on analysis of satellite remote sensing data and data gathered during the field sampling campaign.

Preliminary study results show that the amount of UVA (320-400nm) and UVB (280-320nm) radiation reaching the Earth's surface (using satellite remote sensing measurements) and penetrating water bodies, combined with changes in the water temperature (using field data measurements), may trigger the start of a cholera outbreak on the regional scale. A new indicator was developed in the EO2HEAVEN project to account for this relationship, explained below.

The indicator is a UV index which incorporates the role of water temperature in the response of the *Vibrio cholerae* bacillus when exposed to UV light. UV-induced DNA damage can occur at any temperature but the photo-repair mechanisms enabled by the bacillus after exposure is temperature dependent.

To develop the indicator, a minimum threshold temperature of 26˚C (±0.5˚C) was calculated based on water temperature data for the lakes, measured when field samples were taken in Uganda and from remote-sensing satellite measurements. The amount of UV light reaching the Earth's surface and penetrating water bodies up to certain depths is affected by local atmospheric and water properties. Cloud cover was identified as the most important atmospheric factor attenuating the amount of UVA and UVB radiation reaching the Earth's surface in the equatorial region. Water turbidity is an important water property that controls the amount and depth of UV light penetration in a water body; it is mainly driven by the amount of runoff during the rainy season. This may provide another explanation of the cholera-rainfall season association. The indicator is thus a combination of water temperature threshold, cloud cover percentage and water turbidity and is aimed at gauging when *Vibrio cholerae* that are present in a water body may become toxigenic. Research work on the role of UV light in the change of the virulence status of *Vibrio cholerae* is still ongoing.

7.4 AN INTEGRATED APPROACH TO STUDY CHOLERA DYNAMICS

The EO2HEAVEN project showed that it is useful to approach a local or regional problem such as cholera outbreak with a multi-pronged research programme, with the proviso that different methods and data sources need to be well understood.

Data Integration 7.4.1

It is important to integrate field measurements and laboratory results with satellite remote-sensing data to provide cause-effect explanations for observed patterns and correlations in the cholera case data. The value of a chemical, physical and biological (including microbial) monitoring program for relevant water bodies is shown in the above example of the UV index that was developed.

However, it is important to understand the properties of the different data sources:

- The accuracy and temporal and spatial resolution of satellite data may compromise the results of analytical models developed to simulate pathogen dynamics.
- Changes in identified chemical and physical properties such as pH, salinity and nutrients can only be measured in-situ. The deployment of sensors in water bodies to measure and transmit required chemical and physical data on a continuous basis will enhance monitoring programs. Currently no sensor exists that can measure the presence and concentration of toxigenic and environmental strains of *Vibrio cholerae* in water bodies.
- Satellite data are primarily used to track changes in drivers (e.g. rainfall) of conditions that influence pathogen dynamics (for example its ability to survive, grow, attach itself to sediment particles when nutrients are limited and thus disappear from the water column).

The Role of Climate Factors and Scale Issues in Modelling Cholera Outbreaks 7.4.2

Climate variability as represented by changes in Sea Surface Temperate (SST) in different parts of the Indian ocean (i.e. a regional signal) plays a stronger role in anomalies observed in rainfall, cloud cover and temperature over land in the study area than ocean-atmospheric coupled signals extracted in the Pacific Ocean (i.e. global signals)

Environmental and climate factors play a role in cholera dynamics on the regional level. No significant correlations between rainfall aggregated at different temporal resolutions (e.g. daily, weekly, monthly and seasonal) and local cholera cases, i.e. at the village, parish and sub county levels were observed.

Environmental factors at the larger spatial scale, i.e. on the regional level, correlate with cholera cases on the local level, i.e. within the Kasese district. This can be explained by various factors such as the connectedness of areas via the road infrastructure and waterways such as rivers or shared water bodies such as lakes. Patterns in the cholera case data are also more evident when aggregating the case data at a coarser spatial scale.

These scale issues may imply that for modelling purposes (especially early warning models) the exact location of a cholera case is not as important as the knowledge that cases are reported within the larger area and for a continuous amount of time. Due to human mobility factors and the connectivity of particular areas via water pathways and water bodies, it is highly likely that an outbreak or outbreaks will occur in different parts of the region. This implies that a regional response is required when the first cases are reported in order to reduce the spread and magnitude of outbreaks.

7.5 CORRELATION ANALYSES PROCESSES

Data analysis processes are the mechanism by which the potential role of environmental and human related factors in cholera dynamics in Kasese is determined. There are, broadly speaking, three steps.

First, data about different environmental variables, typically from satellite remote sensing sources, are analysed together with cholera case data. The data are aggregated at different spatial and temporal resolutions and validated against in-situ observational data, before being run through correlation checks (based on statistical, signal processing and visualization techniques) to determine similarities in underlying patterns and if any relationships exist between the case data and the individual environmental variables. Further analyses of individual variables (using the same or similar techniques) is undertaken to tease out which ones are important in relation to each other. The case data and local environmental variables are then related to global environmental variables, such as the ENSO (El Niño – Southern Oscillation) signal, to test if these large scale variables can be linked to cholera outbreak.

Second, data collected in field sampling campaigns are analysed (using for example microbiological techniques). The presence/absence of cholera is checked for, and if present, whether the toxigenic strain is visible. These results are compared with the individual environmental variables and are also analysed in context of seasonality. Essentially, the analyses are searching for environmental and climatic drivers of the presence or absence of the bacteria. An aim is to identify thresholds for important environmental variables, above or below which changes in the cholera pathogen and the disease itself can be expected.

Third, the results from the analyses of case data, environmental variables and sampling data are put into context of social and geographic data - population density, land use practices, demographics, topography, water use practices etc. - and then tied to epidemiological data.

CONCLUSIONS 7.6

A multi-disciplinary and national approach was adopted as part of the EO2HEAVEN project to address cholera in the Kasese district in Uganda. This included expertise from different countries and continents in the fields of microbiology, epidemiology, remote sensing, information technology, data standards, data analysis and modelling. Such a multi-national and multidisciplinary approach provides advantages as well as challenges.

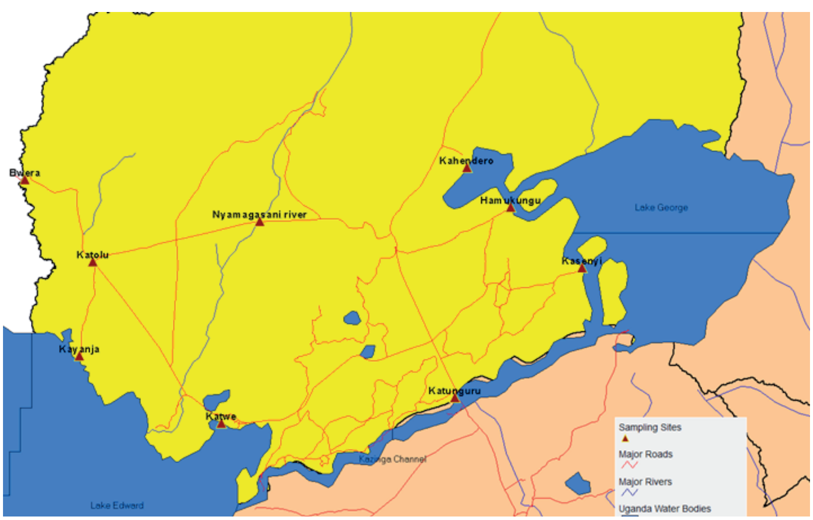
This approach produced valuable insights into a complex environmental related infectious disease such as cholera. New insights were generated that provided new information regarding the dynamics of the disease and the pathogen, *Vibrio cholerae*, for

- an inland area or a landlocked country such as Uganda as compared to a coastal area and
- an area located close to the equator.

The different skills and expertise led to different approaches and ways of identifying underlying problems, developing solutions and integrating the solutions and results. These can form the basis for future research work and support for decision-making concerned with the development of optimal in terms of time, effort and cost, intervention strategies. For infectious disease research and support for developing appropriate, cost-effective intervention policies and strategies, it is important to take advantage of capabilities and advancements in the supply and accessibility of data, computer hardware and software, and ICT tools and expertise in general, e.g. standardisation of data formats, capturing, reporting, provision and accessibility of data.

The major challenge of such a multi-disciplinary approach is the communication between experts from different fields and bias in the interpretation and understanding of what is required and what is technically feasible or not. Experts tend to be strong in a specific field or domain, rather than multiple fields and domains. However, it is becoming increasingly important that health domain practitioners and researchers are trained, from an early stage in their education establishments and workplaces, in aspects of computer science, data analysis and modelling. Likewise, data analysts, modellers and ICT practitioners need to make an effort to understand, to some extent, the application domain, such as health, ecology etc. This can help to limit technical confusion and misunderstanding. Another option is to have a project manager/ technical leader with sufficient understanding of all the fields involved, in order to enable communication between experts from different fields and help with the interpretation of what is required and what can be done.

Especially in complex scenarios, such as the Uganda Cholera one, another important aspect is the requirement that software development and implementation can react quickly to new or altered requirements. Whilst software developers are often concerned about finalising a “release quality” product, the scientist may need a quick prototype to address a new research model requirement. These two positions can create a deadlock very quickly. The software under development may ultimately not be used by a researcher, because a particular form or style of analysis is no longer suitable for the research problem or needs to be altered in some way. Both researchers and developers need to reduce the distance between them and look for small gains, made often. This allows more fruitful evolution of the project software and allows more ICT’s to be used in the end. Nonetheless, the emphasis should always be on building robust, reusable components, rather than ‘quick and dirty’ scripts - so patience and understanding is required on both sides.



Nine field sampling sites in Kasese, Uganda

Taking in situ measurements of the physical properties such as water temperature

8 Cholera Case Surveillance and Health System Enhancement

In the Kasese district of Uganda cholera case detection is currently realized by a passive surveillance system, only proceeding to active surveillance after an outbreak has been confirmed. Yet, the vertical stratification of the health information system over multiple administrative levels confines the prompt dissemination of information. This also introduces error in the information flow, potentially delaying the signalling of an outbreak when it occurs. Even though alternative procedures exist to flag initial cases for follow up investigations, the protocols are fuzzy and the system is prone to human error. Moreover the lack of resources limits the efforts for efficient active surveillance. It appears that the flow of information from the peripheral health facilities to the ministry of health is hampered by lack of tools to collect, store and disseminate this information. Also a considerable amount of resources is invested in collecting information and summarizing data in district reports, which are then again used to report to the national government.

EO2HEAVEN has addressed the requirements for improved information flow and corresponding mitigation responses with the development of a mobile online data collection application, as well as data processing and visualization services, which

- support health workers to collect and retrieve up-to-date cholera case data;
- visualize data and results for research purposes as well as decision making at different levels, using historical and current data.

The activities carried out in this context are closely tied to the results of the previous case study. They provide a link between ongoing research into the triggers and dynamics of cholera outbreaks, the on-the-ground mitigation carried out by health workers and the development of preventive strategies.

EXPECTED STAKEHOLDERS BENEFITS 8.1

The EO2HEAVEN study on cholera case surveilance and health study enhancement offers a range of direct benefits to both health practitioners working in the field as well as health authorities working at the national ministry of health, as it

- reduces the number of human errors induced by the current reporting system,
- reduces the amount of resources spent on data management and dissemination,
- improves response time,
- increases the effectiveness of surveillance by facilitating the exchange and interpretation of information at a sub-district level and
- facilitates communication across all levels of the health system.

BOUNDING CONDITIONS AND CHALLENGES 8.2

There is an urgent need and high potential for the development of a standardized data collection protocol and associated tools, which facilitate data collection and access to health information across all levels of the health infrastructure. However, such developments have to take into account a number of limiting aspects:

- There are only limited technical skills and resources.
- Budget constraints require pragmatic solutions at reasonable costs.
- Lightweigt mainstream hardware has the highest chance of acceptance.
- Information presentation needs to be simple and intuitive.
- Traditional reporting needs to be complemented, allowing for a gradual migration to new processes and technologies.
- A data processing protocol should robust, simple and re-usable in other applications.
- Whilst mobile Internet access is widely available, a reliable backbone, i.e. wide area network or web infrastructure cannot be expected.

Based on these conditions, the development of a smart-phone or tablet-pc based android app with a matching server backend was deemed most feasible.

8.3 DIRA - DISEASE INCIDENCE REPORTING APPLICATION

The solution that EO2HEAVEN implemented to enhance the cholera case surveillance and reporting is based on three main components:

- A mobile app to record new cholera cases and upload it automatically to a server.
- A server that collects and stores all incoming new cholera cases sent by the mobile app devices.
- A simple web client to manage the server side, including export of the data to csv files and user account management.

8.3.1 Server Backend

The server receives the collected patient data from the mobile app as a JSON message over an encrypted HTTPS/REST internet connection. Successful receipt of the data set is acknowledged with a HTTP response 400-OK and the the data is then stored in a database.

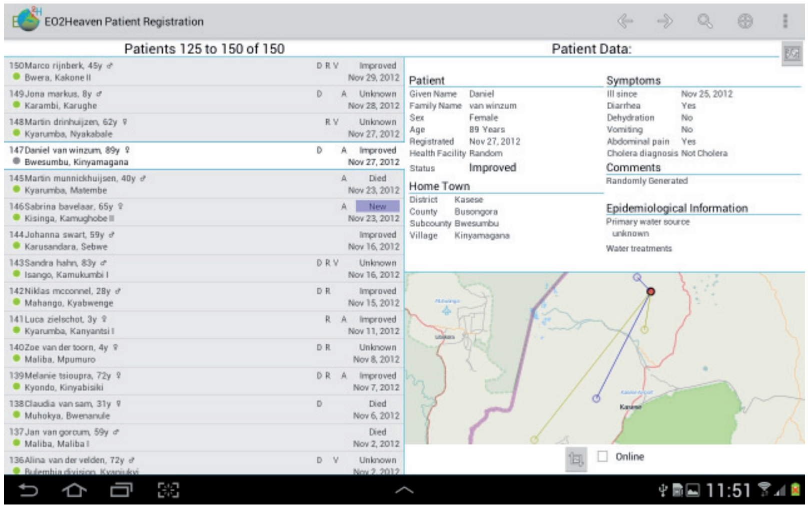
A simple web client on the server side provides administrational and maintenance functionalty, such as the management of user accounts or data export to csv files. This ensures compatibility with the existing analog system and allows users to maintain printed records on the reported incidents. The transmission of data to the client is currently not planned.

8.3.2 Mobile Application

The Disease Incidence Reporting Application, or DIRA, was designed to run on almost all Android devices. In terms of usability, a tablet-pc with a 7”-10” display is recommended. All data is primarily stored locally, with copies transferred to the server when an internet connection, i.e. the mobile phone network or local WiFi-based internet, is available.

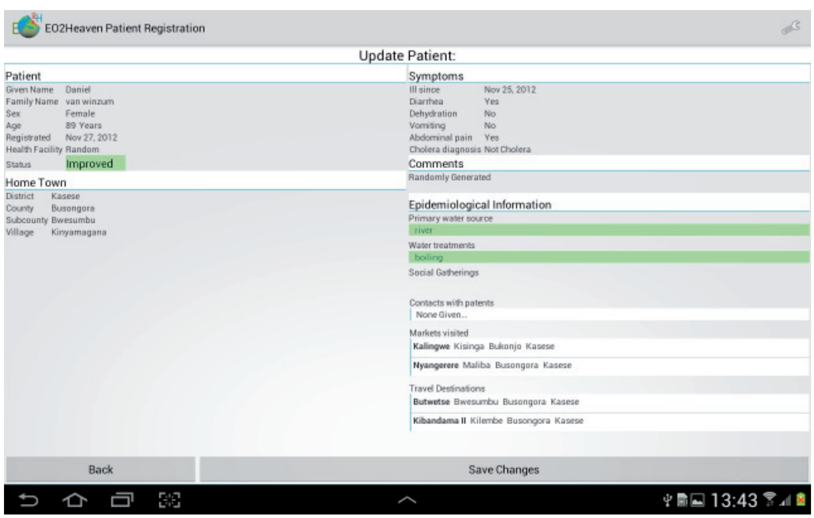
The illustration on the opposite page shows the app start screen. On the left side is the patient list, based on the recorded cases, which are stored on the device. On the righthand side of the app the details of the currently selected case are provided. If a device with a small screen is used, only the case list is displayed and replaced with the case-details screen once a case is selected.

The green and grey icons in the patient list indicate which patients have been synchronized to the server. The status fields of patients that have the status “New” are highlighted in blue.



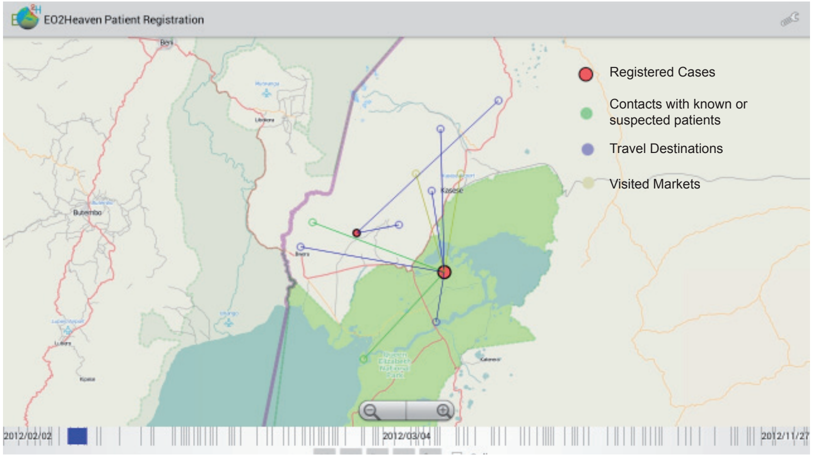
DIRA Start Screen

The basic creation of reports for new patients and updating of the status of patients under treatment is carried out in an intuitive screen dialogue:



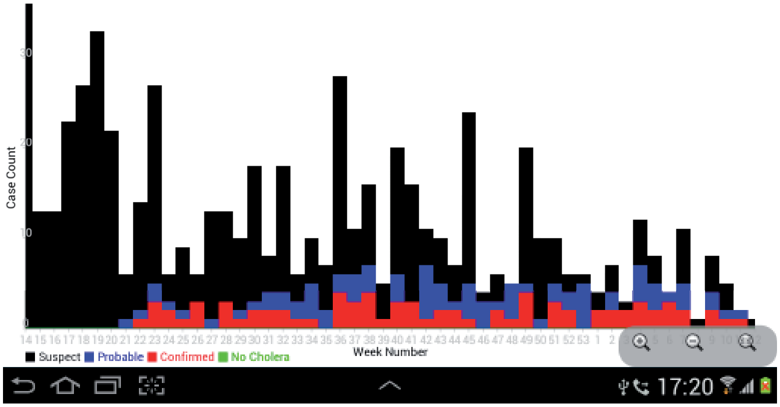
DIRA Patient Data Record

A mapping function of patient movements highlights the towns that a patient visited in the last week with a circle, with a line running to the patient's home town. This is seen as a major improvement compared to the collection of case data on paper, since it provides immediate local information on the spatial dynamics of a cholera outbreaks and thus supports local mitigation activities:



DIRA - Movement of Patients

Registered cholera cases, which are stored on the device, can also be viewed as graph, aggregated per week:



DIRA - Graph of Cholera Cases

8.4 CONCLUSIONS

The flourishing telecommunications markets across Africa have led to a surge in the use of mobile applications for eHealth in many African countries. Throughout East Africa the use of mobile technologies like mTRAC, a SMS system for health surveillance, have been adapted or are being implemented. As a result the national health care structures in Uganda are flooded with new tools and applications, developed by various development agencies, which are adapted to specific health programs or problems. In reaction to the rapid growth in the number of applications MoH-National Resource Centre has endorsed a policy to regulated the implementation of new mobile systems in Uganda. In 2012 they endorsed the mTRAC system in Uganda, as the national system to enable mobile data upload into the DHIS (http://dhis2.org/) for data management, processing and reporting.

As evident from this observation, the tools and system resented by EO2HEAVEN are not new in the sense that they rely on mobile technology to address a specific health issue. The different purpose of the EO2HEAVEN tools as compared to the mTRAC/DHIS system lays in the fact that the EO2HEAVEN tools aim to enhance data quality and to accelerate the reporting process to allow timely response as compared to the regular monitoring activities facilitated by the latter. This added benefit can, however, only be exploited if the system presented by EO2HEAVEN is compatible with the existing Health Information Infrastructure in Uganda. During workshop in Kampala this issue was raised by several participants and was supported by representatives from the MoH. It can therefore be concluded that, in order for these tools to become endorsed and adapted by the MoH, the information recorded and accessed should be interoperable with the existing national surveillance system (e.g. DHIS).

The introduction of a patient based surveillance system for the timely identification of an impending outbreak is widely appreciated health managers. The epidemiological data which is routinely generated in a various vertical disease surveillance programs are not fully analyzed or utilized at the district level. Instead, they are directly forwarded to higher levels, hardly providing any feedback to the lower levels. Real time data access and reporting of standardised patient based data over the whole vertical range of the health system can hence benefit the use and accessibility of this information to various end-users. The existence of a system facilitating solely the registration of cholera patients, however, would limit the extend of use by health workers and hence reduce the relative benefit and ultimately the success of the system. Extending the tools towards the framework for Integrated Disease Surveillance and Response (IDSR), as outlined in the International Health Regulations (IHR, 2005), facilitating the surveillance of prioritized diseases and health services, could reassure the realization of a widely endorsed and interoperable health information system for surveillance and disease control.

9 Disease Mapping and Spatial Analysis for Disease Control

9.1 MAPPING FOR PUBLIC HEALTH

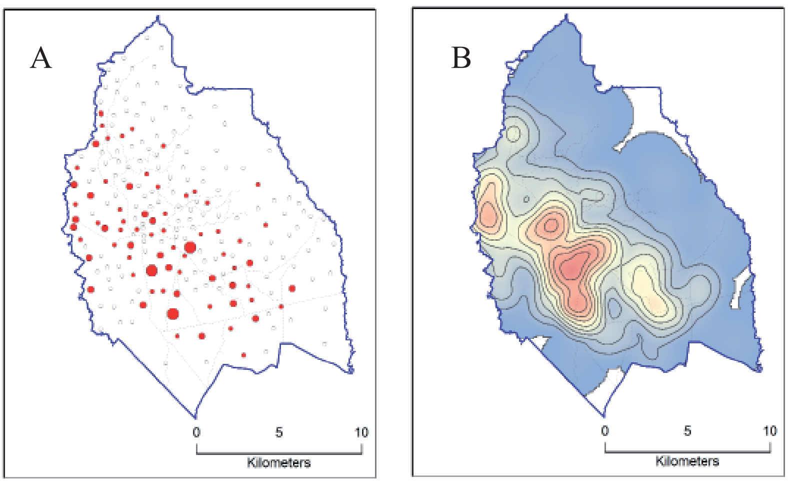
One of the earliest examples of disease mapping in public health was presented by John Snow in 1854. He used maps to trace the occurrence of cholera cases back to a water pump contaminated with cholera. Since then, epidemiologists, public health professionals and medical geographers have advanced the application of geospatial techniques to investigate associations between the public health and the environment. Recently published applications of GIS have shown the value of spatial analysis in the surveillance and monitoring of various diseases (Pfeiffer et al., 2008, Clements et al., 2006), in environmental health modelling exposure external risk factors quantifying lead hazards in a neighbourhood (Root, 2012) and the analysis of disease policy and planning (Rushton, 2003).

The main motivation of the geovisualization process is to uncover unseen patterns in increasingly large and complex datasets, which are otherwise unlikely to be discovered (Orford, 2005). By processing data using various statistical techniques facilitating interpretation of possible trends and patterns the full scope and value of data can be investigated and utilized. Here, the use of GIS and disease mapping in public health is illustrated by a case study conducted in Kasese, Uganda as part of the EC funded EO2HEAVEN project. In this case study, GIS and spatial epidemiologic methods were combined to identify and locate environmental risk factors associated with diarrheal disease resulting from infections with cholera (*Vibrio cholerae*). Based on three core function of environmental health it is demonstrated how GIS and disease mapping techniques, can be used to:

- assess the burden of disease and for risk assessment,
- assess health system coverage and health assurance and
- develop guidelines for decision support, advocacy and policy development.

Mapping the Burden of Disease 9.1.1

Disease mapping and exploratory spatial analysis and techniques allow identifying anomalies in the spatial pattern of disease occurrence. Hence exploratory methods are particularly valuable as a first step to visualize the burden of disease in searching for zones or districts of high disease prevalence. The following illustration shows the spatial distribution of cholera cases per village in an area highly affected by cholera in Kasese over the year 2012:



Spatial point pattern showing a gradual point representation of the observed number of cholera outbreaks per village (A) and a spatial kernel smooth of the same cholera pattern (B) in a heavily affected area over the year 2012 in Kasese district, Uganda

The point representation as shown in map A allows to develop a detailed representation of the health situation and to obtain good general impression of the distribution of cholera infections across the area. Yet, when trying to identify which areas are most impacted or prone to a disease other mapping techniques should be applied. Map B shows the same cholera data as a continuous surface obtained by applying a kernel smoothing technique to the point data. The generalization of cholera point pattern by applying a smoothing technique allows to directly identify which areas have the highest burden of cholera during a particular outbreak. Such information can be used for the planning of interventions, but can also be compared to the spatial pattern of external risk factors to assess whether the occurrence of disease can be related to variations in the external environment.

Smoothing algorithms are popular and widely used to visualize spatial patterns of disease because the ease of use and interpretation on the results. The use of such algorithms, however, requires a priori knowledge of the spatial scale and nature of the spatial relations over which a smoothing algorithm should be applied. Often the selection of the scale of the spatial smooth is made subjectively and is based on some previous knowledge on the scale of transmission of a disease or the size of an area affected by some external risk factor.

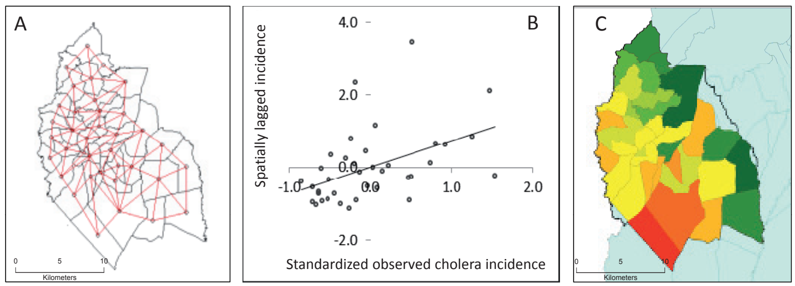
Exploring the spatial pattern and spatial trend can provide more insight into the exact nature of the spatial processes leading to the observed pattern of disease. For example small scale spatial variations in the incidence of cholera found here, could reflect the spatial scale over which a common source or reservoir of cholera results in new infections locally. Similarly, large scale patterns of cholera occurrence over space, and time likewise, could reflect the occurrence and behaviour of cholera in natural reservoirs resulting from climatic and ecological driven processes. These spatial processes should hence be considered when mapping and relating the occurrence of the disease to the behavioural patterns of the host, agent and the surrounding environment.

9.1.2 Spatial Trend Assessment

As was shown previously, many methods for exploratory analysis of disease patterns do not inherently take into account the spatial nature of the disease pattern observed and may assume independence of observations. Since areas may differ greatly in population size, prevalence rates have different levels of variability and reliability. In case of an infectious disease, infections are clearly not independent and the observed incidence of disease is expected to be driven by spatial and temporal processes structured by the adjacency to other affected areas in the neighbourhood. In these situations, one can use various statistical techniques to quantify spatial relations, allowing to explore the spatial and temporal patterns of infectious disease spread. As has been demonstrated, smoothing techniques can be used to generate geographically based regional or local means to which actual rates are smoothed. Yet, these rely on procedures that incorporate proximity as well as population attributes.

By delineating the structure and describing the nature of spatial relations between different regions which ultimately result in the observed pattern of disease, allows quantifying spatial trends and model expected the risk of disease.

The map A below shows a network describing the spatial relations between different administrative units in Kasese, Uganda.



A) Map showing the spatial connectivity between adjacent administrative units (parishes) in an area heavily affected by cholera over the year 2012 in Kasese district, Uganda.
B) Moran's I correlation plot of the area showing the positive relation between the observed incidence of cholera (per 1000 pop) and the spatially lagged, or distance weighed, expected incidence of cholera.
C)The expected incidence of cholera per parish based on the spatially corrected smooth of the observed incidences (Local Bayesian Smooth)

The observed incidence of cholera in the vicinity of an administrative unit was used to estimate the expected incidence of cholera in an area during a cholera outbreak in 2012. The expected risk of cholera infections of each unit was determined based on the observed risk corrected with the risk observed in the surrounding units. Here, an iterative approach was used to determine the optimal risk for each unit, resulting in the spatially lagged or expected incidence based on the surrounding shown in the Moran's I correlation plot (B) above. This analysis showed that, not taking into account other external determinants of cholera incidence, 72% of the observed cases in a certain unit could be accounted for by the observed number of cases in the direct vicinity of an area as illustrated in map c above. This finding has strong implications for the mitigation of cholera outbreaks in this area. When the risk of infections in an area highly depends on the occurrence of cholera in the surrounding areas, interventions should take into account the relative risk of infections in a wider area.

9.1.3 Health Risk Assessment

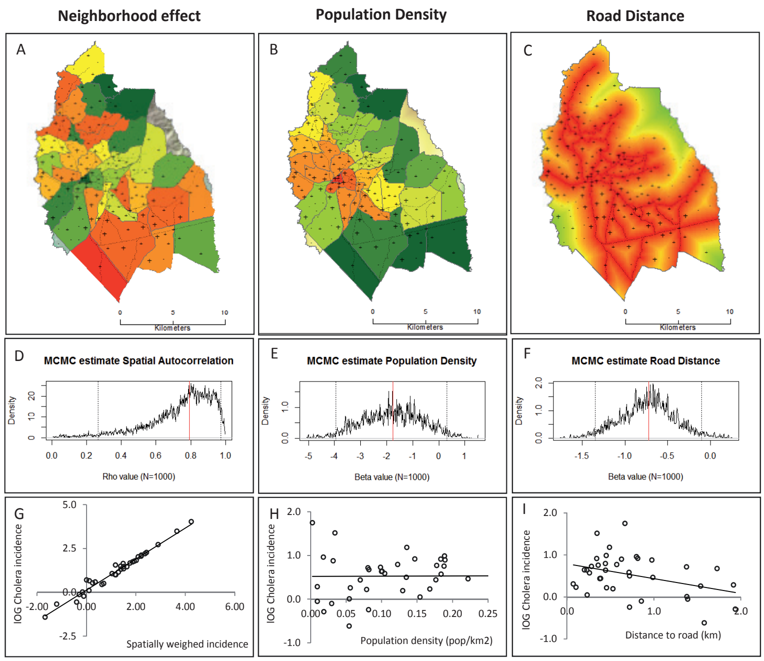
Mapping and estimating the intrinsic clustering of disease over space has many usefull applications when assessing the burden of disease over space and analyzing trends in the spread of disease. Even more, disregarding spatial relations when quantifying the risk of disease, imposed by an external risk factor, can lead to false conclusions. As a sample of cases of an infectious disease might be expected to be clustered in space they cannot be regarded to represent a random, independent sample. Ignoring spatial dependencies could therefore lead to a violation of the statistical assumption of independence. Consequently, disease clustering would lead to a higher occurrence of disease cases as expected based on an underlying risk factor and a bias in statistical inferences (e.g. false positives or type I error).

The use of geostatistical mapping for risk assessement is illustrated on the previous page. Here, three risk factors were mapped to an administrative area (parish) and compared to the incidence of cholera in each parish. To correct for possible spatial dependencies, the ratio between the locally observed incidence and the expected incidence of disease based the spatial surrounding areas were mapped and used to predict the observed incidence of cholera per parish as shown in map A. High positive values indicate that the observed incidence rates were higher than the observed incidences in surrounding parishes and vice versa. This information was then used to quantify and adjust the observed incidence of disease allowing to quantify the spatial denpendencies between neighbouring diseased areas though transmission and dispersion of the disease itself.

Equivalent to the spatial interdependencies of cholera occurrence, the effects of population denisty and the average distance to roads on cholera incidences were analyzed as shown in the Moran's I correlation plot and map C. Visual inspection of these maps can povide a rough impression of the possible relation between the environment might affect and the occurrence of cholera. Moreover, various statistical optimization techniques can be used esitimate these risks (Bivand, 2008). Here, an itterative Bayesian approach was used to estimate and map the relative effects of the three risk factors, e.g. spatial dependencies, population density, average road distance.

The density plots D-F overleaf of the risk estimation for each of the three risk factors are corresponding to the partial risk plots G-I using Monte Carlo optimization (Congdon, 2006). The results of this analysis suggests a strong spatial effect of cholera occurrence in the neighbouhood corresponding to the spatial trend analysis found previously.

After accounting for these spatial dependencies no significant relation was found between the population density and the relative incidence of cholera as shown in figures E/H. The average distance to roads was found to negatively affect the incidence of cholera suggesting a protective effect of increasing distance to roads (figure F/I):



Gradual color maps showing the spatial pattern of risk factors:
A) The ratio between the local incidence and neighboring incidences of cholera,
B) The population density (red=high, green=low) and
C) The distance to the nearest asphalted road (red=near, green=far)
in an area heavily affected by cholera over the year 2012 in Kasese district.
D-F) Density plots of the risk evaluation of each risk factor showing the median (red line) and 95% CI values (grey dashed lines) of 1000 MCMC trials after an initial burn-in period of 2000 iterations.
G-I) Partial effect plots of each risk factor relative to the observed incidence of cholera (natural log transformed).

These results show that using geostatistical mapping and analytical techniques can be used to quantify the risk of cholera incorporating spatial dependencies. Infectious diseases, like cholera, often inhibit intrisic spatial dependencies which could potentially obscure or suggest a (false) disease mitigation planning as targeted interventions are based on false assumptions.

Mapping data and carefully interpreting the observed patterns by comparing various risk factors over various spatial scales can prevent such errors to occur. Therefore, geostatistical mapping techniques provide some powerful tools for health managers to critically assess a ground situation. Even when data is missing or incomplete mapping can be used to identify these voids and can aid in the effective surveillance and survey planning (Waller, 2004).

9.2 HEALTH ASSURANCE

Another application of geospatial mapping for public health is rooted in the field of healthcare geography and health assurance. This specific core function of public health concentrates on the planning, implementation and use of healthcare services, as opposed to the field of spatial epidemiology which has a strong focus on the actual occurrence of disease per se. Health care geography aims to identify spatial patterns in the need and provision of healthcare services and providing a foundation for the performance of national disease programs. Monitoring and evaluation of national disease programs are set up to provide health policy makers and program coordinators with an comprehensive overview of the impact of health programs and the functioning of health systems. The use of GIS and geospatial mapping provides these with a novel collection of analytical and geovisualization techniques to identity patterns and pitfall in the implementation of the program and to plan supervisions and interventions concordantly.

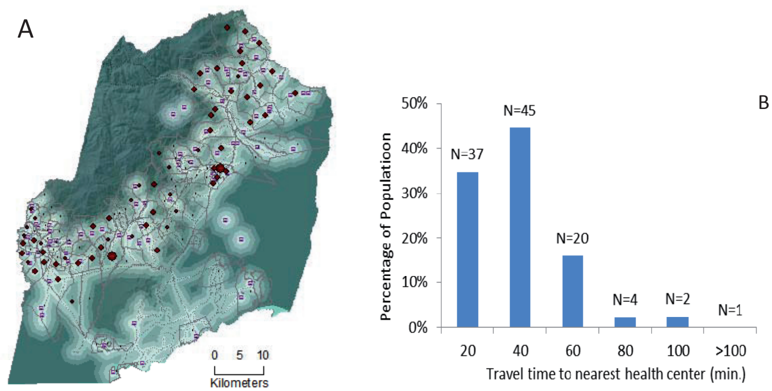
The assurance of health services to a population requiring these services depends on different perspectives set out in local, national or international guidelines and policies. Hence geospatial mapping of health services and health utilization could provide supporting information to develop policies dealing with the equality of services through an equal provision of health services to different regions or the equity of services providing health care to specific entities or groups are addressed. Two core functions that can be addressed with geospatial maps are:

- to link people to needed personal health services and assure the provision of health care when otherwise unavailable and
- to evaluate effectiveness, accessibility, and quality of personal and population-based health services.

Two examples of how this can be done are provided in the next sections.

SYSTEM COVERAGE 9.3

To illustrate how health system coverage can be mapped and quantified, locational data of health centres in the Kasese district in Uganda and population data aggregated to second level administrative units (parishes) were obtained from the Ugandan Bureau of Statistics (UBoS) and plotted. To compare the coverage of the health centre network the travel times to the nearest health centre were used. Travel times were calculated taking into account different travel speeds depending on the travel by roads or off-road and incorporating increased travel times due to topography. The results were then compared to the raw population counts per parish and the percentage of population under a certain travel time threshold were calculated.

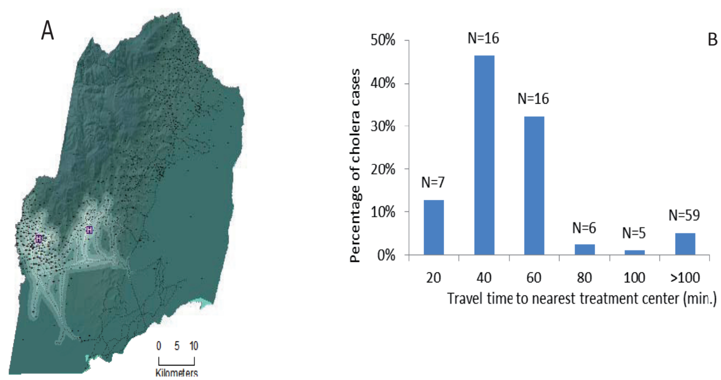


A) Map showing the expected travel times to the nearest health facility (light=low, dark=high) as well as the distribution of the population living in a parish (red dots, sizes represent percentage of the total population).
B) Percentage of the total population living within a certain travel time to a nearest health facility.

The results showed that the majority of the population in Kasese district (>95%) are expected to live within 60 minutes travel time of a health facility (HCIII or higher). Moreover the distribution of the population along the range of travel times implied that the majority of people living in the district have a travel time less than the average over the whole district. This is probably caused by the fact that a small proportion of the population live in remote, mountainous areas with little infrastructure and long travel times as a consequence. Further inspection of the map shows that small settlements along the south of the district have a good coverage of health services even though not many people live in these areas.

9.4 ACCESS TO HEALTH SERVICES

In contrast to the approach used to assess the coverage of the health system as a whole, access to health services can be mapped to quantify the access to health services. Here the number of cholera cases occurring in Kasese district in 2012 was compared to the average travel time to the nearest cholera treatment centre:



Average travel time to the nearest cholera treatment facility (N=2) in Kasese district.

From the map it appears that the majority of the district is insufficiently covered by the treatment services with >80% of the district having travel times of more than one hour corresponding to 64% of the total district population. The delivery of services appears to be highly unequally distributed over the population at risk. Yet, comparing the travel times to the occurrence of cholera shows that >90% of the cholera cases registered in 2012 occurred in the area of less than one hour travel distance.

This example shows that even though the access to services appears to be inequitable in the delivery of services, the limited resources seem to be effectively distributed according to the actual burden of disease. Hence, geovisualization and geospatial analysis of the distribution of services can provide information supporting policies or providing tools for decision support allowing to evaluate and update existing health systems and services.

Using Earth Oberservation Data 10

The science of remote sensing is characterized by the opportunity to obtain information about objects, areas or phenomenon without a direct contact to them. For this task different kind of devices or platforms are available. The most common sensors are mounted on aircrafts or satellite platforms. The fundamental principle of acquisition of information is based on the interaction of electromagnetic radiation with different objects e.g. the earth surface. This electromagnetic radiation however is primarily known as visible light, as the human eye is sensitive to this part of the solar spectrum. Modern recording techniques allow also the use of other wavelength ranges, such as near infrared or thermal radiation and a number of different sensors are available for this.

The major advantage of remote sensing products (e.g. satellite imagery) is the aerial character of data recording. This character is described by means of four parameters or more specifically four different kinds of resolution. To enable the joint use of remote sensing with in situ and health data, the following requirements need to be fulfilled:

Geometric Resolution

The smallest object that can be resolved at the level of the Earth surface defines the geometric or spatial resolution of a remote sensing system. This value is limited by the size of a pixel. For this reason it seems to be more helpful to use systems with a high geometric/spatial resolution (e.g. QuickBird: app. 0,6m panchromatic). However, the determination of air pollution parameter usually demands specialized data-products like the MODIS-AOD-product (Terra/Aqua). But its geometric/spatial resolution is much less (factor: 10) than the original MODIS-product.

Radiometric Resolution

The radiometric resolution of a remote sensing system is defined as the smallest difference of the radiance that can be detected from the system. The most important factor for a high radio-metric resolution is the signal to noise ratio of the systems. In line with the requirements for EO2HEAVEN a preferably high radiometric resolution is recommended.

Spectral Resolution

The spectral resolution of a remote sensing system is characterized by two parameters, the number of spectral channels as well as their bandwidth. For the use in the context of air pollution parameters it is not important how many channels or spectral bands a system offers but the position of a specific channel in the solar spectrum and relative to the absorption band of parameters like O_3 or NO_x. As one example the MODIS-AOD-product is based on the calculation of three specific spectral MODIS-channels.

Temporal Resolution

The temporal resolution is defined as the time between two overpasses of a remote sensing system. Typical temporal resolutions are between 12 hours (day and night overpass) and a few weeks. It must be mentioned that the overpass interval of a system is only a theoretical value of the system in terms of the availability of satellite data because the usability in mid-latitudes regions is strongly affected by cloud coverage.

With the help of Earth observation satellites a massive quantity of diverse information about our planet is being gathered continuously. Various recording instruments measure and photograph the Earth transmitting information on many different environmental parameters, which may have a direct or indirect impact on health.

Direct threats to health, such as atmospheric particulates and reactive gases, and marine pollution can all be monitored in real time. Satellites also provide meteorological data used to predict extreme weather events. For such application a lot of secondary data products (Level 2 products) derived from the origin raw data are usable. Thus, it is for example possible to calculate information about the aerosol optical depth from MODIS-data (Moderate Resolution Imaging Spectroradiometer). For products like these some data providers are available. The most important agencies are the National Aeronautics and Space Administration (NASA) as well as the European Space Agency (ESA). However, also national institutions like the German Aerospace Center (DLR, Germany), the Centre national d'études spatiales (CNES, France) or co-operations between commercial companies and enterprises could offer interesting data for health applications.

REMOTE SENSING DATA MODELLING 10.1

By means of remote sensing data it is possible to extract information about the composite of the atmosphere. Besides permanent trace gases like O_3, CO_2 or SO_2 primarily the emission of particular matter (PM10) was in the main focus of the research in EO2HEAVEN. The realized approach is based on the assumption that there is a significant influence of such particles to the scattering of sunlight during their transmission through the atmosphere. This effect could use as pointer for the present of different particles in the Atmosphere. An established way of using remote sensing data for the determination of PM10 is the use of MODIS AOD-data. These data are provided by the NASA free of charge. It is important to keep in mind that it is not possible to measure PM10 pollutions directly. For this reason it's necessary to find a link between the actual PM10 values observed near the surface and level 2 products based on satellite measurements. For ground truth data the corresponding measurements of the Saxony Insitu-Station network were used.

For this modelling different steps of data processing have to be executed. With the help of the AERONET data it is possible to exclude the influence of different land use and land cover. The predicted result is an over- or underestimated AOD value depending on the reflectance of the terrain and season. This analysis scheme is only an optional step. In the framework of the project EO2HEAVEN we did not use additional data from AERONET.

More important than the calibration of the AOD values by means of AERONET is the consideration of meteorological parameters like the boundary layer height (BLH) or the relative humidity (RH). With the BLH it is possible to consider the vertical variability of the aerosols during the comparison of measurements at different times. The second parameter (RH), describes the vertical distribution of the relative humidity at different pressure levels.

A source for these two meteorological parameters is the European Center of Midrange Weather Forecast (ECMWF).

BLH-data can be downloaded free of charge from the ECMWF. However, within the scope of the project it was unfortunately not possible to get direct access to the RH data of the ECMWF, because this has to be organised through the national meteorological agency.

In the analysis various interpolation methods can be used. This is necessary because the processing of three measurements requires spatial and temporal adjustments. In the simple model the comparison is made on the basis of the nearest neighbour interpolation method. Thereby each of the in-situ measurement is compared with the MODIS AOD which is the nearest one. Due to the geometrical resolution of the MODIS AOD data, spatial differences of up to 5 km arise. By including the boundary layer height (physical model), the interpolation of an additional input parameter is required.

- Nearest Neighbour
  The in-situ measurement is compared with the closest corrected AOD value. This AOD-value is corrected with the closest boundary layer height. The in-situ measurement itself is interpolated to the time of the satellite's overpass.
- Space
  The in-situ measurement is compared with the closest corrected AOD value. The transformed AOD-value is corrected by means of a related to its position bilinear interpolated boundary layer height. The in-situ measurement itself is interpolated to the time of the satellite's overpass.
- Space and Time
  The in-situ measurement is compared with the closest corrected AOD value. The transformed AOD-value is corrected by means of a related to its position bilinear interpolated boundary layer height. The in-situ measurement itself is interpolated to the time of the satellite's overpass. Also a temporal interpolation of boundary layer height to the time of the satellite overpass was done.

The temporal interpolation was performed by a linear approach for the in-situ measurements and the BLH-values of the ECMWF. Here, the in-situ data available for analysis in 2006 in a half-hour resolution were interpolated within this half hour. The analysis of the data shows that the values of the boundary layer height (temporal resolution of 3 hours) may have an extremely high variability.

For the consideration of the relative humidity (RH) some additional information and parameterizations were required. For this and for a detailed description of the underlying models the reader is referred to the public document D3.14. In the project various models and configurations were tested. A Pre-test of modelling by means of the simple linear regression model shows that the quality of the model parameters (independent from the background of the stations or the quality of the MODIS-AOD) is unsatisfactory. The results of the simple model indicate that the use of such an approach is not always possible. The Pearson correlation points out only a weak to moderate linear relationship. At the same time, the standard deviation with over 21 $\mu g / m^3$ is very high. The distinction between *Terra* and *Aqua* data does not lead to any changes. The criteria of the standard deviation and the correlation values are in the same range and do not lead to any improvement of the linear model. Such improvement seems only to be possible if the analysis of data is carried out according to the background of the station and for each individual station separated for the Terra and Aqua satellites, but regardless of the quality of MODIS AOD. Six different categories with a different number of in-situ-stations are available:

Category 1:	urban background	9 stations
Category 2:	rural background	2 stations
Category 3:	rural regional	2 stations
Category 4:	rural suburban	1 station
Category 5:	suburban	1 station
Category 6:	urban traffic	10 stations

Correlations up to 0.7 for data from the Terra-satellite indicate a significant linear relationship between the measurements. Unfortunately the variability of the statistical parameters as well as the model parameters of several stations is comparatively high. For the simple model introduced above it should be noted that analysis of the measured values according to categories and sensors already has a significant improvement of the statistical spread to the resulting linear model. Of the initial 22 $\mu g/m^3$ (standard deviation), this approach can be expected values between 8 and 13 $\mu g/m^3$. Based on this result, it seems to be necessary to improve the model by consideration of meteorological parameters as well as a more robust physical approach. For this a simple algorithm has been used which includes two iterations. The first calculation of the model is followed by an exclusion of measured values that have a deviation of more than two and a half times the standard deviation. Afterwards the parameters of the model and the statistics are recalculated:

- Analysis of all data for all stations, both satellite systems and independent of the quality of the MODIS AOD for different interpolation methods
- Analysis of all data for all stations, separately for the satellite systems and independent of the quality of the MODIS-AOD for different interpolation methods
- Analysis for every individual station (and background), separately for the satellite systems and dependent of the quality of the MODIS-AOD for different interpolation methods and both kind of algorithm.

The results show that the robust approach leads to an improvement of the standard deviation and the correlation. It also becomes clear, that the interpolation method has a significant influence to the model parameter (a). The gain (slope) of the model arises by factor two between the nearest-neighbour- and the space-interpolation.

The second step distinguishes between the data of the two satellite systems. It is striking that the correlation is a little lower than that of the joint analysis of different satellite data. This is not confirmed for the Terra data. By means of the robust algorithm the correlation arises up to 0,5, a threshold to a clear linear relationship. As in the first case, the influence of the interpolation method especially for the Terradata is comparatively high.

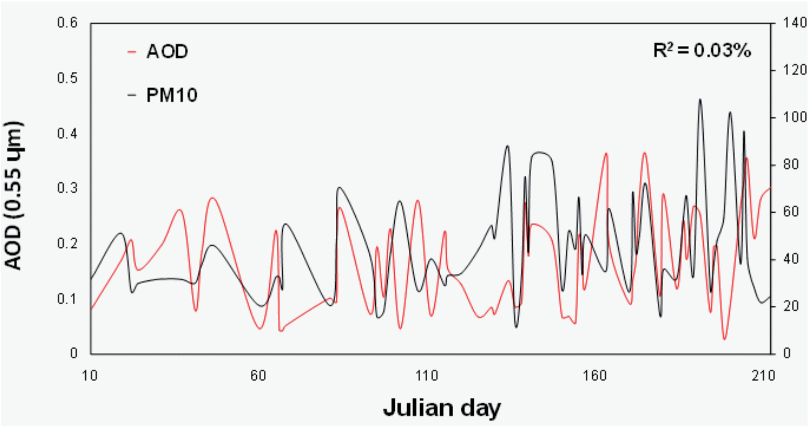
For the reason that at this particular time of analysis, no clear connections could be detected between the model parameters, the statistical parameters and the model configuration, it was necessary to analyse the different configurations of the model in all combinations.

- Two satellites (Terra / Aqua)
- Three interpolation methods (nearest neighbour / space / space and time)
- Two algorithms (all data / robust)
- Four MODIS-AOD-Qualities (≥QA0 / ≥QA1 / ≥QA2 / QA3)

These boundary conditions allow analysing of 48 combinations / configurations. The visualization of these results is difficult and comprehensive and described in detail in the public EO2HEAVEN Deliverable 3.14. Finally the analysis of all stations shows, that the behaviour of the parameters and statistical criteria dependently the described configuration of the model is very heterogeneous. Currently, there is no possibility for a final judgment.

Nevertheless, there are a few basic recommendations. First of all it should be mentioned that the geometrical resolution of the MODIS-AOD data is considerably worse than the activity area of the in-situ stations. It can be assumed that an improvement of the satellite technology, as well as of the algorithms for the determination of the AOD, will lead to a better modelling of remote sensing and insitu-data. Next to that one limiting factor will be an obstacle independent from technological developments. The influence of cloud coverage in mid-latitude regions to the availability or quality of remote sensing data is important. Analyses of the year 2006 show that only app. 8 % of the theoretical available MODIS AOD data in the case study of Saxony have a quality level of QA3 (cloud free).

For the Durban case study, AOD data was extracted from the MODIS aerosol data for 2010. PM10 data was used from the Wentworth air quality monitoring station. The approach used was a point to pixel comparison. It was expected that observations made from this analysis would yield considerations for future analysis. Hourly correlations between MODIS AOD with PM10 were weak as illustrated below:



Trends of AOD over South Durban and In situ PM10 measured at Wentworth

MODIS AOD did not show correlation for all days in the year. The degree of correlation fluctuated greatly day to day. The correlation is insufficient as an input to any air quality or health risk model as there is no clear trend between AOD and PM10.

MODIS AOD had generally good performance at global levels but performance varied greatly at finer resolutions where its performance in urban areas was poor (Hyer et al. 2010). Previous work had successfully used a modified aerosol retrieval algorithm to extract aerosol data at high spatial resolution for example Kumar et al. (2008), in order to correlation to PM10.

For MODIS AOD to be of use, the level of uncertainty propagated by the relationship of AOD with PM10 needed to be small. The high uncertainty related to untransformed MODIS AOD over South Durban, was unsuitable as an input for an air dispersion model. Related to this, Mölter et al. (2010) noted the uncertainty related to the input modelled data from an air dispersion model, had a significant impact on the uncertainty from the results of the land use regression model used.

The South Durban case study targeted acute health exposure at the individual level. Explicit use of AOD data for exposure is linked to population level exposure and not individual level exposure. For example, Hu (2009) used AOD and PM2.5 to create estimated PM2.5 surfaces for the whole of the United States. This was used as input to model the relationship of population exposure to PM2.5 with heart disease. The results could not be disaggregated to understand individual exposure (Hu, 2009).

Thus, the MODIS AOD is unsuitable to be used as a proxy for PM10 in South Durban. The product is provided only at medium resolution. The eThekwini monitoring network is not spatially complementary at this resolution. Performance of the MODIS AOD at the intra-urban scale is highly uncertain. The MODIS aerosol product generally has links to population and not individual health exposure due to the level of spatial aggregation used. With consideration of the specifications of the South Durban Case study, the use of MODIS AOD in its untransformed provision was deemed inappropriate.

10.2 STUDY ON USING UAVS FOR GATHERING EO DATA

One of the most important phenomena of urban climatology is the manifestation of the so-called heat islands. That means that the climate of an urban area stands out clearly from the surrounding area. This is particularly evident on hot windless summer days. Especially older and vulnerable people suffer from these conditions.

One possibility for the determination of the urban climate is the calculation of surface temperatures from thermal images. This is based on the physical principle that each object emits thermal radiation based on its surface temperature. Thus, technically only radiation temperatures can be captured by thermal cameras. With knowledge about emissivity values, surface temperatures can be deduced. So far, thermal maps come from satellite images or airborne flight missions. The advantage of satellite data is the almost simultaneous capture of the whole urban area. However, at which time the images are taken cannot be influenced. Furthermore, the low spatial resolution (e.g. 90m Aster, 120m Landsat 4/5, 60m Landsat 7) is a limiting factor. Concerning this, plane-based methods are appropriate, however, involved with high costs.

The current developments in the field of unmanned aerial vehicles (UAV) allow us to collect data for applications at a local scale. Equipped with a thermal camera, UAVs can be used for studying micro urban heat islands. With UAVs a high spatial and temporal resolution is feasible, adapted to the user requirements. Compared to satellite techniques, UAVs have no orbital restrictions and the flight path is easy to manipulate. Due to the low altitude, clouds are no problem. Furthermore, there are differences in data acquisition and data processing. This is due a number of facts: compared to airplanes, the platform is more instable and susceptible to interferences, e.g. blurring caused by wind or rotor movement. Also smaller and mostly more inaccurate sensors have to be chosen because of payload limitations. In addition, the smaller scale and the special character of thermal images have to be considered.

A UAV is only one part of an overall system with lots of sensors. To navigate a UAV also a ground station is needed. This enables waypoint navigation, but also intervention should be possible. The overall system is called Unmanned Aerial System (UAS). For our study, the octocopter called HORUS (Hovering Remote controlled Ultra-light Sensor platform) was used which was developed by the Fraunhofer Institute for Transportation and Infrastructure Systems IVI .

The ground station consists of the following components:

- Computer for flight planning and monitoring
- Remote control to transfer the mission parameters
- Live monitor to assess the recordings of the UAV
- Standby generator for powering the base
- An antenna module with tracking to receive data from the UAV and transfer to the base

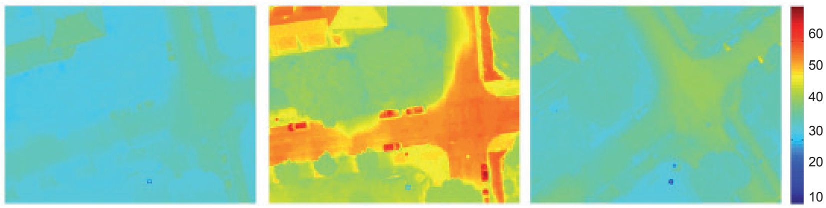
The unmanned aerial vehicle (UAV) itself has a dimension of 1300 mm x 1000 mm x 80 mm, weighs 1800 g and consists of the following main components:

- GPS (Global Positioning System)
- IMU (Inertial Measurement Unit)
- Telemetry
- Remote control unit
- Video downstream
- Cameras and other sensors

The UAV was equipped with the calibrated hand-held thermal camera InfraTec mobileIR E9 and the small and light RGB camera GoPro HD Hero.

Within the project, an urban area of 260 x 200 m was recorded. The area is part of the TU Dresden campus, containing different kinds of buildings, streets and vegetation. To consider diurnal temperature variations, three flight missions were carried out at one summer day (30.06.2012): one in the morning (7 am), one at noon (2 pm) and one in the evening (9 pm).

The following illustration shows the areal temperature distribution of a road junction in Dresden at different times of day (without local emissivity correction):



7:00 14:00 21:00

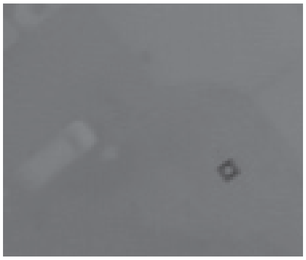
The aim was to generate a thermal orthophoto, which represents the captured radiation temperatures with the geometry of a map. Therefore, in the planning and realisation of the flights some things had to be considered to ensure the complete coverage of the area and the intended geometrical resolution of 15 cm, including:

- Camera specifications (focal length, image size (pixel), pixel size, frame rate of videos)
- UAS specifications (battery service life, speed, visual flight)
- Characteristics of the test area (size and terrain slope)
- Legal regulations

The illustration on the following page shows the planned flight path. During the flight, the GPS positions and orientations of the copter and some other information are logged and stored in a GPX file. However, the accuracy is too low for direct geo-referencing. Thus, the aero-triangulation has to be performed to orientate and geo-reference the image block. Subsequently, products like digital terrain models (DTM) and orthophotos can be derived. Nevertheless the logged positions and orientations act as approximate values. A main component of aero-triangulation is the definition of control points. Control points are discrete natural or artificial points which should be distributed around the whole test area and which positions are determined by geodetic methods. In the aero-triangulation, these known points stable the image block and allow the transformation of the image block to the reference system.

Planned flight path (red) and positions of control points (green triangles)
data source orthophoto: ATKIS®-DOP © Staatsbetrieb Geobasisinformation und Vermessung Sachsen 2013 (GeoSN)

In the course of designing the control points (defining size and composition), the challenge was that they had to be visible in the thermal as well as in the RGB images from 150 m flight height. Consecutively, around the test area, eight control points were arranged and their positions were determined by dGPS. To realize the visibility in the thermal image data from 150 m height, the control points were designed using a 1 x 1 m flake board which was glued with crinkled aluminum foil. The centre was marked with a black paper circle of 45 cm diameter. As aluminum reflects the cold sky radiation, the control points appear dark with a brighter circle in the image data and stand out against their surroundings:



Control points (left: RGB image; right: thermal image from 150 m height)

The following steps were involved in creating a thermal orthophoto based on the acquired data:

- Checking the acquired data (the compliance of the flight parameters, the quality and completeness of the videos itself and the visibility of control points within the videos)
- (Semi-) automatic synchronisation of the flight trajectory with the image data
- Automatic extraction of each flight strip
- Automatic reduction of the image data specifying a longitudinal overlap
- Colour-coded visualisation of the raw data
- Aero-triangualation
- Digital terrain model generation

A more detailed description of the methods is shown in Pech et al. 2013 and the public EO2HEAVEN document "Environmental Monitoring for Health Applications", which is available on the EO2HEAVEN website.

EO2HEAVEN Software 11

TIME2MAPS 11.1

For the visualization of available datasets, a web application for the interactive mapping and display of spatial data and additional context information called Time2Maps has been implemented. Apart from the visualization of static data, the user can also explore maps provided by time-aware services to examine the temporal progress of measured data. The application has been deployed in two of the project studies:

- To access information on ozone, particulate matter and the modeled load of diseases caused by these environmental parameters in Saxony, Germany
- To access information on cholera cases, as well as prevailing environmental conditions in Kasese, Uganda

In both cases the underlying technology is the same, with adaptations to accommodate the specific datasets and analyses results.

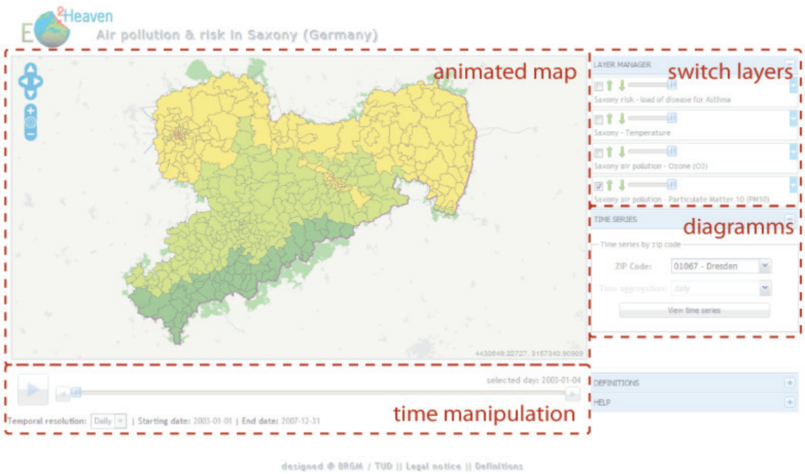
Architecture 11.1.1

The web application consists of three major building blocks:

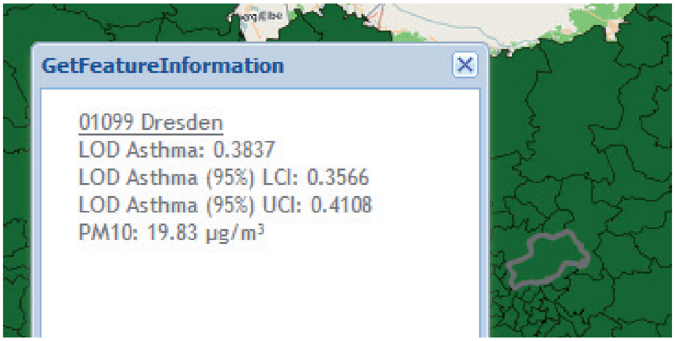
1) On the server-side, spatial data is stored in a spatial database (PostgreSQL with PostGIS extension).
2) Based on this, a UMN-Mapserver offers different web mapping (WMS) and web feature (WFS) services.
3) On the client-side different JavaScript Frameworks are used to build the web application, including:
   - Sencha Ext JS to build the skeletal web app structure (http://sencha.com)
   - Openlayers for creating the map and all interactions within (http://openlayers.org)
   - Highstock for creating the interactive charts (http://highcharts.com)

11.1.2 Functionality

The application primarily consists of a map area and tabs, which allows the user to select the datasets to be accessed and displayed, their legend and other metadata. Each available and selected dataset is listed and options include e.g. the definition of the layer order, drawing styles or layer opacity. For time series data, a time slider clearly indicates the current period displayed and enables the user to view a visualisation for a fixed point in time, or an animation for a time period.

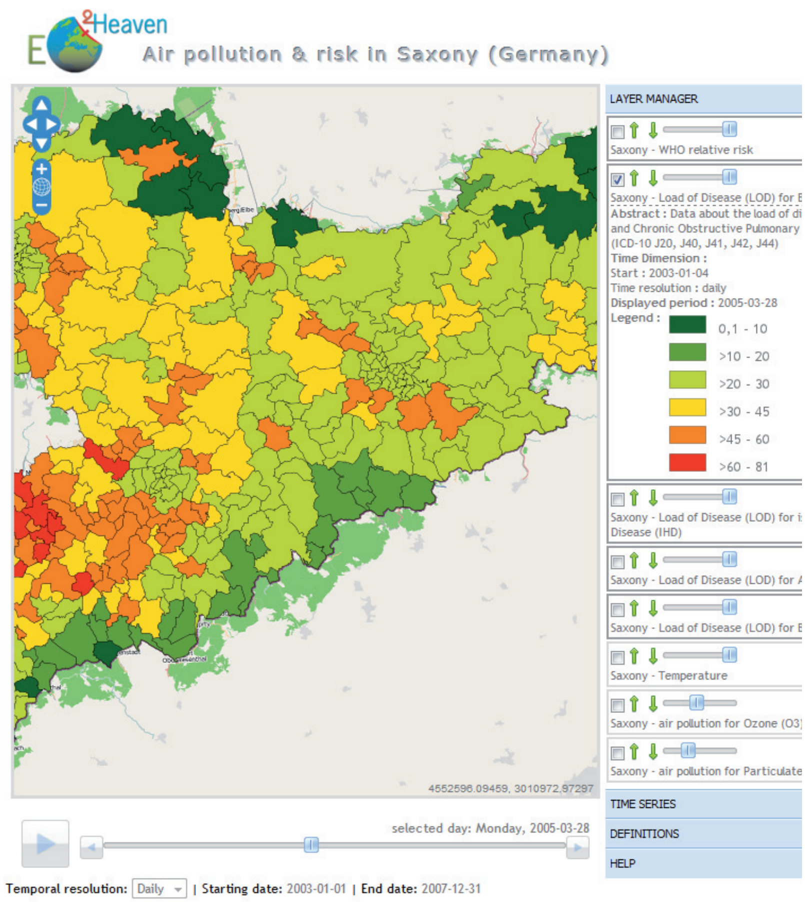


Depending on the selected data, the user can also retrieve additional information, such as statistics or detailed reports by simply clicking on according map features. The following example shows detailed information the of PM10 value and Asthma LOD for a selected day, which are linked to displayed postal code areas:



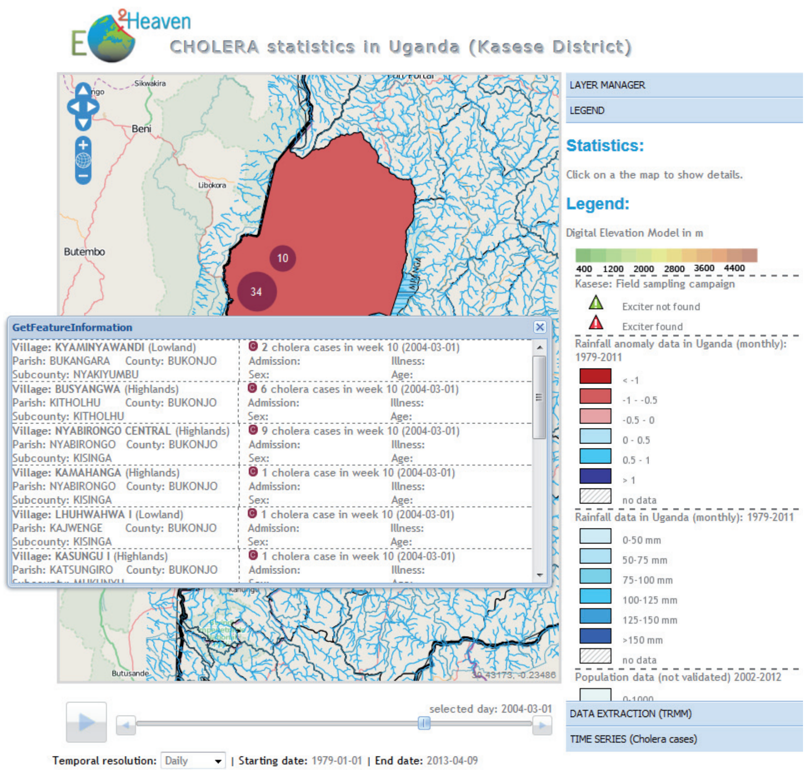
In addition to data visualisation and the retrieval of linked associated information, a data extraction module was implemented for the cholera scenario. This component allows the user to download TRMM datasets for a selected area of interest by drawing a polygon on the map, selecting a time period and the desired file format.

The EO2HEAVEN client implementation for the case study 1 client focuses on the presentation of measured and recorded data of air pollution and resultant the computed risks of respiratory diseases for the selected postal codes:



The web client for the case study on air pollution & risk in Saxony is accessible at the following URL: http://time2maps.dyndns.org/saxony/

The EO2HEAVEN client implementation for the case study 3 client shows the time aggregated total recorded cases of cholera diseases for the selected parish/county:



The web client for the case study on cholera cases in Kasese, Uganda is accessible at the following URL: http://time2maps.dyndns.org/uganda/

11.1.3 Licencing & Availability

This Time2Maps software is an open source project and available under the GPLv3 license for open source projects. For further information on using the client in your own projects, don't hesitate to contact the EO2HEAVEN project team (eo2heaven.org).

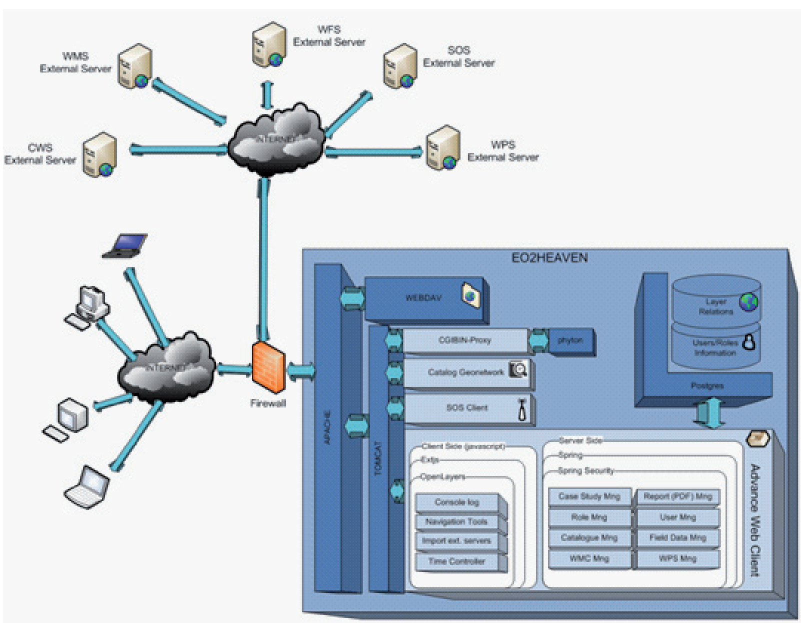
All implementations are available online at: http://time2maps.dyndns.org/

EO2HEAVEN GENERIC ADVANCED CLIENT 11.2

Whereas the Time2Maps application has a primary focus on providing easy to use visualisation and retrieval capabilities, an alternative approach was evaluated in the context of the project. Here, a web client tool was designed to integrate further functionalities, on the fly addition of external layers, visualisation of time series in a video, visualisation of detailed information of the geometries of the layers, saving map contexts, downloading the results, execution of pre-defined web processing services (WPS) and exploring the outputs produced by the other services.

Architecture 11.2.1

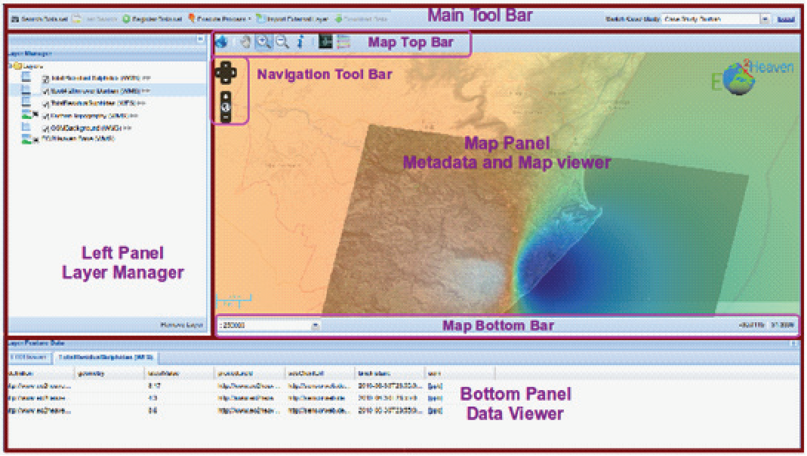
The following illustration outlines the architectural concept and components:



The Extjs and Openlayers frameworks are again used as the base of the client implementation and PostgreSQL with PostGIS extension is used for data storage. The extended functionality allows dynamic access to external map and feature services based on the WFS, WMS, WMS-t and SOS standards of the Open Geospatial Consortium. A cgi-bin Proxy component developed in python enables access to servers in different domains. Web Distributed Authoring and Versioning (WebDAV) is used for data and report sharing.

11.2.2 Functionality

The EO2HEAVEN generic advanced client includes a user management component to support different behaviour of the client application towards individual user groups. Following the entry of the login credentials and the predefined topic of interest, the client loads predetermined layers, e.g. pollution concentration maps (SO_2, PM10 and NO_2), remote sensing images, case data, sampling data, social and geographical data such as population density, land use practices, demographics, topography or water use practices:



The user can add supplementary layers by searching new datasets in the EO2HEAVEN catalogue, recovering the last search in the project catalog or import external layers. Layers from an external WMS, WFS or SOS services can be added to the web client's current map and new datasets can be registered in the project catalogue. The Web Processing Service (WPS) Manager is currently predefined to run the Compute NDVI process. However, new processes can be added by changing a configuration file.

A client with reduced functionality was used in the Durban Case Study, as described in Section 6.5.

11.2.3 Licencing & Availability

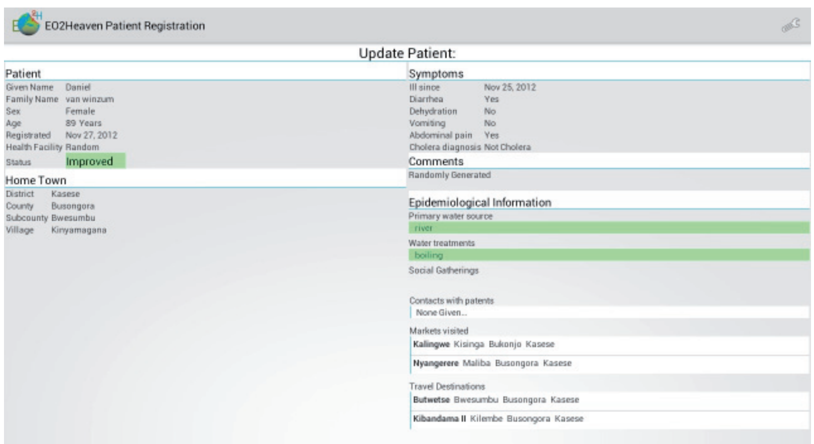
The EO2HEAVEN Generic Advanced Client software is available as open source under the GPLv3 license for open source projects. For further information on using the client in your own projects, don't hesitate to contact the EO2HEAVEN project team (eo2heaven.org).

DIRA - A MOBILE APP TO REGISTER CHOLERA CASES 11.3

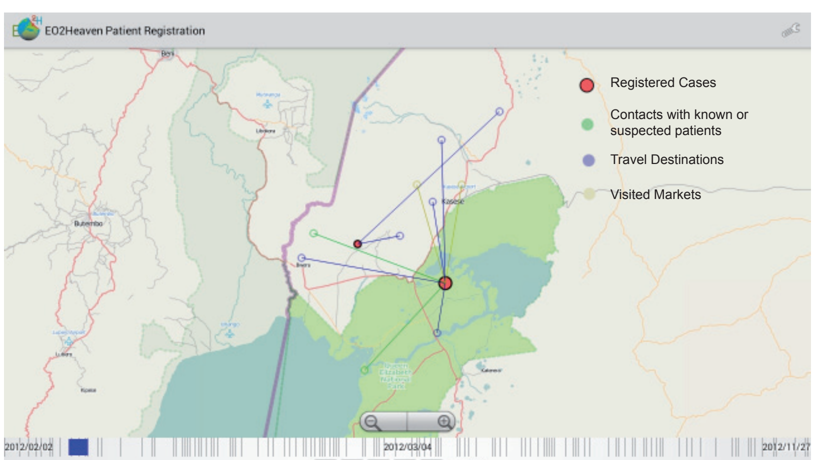
To improve the cholera reporting in Uganda, an application was developed for mobile devices that allows the field registration of patients to be done quickly and easily. The vision is to incorporate this application into an early warning system.

Functionality 11.3.1

The application has an easy-to-use interface for entering patients. It also allows for the collection of epidemiological information, such as the patient's primary water source, and which villages he or she has visited recently:



The application has several ways to visualise the data, for instance patient movements animated over time on a map:



To gather the data in a central place the application can send the patient data to a central server over any available internet connection, e.g. GSM. If the data transfer is interrupted, the application will automatically resend the data from the point of failure. The application will also automatically resend patient records that have been updated. The application has been made robust for field work with intermittent internet connectivity.

Since land-line internet connectivity may not always be available, hospitals expressed the requirement to keep local paper records of patients. To facilitate this, the application can export the data records available on the device in various file formats, including PDF for easy printing.

In summary, the Mobile Patient Recording App offers the following functionality:

- User-friendly entering and updating of data
- Local storage of data
- Export of local data to various file formats
- Synchronisation of local data to a remote server with simple JSON formatted messages
- Visualisation of data in graphs and on maps
- Runs on tablets with Android 2.3.3 or higher

11.3.2 Licencing & Availability

The DIRA App is available as open source. For further information on using the app in your own projects, don't hesitate to contact the EO2HEAVEN project team (eo2heaven.org).

11.4 WEB-BASED COMMUNITY COLLABORATION AROUND MODEL-BUILDING

Scientists often work for themselves in their own desktop environment, with their own tools and software used to perform analysis using datasets created or prepared by them. They are often confronted with numerous data pre-processing tasks during this process. Scientists may try to ease these tasks by basic shell, batch or similar scripting, but often perform tedious manual pre-processing steps. Ultimately, a research result is generated, which is usually published. Difficulties emerge, however, when (and if) they wish to share their research process (the models, scripts and pre-processing tasks) with fellow researchers, since the steps followed and artefacts produced are so particular to the researcher's local environment. This leads to missed opportunities:

- If researchers could communicate, collaborate and discuss issues relating to their models, scripts and processes at an early stage, they could shorten the learning process or overcome initial challenges more readily.
- Research could be reproduced and extended more readily if there were ways to share the objects of that research, i.e. the models and scripts, rather than the textual publications alone.
- Research teams could form more easily around research topics if the objects of research where easily reached and utilised.

These opportunities may seem artificial, e.g. in some cases researchers wish to protect their intellectual endeavours. Nevertheless, these opportunities are worth further consideration. Two emerging approaches are required to address these opportunities:

- Researchers and research institutions should foster the use of social media and the broader Web platform for communication purposes.
- Use of the Web to host geo-processing repositories and services for the purpose of storing, sharing and providing access to models, functions and scripts. This would support collaboration between researchers on the same or different projects. This is illustrated best through an example. A researcher may upload the first version of a geo-processing model to a repository. Partners on the research project may then test the model by downloading it, deploying and running it from their own servers or running it remotely to avoid configuration effort. Improvements could then be made and the second version of the model re-uploaded to the repository. The improved and verified model could then be shared more widely with the research community, and its use be sustained. The concept is described in detail in the public EO2HEAVEN Specification of Advanced Geo-Processing Services and Müller et al., 2010 & 2013.

Sharing of models requires a common understanding of the underlying components or building blocks. Technical descriptions of necessary database systems, component frameworks and scripting languages need to be provided, for example. In order to support the shared understanding of business logic inside models, the semantics of the logic components needs to be mutually understood and agreed upon, such as what it menas to use a particular geooperator. Further research in this direction is required. This vision of sharing and collaborating on models in a wide research community would provide benefits to the GEOSS community, for instance – the existing data centric collaboration would be raised to a higher level, that of process-centric collaboration. The GEOSS brokers would not only provide a registry and collection of datasets but also a set of tools and models that can be applied to them.

Approaches such as this are favoured in communities outside of the geo-informatics domain. Scientific Workflow tools, e.g. VisTrails (Vistrails, 2013) or Taverna are used to support the easy creation, sharing and execution of models in various research communities, e.g. eco-informatics. Special emphasis is placed on tracking scientific provenance and ensuring repeatability of experiments. Spatial analysis and data processing components for scientific workflow environments exist but are still in their infancy, e.g. the EO4VisTrails project (EO4Vistrails, 2013).

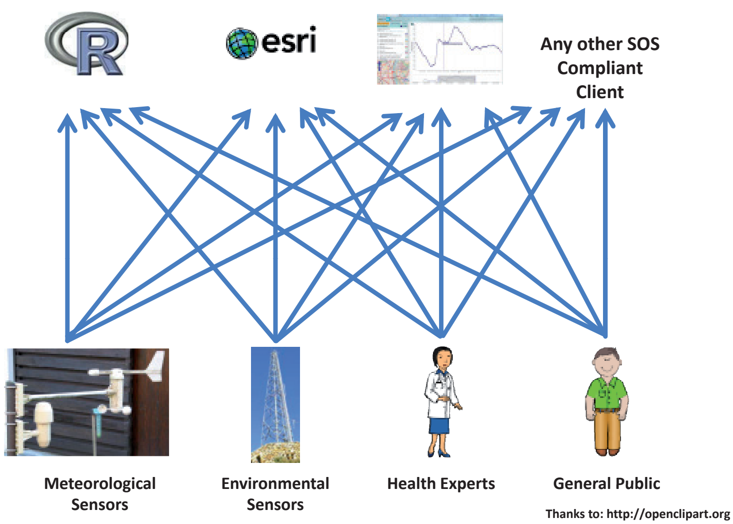
The use of scientific workflow environments/ workbenches in the geo community holds great potential for further research. These workbenches are oriented around model building and allow the integration of functionality supplied from different legacy systems, e.g. stemming from desktop based GIS or statistical analysis software, distributed Web Services and varied programming and scripting environments.

11.5 OGC WMS WRAPPER FOR WPS

Geospatial web processing is often implemented using the OGC WPS specification, which is currently only supported by a few clients are. However, if a WPS provides visualizable data like feature collections or raster images, it can be wrapped by the more widespread OGC WMS interface to allow WMS clients to access those components. The developed WMS wrapper requests and receives data from a WPS providing either an air quality modelling or zonal aggregation process. This allows for the client to access and visualize processed data via the WMS interface. The wrapper is implemented in Java and can be configured to serve as a wrapper for any WPS delivering raster or feature data.

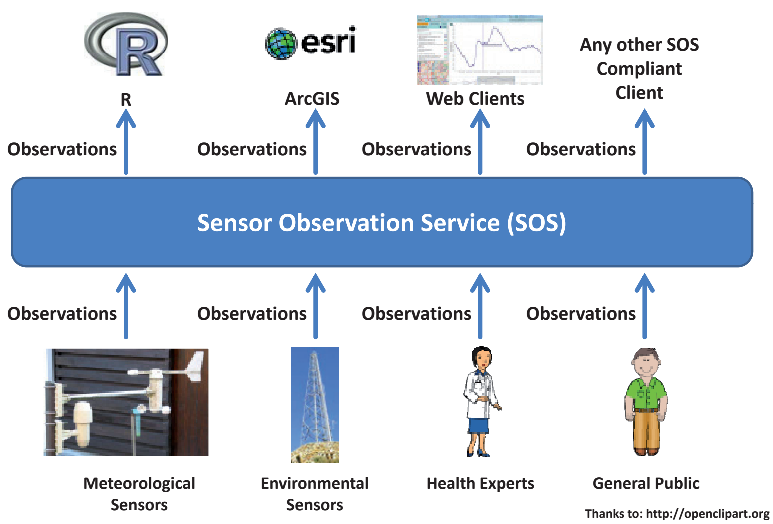
SENSOR WEB COMPONENTS 11.6

Within the EO2HEAVEN project scenarios there are lots of different sensors utilised (e.g. meteorological sensors, environmental sensors, observations by health experts such as illness cases, observations of the general public, observations from field sampling campaigns, etc.). However, a common way to integrate sensor data into applications was lacking. For each sensor type and for each client, particular adjustments to the software components were necessary in the past. If one sensor type was to be used by different client applications, each client application had to be adjusted separately to each sensor. The illustration below shows an example for this integration overhead: There are four client applications and four sensor types. This means that each of these four applications needs to be enhanced by four connectors to the different sensor types (in total 16 software modules need to be programmed). If one new client application were to be added, four new connectors to the sensors would need to be created.



Integration of sensor data without the EO2HEAVEN Sensor Web approach

There exists a mechanism whereby all sensor types (e.g. meteorological sensors, environmental sensors, observations by health experts such as illness cases, observations of the general public, etc.) can use the same approach to for publishing their data: the Sensor Observation Service or SOS. The SOS provides a common interface and data format for all these different types of observation data. This provides an interoperable mechanism for accessing these different data sets. Each implementation of SOS client software is able to access the same data from the service. For example, R, ArcGIS, Web Clients and all other applications supporting the SOS standard can use the same data access method. The following illustration shows the benefit of the Sensor Web solution of EO2HEAVEN. By using the standardised Sensor Observation Service (SOS) interface, the integration overhead is significantly reduced:



Easier integration of sensor data through the EO2HEAVEN Sensor Web approach

For accessing the data, the actual sensor type does not matter: asking for the data happens always in the same way through the SOS standard. Thus, if a client supports the SOS interface, it can easily combine data from lots of different domains (e.g. combining health data directly with environmental and meteorological data). Consequentially, the necessary implementation effort is reduced: For every sensor type it is necessary to build only one SOS adapter and for every client only one SOS adapter needs to be created (only if they are not yet already available). If a new client application is added only one SOS adapter needs to be created.

Sensor Observation Service Server 11.6.1

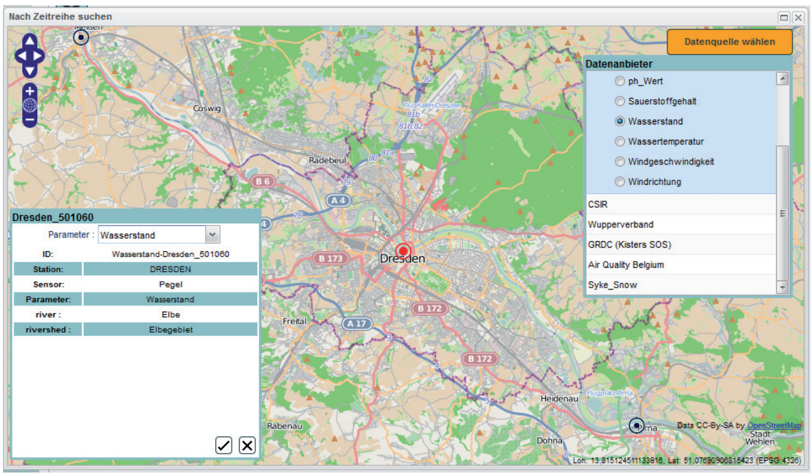
Standardized access to sensor observations and sensor metadata is provided by the OGC Sensor Observation Service (SOS). Observations offered by the SOS can be collected by physical or virtual sensors (i.e. simulations). Observations can even be collected indirectly, for example, laboratory analysis results from field sampling campaigns. The service acts as a mediator between a client and the sensor data archive or a real-time sensor system. The heterogeneous communication protocols and data formats of the associated sensors are hidden by the standardized interface of the SOS. Sensor data requested by a client are returned as observations. The interface of the SOS supports access to heterogeneous sensor types, stationary as well as mobile sensors which gather their data in-situ or remotely. Within EO2HEAVEN, a significant contribution to the 52°North SOS project was made. Important features of this SOS implementation comprise:

- Implementation of the SOS 1.0 and 2.0 standard
- Standardised access to sensor data
- Support of the transactional SOS operations for publishing new sensor measurements in a common manner
- Support of the EO2HEAVEN lightweight SOS profile for stationary in-situ sensors to increase interoperability
- Modular architecture for easier customisation of the SOS to existing databases
- GUI based tools for setup and administration of the SOS server

Sensor Observation Service Client 11.6.2

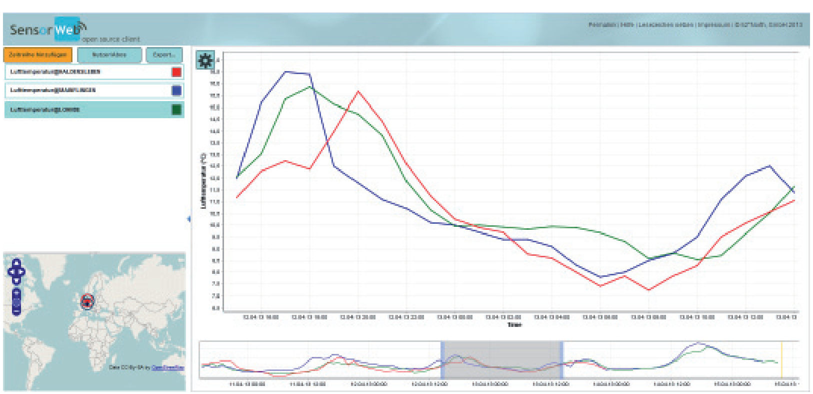
The 52°North Sensor Web Client that was developed with the co-funding of EO2HEAVEN, is a web-based client application which offers an easy to use interface to generate graphs from near-realtime sensor data. The client hides technical details of SWE services and protocols, so that even non-experts can use components of the Sensor Web transparently. To enable maximal platform independence the client is developed as a web-based browser application, making use of newest web technologies such as asynchronous client/server communication.

The client is capable of loading multiple time series from multiple SOS instances. Sensor data can be displayed as graphs and exported as various formats. A table of contents shows information about color and hatching settings of the displayed measurement series and offers options to manipulate the visualisation. The application allows exploring sensor metadata and "saving" current view as permalink. A map base menu is available for selecting the time series data the user wants to view. Within this menu, the user first selects the parameter (observed property) he is interested in. In the next step all available measurement stations for this parameter are displayed on a map:



The SOS Client data selection menu

After selecting one station, the relevant data is retrieved from a SOS server and displayed as a diagram as illustrated below. By repeatedly calling the time series selection menu, the user is able to overlay multiple time series (for different parameters and/or stations) to conduct a visual analysis of the available in-situ observation data:



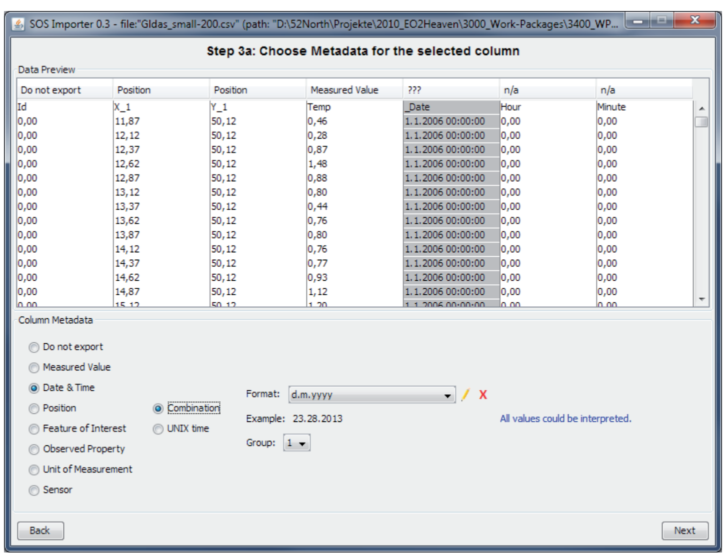
The SOS Client time series display

Sensor Observation Service Publication Tool 11.6.3

Within the EO2HEAVEN project scenarios, many existing (sensor) datasets are used which cover the appropriate period of time. However, these datasets were often only available in form of existing files (e.g. comma separated value or CSV) or local databases. In order to integrate these datasets into the EO2HEAVEN system, which will make use of distributed Sensor Web components, it was necessary to provide the existing sensor data though standardised web based interfaces, i.e. the OGC Sensor Observation Service (SOS). The Sensor Observation Service Publication Tool facilitates the import of these sensor data archives into SOS instances. As a result, the publication of observation data sets in on a SOS server becomes a task manageable for Non-IT-experts.

The application makes use of the wizard design pattern which guides the user through different steps. Within these steps the user describes the existing data files by providing information such as

- Description of each column meaning (e.g. timestamp, measured values)
- Geoferencing of the sensor locations
- Data types (e.g. format of time stamps, types of the measured data)
- Additional metadata that might be missing (e.g. sensor type, time zone of the sensor)



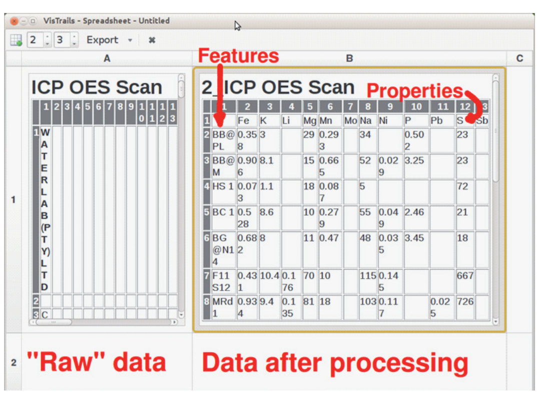
The SOS Publication Tool

The tool uses the provided information to convert the content of the existing data files into the Sensor Web data models/XML representations automatically and inserts them into a selected SOS server.

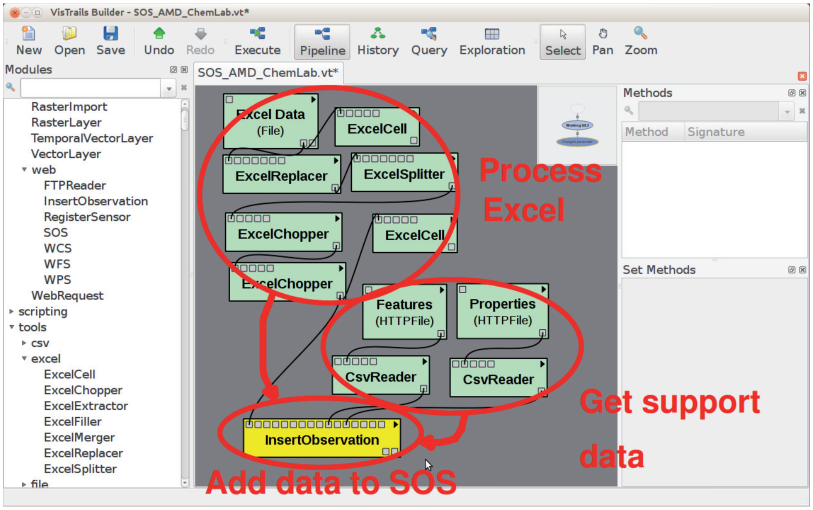
11.6.4 Scientific Workflow Components for publishing laboratory results to SOS

In the EO2HEAVEN project, water and sediment samples were collected during the Ugandan field sampling campaigns, which concerned the presence/absence of *Vibrio cholerae* as well as biophysical measurements taken at the same time. The project treated lab results and biophysical measurements as if the samples were directly analysed in-situ, at the sampling location, hence following the SOS approach, to make results from lab analysis accessible over the web..

This approach required sets of pre-processing software to be developed for parsing semi-structured data from laboratory result spreadsheets into a structure and format suitable for insertion into a SOS. This is similar to the requirement described on section 11.6.3. Yet, it is more complex, since the data had to be given structure during several supervised pre-processing steps. Field sampling campaigns may return results irregularly; it is useful for the scientist who is responsible for loading the SOS with data to be able to repeat the pre-processing steps without having to rebuild scripts. The software utilised was EO4Vistrails (EO4Vistrails, 2013), a geospatial extension to the Vistrails (Vistrails, 2013) scientific workflow software suite, which defined several repeatable/reusable workflows for performing the translation between laboratory and SOS format. A number of modules for EO4VisTrails were developed to support these workflows, which now form part of the main release of the software:



EO4VisTrails



EO4VisTrails Tools

Licencing & Availability 11.6.6

The Sensor Web Components are open source projects.

- The current implementation of the SOS server is available online at: http://52north.org/sos
- The current implementation of the SOS Import tool is available online at: https://wiki.52north.org/bin/view/SensorWeb/SosImporter
- The current implementation of the client is available online at: http://www.52north.org/sensorwebclient
- The current implementation of the scientific workflow SOS publishing components can be found in the EO4VisTrails project, available online at: http://code.google.com/p/eo4vistrails

For further information on using the client in your own projects, don't hesitate to contact the EO2HEAVEN project team (eo2heaven.org).

12 References

Chapter 2

Deutscher Allergie- und Asthmabund, 2010

Asthma beim Kind. Available online at: www.daab.de/atemwege/asthma-im-kindesalter

Directive 95/46/EC

Directive 95/46/EC of The European Parliament and of the Council of the 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data

Franck et al., 2011

The effect of particle size on cardiovascular disorders — The smaller the worse. Science of the Total Environment 409, 4217–21

Heinrich et al., 2002

Die europäische Studie zu Atemwegserkrankungen bei Erwachsenen, Pneumologie 2002

Kistnasamy et al, 2008

The relationship between asthma and ambient air pollutants among primary school students in Durban, South Africa. Int. J. Environment and Health 2008; 2:365-385.

McConnel et al., 2010

Childhood Incident Asthma and Traffic-related Air Pollution at Home and School. Environ Health Perspect. 2010 Mar 22

Meyer et al., 2007

Particulate Matter and Cardiovascular Health in the European Union: Time for Prevention?, Springer

Naidoo et al., 2007

South Durban Health Study. Final Project Report. Centre for Occupational and Environmental Health, University of KwaZulu-Natal, Durban, South Africa. 2007

Nsubuga, Peter et al, 2006

Public Health Surveillance: A Tool for Targeting and Monitoring Intervention." 2006. Disease Control Priorities in Developing Countries (2nd Edition),ed. , 997-1.018. New York: Oxford University Press. DOI: 10.1596/978-0-821-36179-5/Chpt-53.

Robert-Koch-Institut (ed), 2011

Gesundheitsberichterstattung des Bundes, Heft 52 - Sterblichkeit, Todesursachen und regionale Unterschiede, available online at: http://www.gbe-bund.de

Schlaud et al., 2007

Allergische Erkrankungen - Ergebnisse aus dem Kinder- und Jugendgesundheitssurvey (KIGGS). Bundesgesundheitsbl Gesundheitsforsch Gesundheitsschutz 50:578-599

Strickland et al., 2010

Short-term Association between Ambient Air Pollutants and Pediatric Asthma Emergency Department Visits. Am J Respir Crit Care Med. Apr 8

Trtanj, 2009

Integrated Surveillance and Monitoring for Vibrio Early Warning Systems. Presentation at the IEEE GEOSS Workshop XXXI - Using Earth Observation for Health; a workshop of the GEO Health and the Environment Community of Practice, 2009. Available online at: http://www.ieee-earth.org/event/geoss-workshop-xxxi-health

WHO, 2009

World Health Organization, Weekly epidemiological record. 2009. No 31, 84th year, pp 309–324,, available online at: www.who.int

Chapter 3

GEOSS for Health, The GEO Health Societal Benefit Area

EUROPEAN COMMISSION, Directorate-General for Research and Innovation, Directorate I – Environment, Unit I.3 – Management of natural ressources, Contact: Jane Shiel, European Commission, Office CDMA 03/157, B-1049 Brussels

INSPIRE, 2010a

About INSPIRE, available online at http://inspire.jrc.ec.europa.eu

THE EUROPEAN PARLIAMENT AND COUNCIL OF THE EUROPEAN UNION, 2007

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).

Chapter 4

Erl, 2008

SOA: Principles of Service Design. Prentice Hall. ISBN 0-13-234482-3.

Gordis, 2009

Epidemiology, 4th edition, Saunders Elsevier

ISO 10746-1, 1998

Information technology — Open Distributed Processing — Reference model: Overview, available online at: www.iso.org

Kirch (ed), 2008

Encyclopedia of Public Health, Springer

Novick, Robert (ed.), 1999.

Overview of the environment and health in Europe in the 1990s. World Health Organization.

OGC 07-097

Usländer, T. (ed.). Reference Model for the ORCHESTRA Architecture Version 2. OGC Best Practices Document 07-097, 2007, available online at: www.opengeospatial.org

OGC 09-132r1

Usländer, T. (ed.). Specification of the Sensor Service Architecture Version 3.0 (Rev. 3.1). OGC Discussion Paper 09-132r1, 2009, available online at: www.opengeospatial.org

Tze Wai Wong et al., 2012

Developing a risk-based air quality health index, Atmospheric Environment, Available online 4 July 2012, ISSN 1352-2310, 10.1016

Uganda Ministry of Health, 2010

The Health management Information System. District/HSD Procedure Manual, Volume 3

Wei-Zhen Lu et al., 2011

Assessing air quality in Hong Kong: A proposed, revised air pollution index (API), Building and Environment, Volume 46, Issue 12, December 2011, Pages 2562-2569, ISSN 0360-1323, 10.1016

WHO, 2009

Global Health Risks – Mortality and Burden of Disease attributable to selected major risk factors, Technical report, World Health Organization, 2009, available online at: www.who.int

Chapter 5

APMoSPHERE, 2005

6th Detailed report, related to overall project duration. APMoSPHERE project consortium, 2005

Janssen et al., 2008

Spatial interpolation of air pollution measurements using Corine Land Cover data. In: Atmospheric Environment, 42(20), pp. 4884–4903.

Lambert, 1992

Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics 34(1): 1-14

McGregor, 1996

Identification of air quality affinity areas in Birmingham, UK.
In: Applied Geography, 16(2): 109–122.

Tobler, 1970

A computer movie simulating urban growth in the Detroit region. In: Economic Geography, 46(2): 234–240.

Wiemann et al., 2012

Classification-driven air pollution mapping as for environment and health analysis. 6th International Environmental Modelling and Software Society (iEMSs), 2012, Leipzig, Germany

Chapter 6

Strickland et al., 2010

Short-term Association between Ambient Air Pollutants and Pediatric Asthma Emergency Department Visits. Am J Respir Crit Care Med. Apr 8

Chapter 7

Safa et al., 2008

Vibrio cholerae O1 Hybrid El Tor Strains, Asia and Africa. Emerg Infect Dis. 14(6): 987–988.

Chapter 8

IHR, 2005

International Health Regulations. WHO Official Records. World Health Organization, available online at: www.who.int/ihr/IHR_2005_en.pdf

Chapter 9

Bivand, 2008

Applied spatial data: analysis with R. Springer Science+Business Media, LLC.

Clements et al., 2006

Bayesian spatial analysis and disease mapping: tools to enhance planning and implementation of a schistosomiasis control programme in Tanzania. Tropical Medicine & International Health 11: 490-503.

Congdon, 2006

Bayesian statistical modelling. John Wiley & Sons.

Orford, 2005

Cartography and Visualization. In: Castree N, Rogers A, Sherman D, editors. Questioning Geography:Fundamental Debates. Malden: Blackwell Publishing. pp. 189-205.

Pfeiffer et al., 2008

Spatial analysis in epidemiology. Oxford University Press New York.

Root (ed), 2012

Moving Neighborhoods and Health Research Forward: Using Geographic Methods to Examine the Role of Spatial Scale in Neighborhood Effects on Health. Ann Assoc Am Geogr 102: 986-995.

Rushton, 2003

Public health, GIS, and spatial analytic tools. Annu Rev Public Health 24: 43-56

Waller, 2004

Applied Spatial Statistics for Public Health Data. Wiley.

Chapter 10

Hu, 2009

Spatial analysis of MODIS aerosol optical depth, PM2.5, and chronic coronary heart disease. International Journal of Health Geographics, 8:27.

Hyer et al., 2010

An over-land aerosol optical depth data set for data assimilation by filtering, correction, and aggregation of MODIS Collection 5 optical depth retrievals, Atmospheric Measurement Techniques Discussions, 3: 4091-4167.

Kumar et al., 2008

Remote sensing of ambient particles in Delhi and its environs: estimation and validation. International Journal of Remote Sensing, 29: 3383-3405.

Mölter et al., 2010

Modelling air pollution for epidemiologic research — Part I: A novel approach combining land use regression and air dispersion. Science of the Total Environment, 408: 5862-5869.

Pech et. al., 2013

Generation of multitemporal thermal orthophotos from UAV data. Will appear in the Proceedings of the International Conference on Unmanned Aerial Vehicle in Geomatics (UAV-g) in September 2013

Chapter 11

EO4Vistrails, 2013

http://code.google.com/p/eo4vistrails

Vistrails, 2013

http://www.vistrails.org

Müller, M. et al., 2010.

Moving Code in Spatial Data Infrastructures - Web Service Based Deployment of Geoprocessing Algorithms. Transactions in GIS, 14(s1), 101-118. doi:10.1111/j.1467-9671.2010.01205.x

Müller, et al., 2013

Moving code - Sharing geoprocessing logic on the Web. In: ISPRS Journal of Photogrammetry and Remote Sensing, 2013.

Why are you reading this book?

- Perhaps you are dealing with the impact of environmental factors on human health
- Perhaps you are wondering about putting medical data in a geographical context
- Perhaps you are looking for options to share existing data across domain boundaries

Or perhaps you are keen to understand how current trends in geospatial technology and global initiatives can help to bridge the divide between environmental studies and health research. If this is the case, then yes: this book is for you!

EO2HEAVEN stands for **Earth Observation and Environmental Modelling for the Mitigation of Health Risks** and is a major collaborative project funded under the 7th Framework Programme of the European Commission.

EO2HEAVEN designed and developed methods and tools to correlate environmental data with exposure and health data, to support the collection and integration of data and to visualise results in their geographical context.

SEVENTH FRAMEWORK PROGRAMME

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 244100