Jul 1st, 12:00 AM

Ontology-Based Approach to the Discovery of Human Health and Environmental Risks Assessment

Jiří Hřebíček

Ladislav Dušek

Jiří Kalina

Miroslav Kubásek

Jana Klánová

See next page for additional authors

Presenter/Author Information
Jiří Hřebíček, Ladislav Dušek, Jiří Kalina, Miroslav Kubásek, Jana Klánová, and Ivan Holoubek

International Environmental Modelling and Software Society (iEMSs)
2012 International Congress on Environmental Modelling and Software
Managing Resources of a Limited Planet, Sixth Biennial Meeting, Leipzig, Germany
R. Seppelt, A.A. Voinov, S. Lange, D. Bankamp (Eds.)
http://www.iemss.org/society/index.php/iemss-2012-proceedings

Ontology-Based Approach to the Discovery of Human Health and Environmental Risks Assessment

Jiří Hřebíček[a], Ladislav Dušek[a], Jiří Kalina[a], Miroslav Kubásek[a], Jana Klánová[b], Ivan Holoubek[b]
[a]*Masaryk University, Institute of Biostatistics and Analyses, Kamenice 125/3, 625 00 Brno, Czech Republic, {hrebicek,dusek,kalina,kubasek}@iba.muni.cz*
[b]*Masaryk University, Research Centre for Toxic Compounds in the Environment, Kamenice 125/3, 625 00 Brno, Czech Republic, {klanova,holoubek}@recetox.muni.cz*

Abstract: The Global ENvironmental ASsessment Information System (GENASIS) is a system developed by the Research Centre for Toxic Compounds in the Environment and the Institute of Biostatistics and Analyses at the Masaryk University in Brno. Using the IT tools which have been developed in the FP7 project Tagging Tool based on a Semantic Discovery Framework (TaToo), the GENASIS project aims to set up a semantic web solution to close the discovery gap which prevents a full and easy access to environmental resources on the web. The aim of the GENASIS project is to compile discovered and validated data on persistent organic pollutants (POPs), including their properties, sources, long-term levels, lifetimes, transport mechanisms, effects and risks; these data have been scattered across various institutions and ministries. The project will provide tools for the visualization, analyses and interpretation of these data, and will make it possible to assess environmental and human risks, or to model the fate of these toxic compounds. It can be used to facilitate the discovery, integration and analysis of environmental data in cancer-related risk studies. POPs were chosen as a model because of their persistence, bioaccumulation potential and genotoxicity. The paper introduces GENASIS and its analytical tools connected with TaToo that significantly enhance comprehensive discovery, tagging and understanding of information resources about the fate of POPs in the environment and their impacts on ecosystems and the human population.

***Keywords**:* GENASIS; TaToo; POPs; cancer risk; information discovery

1 INTRODUCTION

Contamination of the environment is one of many factors widely agreed to have a negative impact on human health. Over the last few decades, a number of negative effects related to environmental contamination have been confirmed. There is increased incidence of various diseases in populations exposed to contaminated environment; these include health issues such as cardiovascular and respiratory diseases, reproductive problems and occurrence of malignant tumours[1]. Toxic chemicals play an important role in environmental contamination. Therefore, the United Nations (UN) conventions on toxic chemicals are focused on the protection of human and environment against toxic chemicals, wastes and products of these chemicals, their illegal transport and illegal disposal, see Holoubek et al. [2011].

The most important conventions from this point of view include the Stockholm Convention[2] (SC), the Basel Convention[3] (BC) and the Rotterdam Convention[4]

[1] http://www.svod.cz
[2] http://www.pops.int
[3] http://www.basel.int/

(RC). The synergy principles in the cooperation among these Conventions have been frequently discussed in recent times. Other relevant chemical conventions include the Montreal Protocol on Substances That Deplete the Ozone Layer[5] and the UN ECE convention on long range transboundary air pollution[6] (CRLTAP). The synergy process currently taking place within the Stockholm, Rotterdam, and Basel conventions creates demand and opportunity for a more comprehensive approach that extends support beyond Persistent Organic Pollutants (POPs) and Ozone Depleting Substances (ODS). One of the most important scientific goals of the SC, BC and RC is focused on the development of new approaches to the evaluation of multiple chemical exposures, as well as toxicological interactions of candidate POPs, see Klanová et al. [2012], Synergies Success Stories [2011].

IT tools are needed for the solution of problems related to chemical pollution, which should finally support the decision-making process in order to minimize the impact of anthropogenic activities on human and ecosystem health (Dušek et al. [2011], Kubásek et al. [2011], Reis et al. [2010], Sabel et al. [2009]). These include tools for a safe and acceptable disposal; tools for the acceptance of best available techniques, waste management, remediation of contaminated sites, monitoring of temporal and spatial trends; tools for the evaluation of effectiveness of the conventions measures; and finally, tools for the visualization and interpretation of data from national and international inventories. All these aspects are the key points of POPs-related issues solved by the Research Centre for Toxic Compounds in the Environment[7] (RECETOX) and the Institute of Biostatistics and Analyses[8] (IBA) at the Masaryk University (MU). RECETOX is also the Czech National Centre for Persistent Organic Pollutants[9] (hereinafter referred to as the "Centre"). The Centre is located in the new campus of MU in Brno, Czech Republic.

The Centre has obtained the official status as the „Stockholm Convention Regional Centre for capacity building and transfer of technology in Central and Eastern Europe“ (Regional Centre) by nomination of fourth Conference of the Parties of the Stockholm Convention[10] held in Geneva in May 2009. The Regional Centre[11] is a joint project of countries in Central and Eastern Europe (CEE) and brings together representatives of countries, institutions and experts from these countries as well as interested parties outside the region. It is mainly focused on assistance and cooperation in the field of monitoring, sampling and analytical techniques, environmental modelling and data analysis, ecotoxicology and risk assessment. The activities of the Regional Centre are financially supported by the Ministry of Environment of the Czech Republic (MoE).

The activities of the Regional Centre are carried out in accordance with the decisions of the Conference of Parties to the SC No. SC-4/23 and SC-2/9. It offers the strongest expertise in the area of POP analysis ranging from the development of sampling techniques (for air as the key matrix in the Global Monitoring Plan[12] of the SC, but also for water, the need of which increases with new and more soluble substances added to the list of SC), through their application in the field, the development and application of analytical methods (not only) for new chemicals put on a list of the SC, up to the implementation of pilot studies and long-term monitoring programs, management and data processing and development of environmental databases (Dušek et al [2011], Klánová et al [2012]). The Global Monitoring Plan as one of the effectiveness evaluation mechanisms under the SC was designed to assess the trends in ambient air and human milk levels of POPs, and the involved countries were invited to establish long-term monitoring programmes. Complementarily, the Regional Centre also provides a base for

[4] http://www.pic.int/
[5] http://ozone.unep.org/new_site/en/montreal_protocol.php
[6] http://www.unece.org/env/lrtap
[7] http://www.recetox.muni.cz
[8] http://www.iba.muni.cz
[9] http://www.recetox.muni.cz/index-en.php?pg=national-pops-centre
[10] http://chm.pops.int/Implementation/RegionalCentres/Overview/tabid/425/Default.aspx
[11] http://www.recetox.muni.cz/index-en.php?pg=regional-pops-center
[12] http://chm.pops.int/Implementation/GlobalMonitoringPlan/Overview/tabid/83/Default.aspx

ecotoxicological studies focused on the effect of chemicals on living organisms, the analysis of environmental and human risks as well as the impacts of chemical pollution on biodiversity.

The main goal of the Regional Centre is to develop the *Global ENvironmental ASsessment Information System*[13] (GENASIS), as described in Dušek et al. [2011] and Holoubek et al [2011]. This paper introduces GENASIS functionalities with its analytical tools that enable human health and environmental risks evaluation. It also introduces ontologies[14] of the FP7 project "Tagging Tool based on a Semantic Discovery Framework"[15] (TaToo) Dihé et al. [2011], Kubasek et al. [2011]. They are used in the GENASIS project in order to provide a user-friendly interface for the semantic annotation of environmental resources, more specifically environmental data, information, web services and models.

2 GENASIS

2.1 Challenges of GENASIS

The synthesis of existing (air) pollution monitoring databases with epidemiological data is required in order to identify the effects of pollution on human health (Dušek et al [2011]) and to evaluate environmental and human health risks. This task requires a new approach to data discovery, using the maximum of available knowledge. Partners of IBA and RECETOX ask for new anthropogenic impact studies which require data discovery from a multitude of monitoring networks and resources. Proper use of such data requires contextual information, which will be delivered via TaToo by the means of tagging and enhanced information description (meta-data). In this context, MU intends to employ TaToo tools and validate their performance in tagging and semantic rich discovery throughout resources related to effects of pollution on human health. Therefore, a new version of the information system GENASIS has been developed, taking into account requests on interoperability as defined by the eGovernment[16] of the Czech Republic and SEIS[17].

2.2 Implementation of GENASIS

GENASIS is a web information system developed by the RECETOX/IBA with the usage of IT tools developed within the TaToo project. The aim of the GENASIS project is to compile discovered and validated data on POPs, including their properties, sources, long-term levels, lifetimes, transport mechanisms, effects and risks; these data have been scattered across various institutions and ministries. The project will provide tools for the visualization, analyses and interpretation of these data, and will make it possible to assess environmental and human risks or modelling of fate (Holoubek et al [2011]). Such IT tools should significantly enhance a comprehensive understanding of the fate of POPs in the environment and their impacts on ecosystems and the human population. The combination of expert knowledge and validated data from several cooperating institutions create an opportunity for a broad spectrum of visualizations, analyses and modelling.

The initial phase of the GENASIS development was focused on data from regular monitoring programmes, providing a general overview on spatial patterns and temporal trends of pollutants concentrations distributed in various environmental matrices (air, soil, water, biota)(Holoubek et al [2011]). The developed web portal GENASIS provides a unique information source to be used by both the public and

[13] http://www.genasis.cz

[14] http://www.w3.org/standards/semanticweb/

[15] http://www.tatoo-fp7.eu/tatooweb/

[16] http://www.mvcr.cz/mvcren/scope-of-activities-egovernment.aspx

[17] http://www.eea.europa.eu/about-us/what/information-sharing-1/shared-environmental-information-system

the scientific community. It serves as an information support for the SC implementation on the international level. Linking to additional information sources of the MoE provides the potential for a complex assessment of anthropogenic impacts on the environment and related ecological and human health risks of the Czech Republic and the CEE region.

GENASIS has been built as a modular structure providing web services to a wide range of potential users. The initial version of the database, which was launched in 2010, contains data from the long-term integrated monitoring at the Košetice observatory, which is part of the European Monitoring and Evaluation Programme[18] (EMEP), and long-term data from the *MOnitoring NETwork* for determination of POPs in ambient air using passive sampling (MONET) in the Czech Republic, Dvorská et al. [2008].

3 GENASIS as TaToo validation scenario

GENASIS is a part of the validation scenario "Anthropogenic impact and global climate change" (abbreviated "MU Scenario") of the FP7 project TaToo. The validation scenario is led by the MU and is focused on the correlation of environmental pollutants (POPs) and their impact on health of the human population. The development of a set of TaToo tools will provide a preliminary semantic analysis of the content of web resources for the MU Scenario, with the ability to access different published ontologies that describe the available knowledge basis. The aim of MU Scenario is to create a central place for researchers, domain experts and decision makers to discover and access interdisciplinary knowledge in a more efficient and usable way that is the current state of the art (Dihé at al, Kubásek et al [2011]).

The MU Scenario is not limited to the use of TaToo tools in order to improve the discovery of scientific resources for one particular domain, but also tries to discover and create new relationships among different domains (environmental pollution and cancer epidemiology) (Kubasek et al [2011]). The correlation of POPs and their impact on the human health is only one significant example of creating new relationships among different domains. These correlations could represent new scientific insights into already available resources and connect the knowledge of the single domains. These relationships should facilitate further discovery process to deliver matching resources of multiple domains.

3.1 Domain ontology and GENASIS

The objectives of the TaToo project at semantic tagging and searching environmental resources are based on the principles of the Semantic Web using ontologies as the underlying model for tagging and searching resources. The main objective of the ontologies within the TaToo project Framework is to allow formal and contextual cross-domain tagging of searched environmental resources (Dihé et al. [2011]). Within the process of ontology development, we kept existing W3C standards such as RDF[19], RDFS[20] and OWL[21] and, in particular, a subset of OWL 2 called OWL2 RL. Within the TaToo project, the use of the NeOn methodology[22] has been fostered. As regards the specification of ontologies, we have used current tools for ontology engineering, such as the NeOn-Toolkit[23] or Protégé[24].

[18] EMEP (http://www.emep.int/) is a scientifically based and policy driven program under the Convention on Long-range Transboundary Air Pollution (http://www.unece.org/env/lrtap/) for international co-operation to solve transboundary air pollution problems.

[19] http://en.wikipedia.org/wiki/Resource_Description_Framework

[20] http://en.wikipedia.org/wiki/RDFS

[21] http://en.wikipedia.org/wiki/Web_Ontology_Language

[22] http://www.neon-project.org/nw/NeOn_Book

[23] http://www.neon-project.org

By the analysis of both domains, i.e. environmental and human risk assessment areas, we have identified four levels of developed ontology system:

- *Nomenclature*, which contains a set of nominal descriptors identifying key objects (variables); these should be mostly identified in investigation or in research searching for information (POPs – chemical compounds and diseases – cancer diagnoses). This level is highly standardized, adopting internationally unified, extensively translated nomenclature systems. In the developed ontology, we used the International Classification of Diseases[25] (ICD-10) class hierarchy and recommended POPs and Matrix taxonomy based on the SC.
- *Classifiers*, which present attributes determining some key properties of the examined objects (chemical compounds, diseases). Only classifiers highly relevant for studies concerned with environmental exposure and risk assessment studies are adopted. As a result, the classifiers represent binary codes or multiple categories, typically derived on the basis of some external information accessible from a standardized database, evidence-based literature, a thesaurus or an encyclopaedia (properties of a given chemical compound, properties of a certain disease at the time of diagnosis, etc.). The set of attributes is flexible according of used classifiers; e.g., in case of studies focusing on some special topics etc.
- *Identifiers*, which are necessary descriptors of the source of information that is processed or needed. These attributes also refer to some type of validity scoring because they describe the types of studies and other information sources which can be regarded as relevant. Furthermore, this set of attributes allows the users to specify studied problem or scientific field to be inspected.
- *Obligatory descriptors*, which represent a family of variables describing the lowest level of the information processed or searched for. It means set of variables which are obligatory to understand the numerical values achieved by search engines (units, sampled, matrices, epidemiological measures, time and site specification, etc.). These descriptors are strictly obligatory and represent valid information, i.e. only information; numerical value supplied with these descriptors can be regarded as valid and trustworthy.

3.2 GENASIS mapping onto the TaToo Framework Architecture

The main entry point of the TaToo Framework Architecture for accessing tags and annotations of environmental resources is the TaToo Web Portal and TaToo Web Services, which both provide access to the TaToo repository. The TaToo Web Portal is built on Liferay Portal[26] which is an implementation of JSR-286[27] enterprise portal. TaToo Web Services include a set of services based on the SOAP[28] and WSDL[29] standards, and the TaToo repository is build on Sesame framework - a de-facto standard framework for processing RDF data which fully supports the SPARQL query language[30] for expressive querying.

Within GENASIS, we have developed the MU Specialized Tagging Portlet. This portlet is part of the TaToo Web Portal and provides domain-specific tagging functionality. It supports a variety of meta-information schemes based on a very generic user interface, and is designed for tagging and annotating a specific set of

---

[24] http://protege.stanford.edu

[25] http://www.who.int/classifications/icd/en/

[26] http://www.liferay.com

[27] http://en.wikipedia.org/wiki/Java_Portlet_Specification

[28] http://en.wikipedia.org/wiki/SOAP_%28protocol%29

[29] http://en.wikipedia.org/wiki/Web_Services_Description_Language

[30] http://en.wikipedia.org/wiki/SPARQL

domain-related resource types with POPs and other related concepts from the Domain Ontology of MU Scenario.

The Discovery Component of TaToo acts as a Web Client of the TaToo Discovery Web Service for GENASIS and is also able to display the retrieved information in an appropriate manner. It will become part of current analytical tools of GENASIS.

The Discovery Component of TaToo is designed to discover and visualize POPs and related resources and thus is focused on a set of predefined topics from the MU Domain Ontology.

Figure 1 describes the sequence diagram of Discovery Component of TaToo, where we illustrated a user considering first processing analyses on GENASIS portal and then deciding to display all similar resources from the TaToo repository. This evokes the *2.1. getResources* action and the user can see a list of similar resources. The user can then get a detailed description of a specific resource and see them (display, download... depending of type of resource). The *1.1.1 get NumberOfResources* action is evoked continuously during processing analyses, to give the user an idea of how many similar resources are available in the TaToo repository.

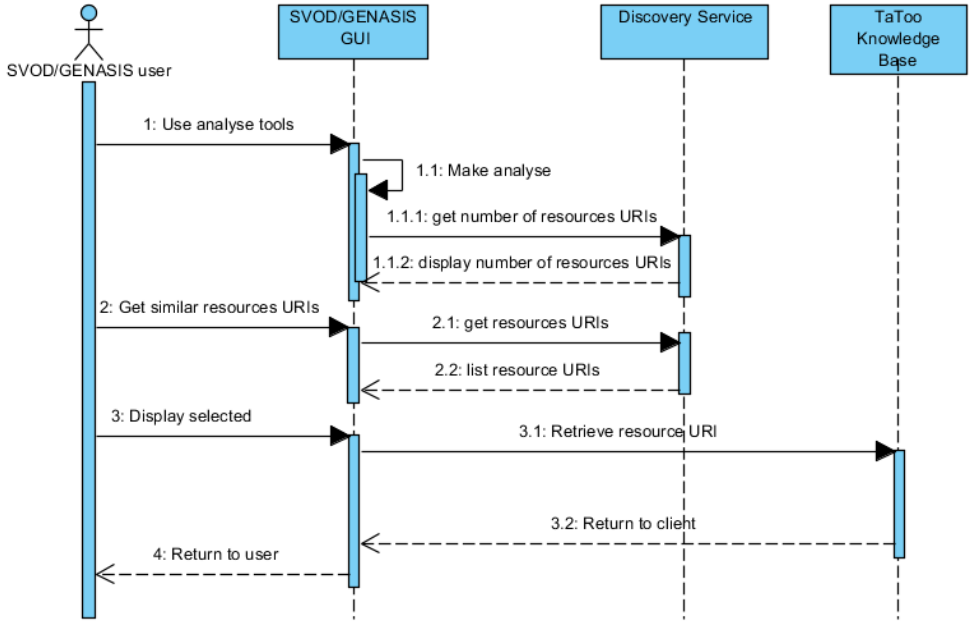

Figure 1. Sequence diagram of Discovery Component.

3.3. Analytical tools of GENASIS

Analytical tools of GENASIS are the most important parts of the whole project. These tools allow the comprehension and basic processing of measured environmental data by the GENASIS Data Browser, which enables the user to select the kind of data entering the analysis (e.g. region, project name, sampling time, matrix, chemical compound, type of monitoring etc.). The GENASIS Data Browser provides the selected set of data for further analyses, see Figure 2. It is possible to visualize the location of each sampled site by means of synoptic maps and examine general and/or detailed information about the sampling frequency. It is also possible to sort and select/deselect localities and view measured concentrations of selected compounds at those localities (data overview menu).

Using additional modules, it is possible to obtain descriptive statistics for selected data set, to monitor changes in the concentration of user-selected chemicals over a specified time period and to display seasonal and long-term trends. Each module includes an option to use additional criteria that restrict entry data (e.g. the selection of explicit altitudes). Another integral part of the analytical modules is the stratification of localities according to various parameters (land use, altitude, distance to roads, sources of pollution, inhabited areas), which enables a more detailed view and discrimination of localities. More complex analyses and models are currently and continuously being prepared.

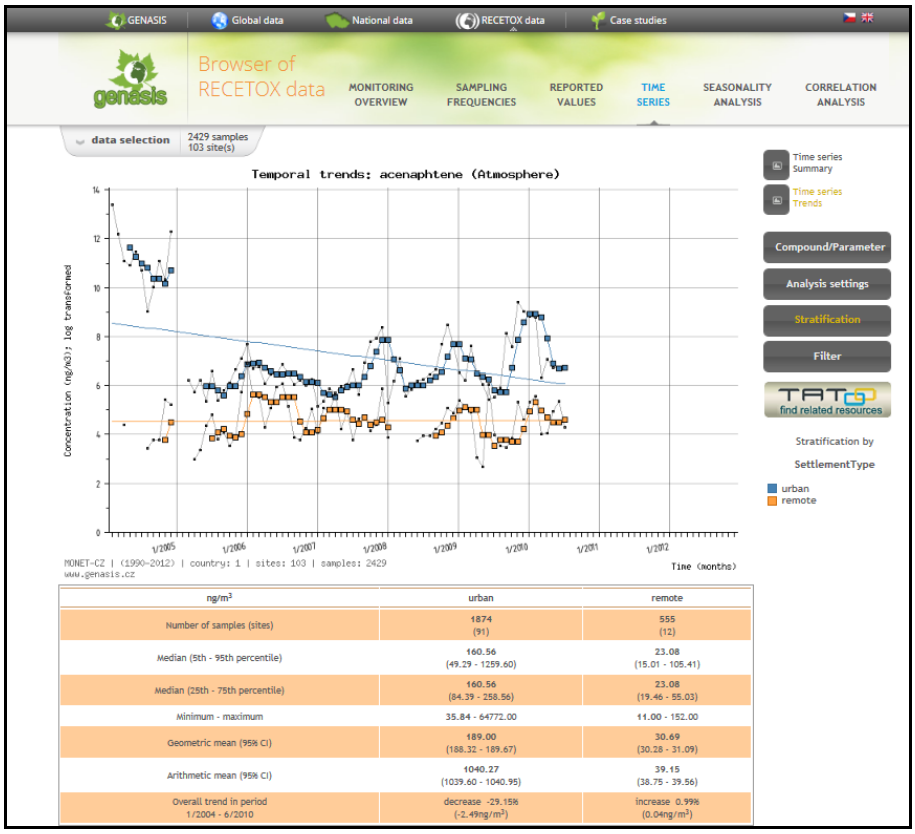

Figure 2. Analytical tools of GENASIS.

4 Conclusion

We have introduced GENASIS functionalities as an ontology-based approach to the discovery of human health and environmental risks assessment together with its analytical tools. GENASIS discovery tools are connected with TaToo tools. They significantly enhance a comprehensive discovery, tagging and understanding of information resources about the fate of POPs in the environment, as well as their impact on human health. They make it possible to assess environmental risks. We have also introduced the MU Scenario of the FP7 project TaToo which has been specifically designed to provide a user-friendly interface for the semantic annotation of environmental resources, more specifically environmental data, information, web services and models. The paper also contains a short overview of domain ontologies for POPs contamination and cancer epidemiology domains. We have presented the TaToo framework and described how the GENASIS in MU Scenario is mapped onto the TaToo Framework Architecture, and have described in more detail two cases of selected validation use.

Acknowledgements

"The research leading to these results has been funded from the European Community's Seventh Framework Programme (FP7/2007-2013) under Grant Agreement Number 247893."

REFERENCES

Dihé, P., S. Frysinger, R. Güttler, S. Schlobinski, L. Petronzio, R. Denzer, S. Nešić, T.P. Lobo, G. Schimak, J. Hřebíček and M. Donatelli, An Architecture for the Semantic Enhancement of Environmental Resources, 9th IFIP WG 5.11 International Symposium on Environmental Software Systems: Frameworks of eEnvironment, ISESS 2011. Brno, Czech Republic, June 27-29, 2011.

Dušek, L., J. Hřebíček, M. Kubásek, J. Jarkovský, J. Kalina, R. Baroš, Z. Bednářová, J. Klánová, I. Holoubek, Conceptual Model Enhancing Accessibility of Data from Cancer–Related Environmental Risk Assessment Studies, 9th IFIP WG 5.11 International Symposium on Environmental Software Systems: Frameworks of eEnvironment, ISESS 2011. Brno, Czech Republic, June 27-29, 2011.

Dvorská, A., G. Lammel, J. Klánová and I. Holoubek, Košetice, Czech Republic – ten years of air pollution monitoring and four years of evaluating the origin of persistent organic pollutants, *Environmental Pollution,* 156 (2), 403-408, 2008.

Grossman, D.A. and O. Frieder, *Information Retrieval: Algorithms and Heuristics*. Springer, 332 pp., Heidelberg, 2004.

Holoubek, I., L. Dušek, J. Klánová, M. Kubásek, J. Jarkovský, R. Baroš, K. Komprdová, Z. Bednářová, R. Hůlek and J. Hřebíček, GENASIS Information System: A Global Environmental Assessment of Persistent Organic Pollutants, 9th IFIP WG 5.11 International Symposium on Environmental Software Systems: Frameworks of eEnvironment, ISESS 2011. Brno, Czech Republic, June 27-29, 2011.

Klánová, J., L. Dušek, J. Borůvková, R. Hůlek, K. Šebková, J. Gregor, J. Jarkovský, L. Kohút, J. Hřebíček and I. Holoubek, The initial analysis of the Global Monitoring Plan (GMP) reports and a detailed proposal to develop an interactive on-line data storage, handling, and presentation module for the GMP in the framework of the GENASIS database and risk assessment tool. Research report RECETOX, IBA MU Brno, 2012.

Kubásek, M., J. Hřebíček, J. Kalina, L. Dušek, J. Urbánek and I. Holoubek, Global Environmental Assessment Requires Global Functional Searching Engines: Robust Application of TaToo Tools, 9th IFIP WG 5.11 International Symposium on Environmental Software Systems: Frameworks of eEnvironment, ISESS 2011. Brno, Czech Republic, June 27-29, 2011.

Reis, S., T. Oxley and E. Rowe, Integrated Modelling for Health and Environmental Impact Assessment of Air Pollution and Climate Change. International Congress on Environmental Modelling and Software, Modelling for Environment's Sake: iEMSs Fifth Biennial Meeting.iEMSs 2010. Ottawa, Canada, July 5-8, 2010

Sabel, C.E., P. Boyle, G. Raab, M. Löytönenm and P. Maasilta, Modelling individual space–time exposure opportunities: a novel approach to unravelling the genetic or environment disease causation debate. *Spatial and Spatio-temporal Epidemiology*, 1, 85-94, 2009.

Synergies Success Stories. Enhancing cooperation and coordination among the Basel, Rotterdam and Stockholm conventions. Report of the Division for Sustainable Development, United Nations Department of Economic and Social Affairs (UNDESA), Secretariat of the Basel Convention, Secretariat of the Rotterdam Convention, Secretariat of the Stockholm Convention, New York: UN, 2011. < http://chm.pops.int/Portals/0/download.aspx?d=UNEP-FAO-DESA-SYN-COMPI-SynergiesSuccessStories.En.pdf>