GENASIS: On-Line Data Browser for Environmental Monitoring and Associated Information Systems

Jiří Jarkovský, Ladislav Dušek, Miroslav Kubásek, Richard Hůlek, Jakub Gregor, Jiří Hřebíček, Jana Klánová, Kateřina Šebková, Jana Borůvková, Ivan Holoubek

GENASIS: On-line data browser for environmental monitoring and associated information systems

Jiří Jarkovský[1], Ladislav Dušek[1], Miroslav Kubásek[1], Richard Hůlek[1,2],
Jakub Gregor[1,2], Jiří Hřebíček[1], Jana Klánová[2], Kateřina Šebková[2], Jana Borůvková[2],
and Ivan Holoubek[2]

[1]Institute of Biostatistics and Analyses, Masaryk University,
Kamenice 126/3, 625 00 Brno, Czech Republic
{jarkovsky, dusek, kubasek, hulek, gregor, hrebicek}@iba.muni.cz
[2] Research Centre for Toxic Compounds in the Environment, Masaryk University,
Kamenice 753/5 (A29), 625 00 Brno, Czech Republic
{klanova, sebkova, boruvkova, holoubek}@recetox.muni.cz

Abstract. The current environmental research is producing large quantities of data on all environmental matrices and number of chemical and biological endpoints. However, there is a paradox in this situation that we suffer from lack of representative data, since their production has not been accordingly supported by development of systems that make them accessible for analysis. The introduced GENASIS data browser aims in its current version on the problem of persistent organic pollutants and the analysis, visualization and interpretation of data from their monitoring networks. The development of the GENASIS involves problems of the so called information pyramid: monitoring (obtaining environmental data), data validation (including uncertainty quantification), data analysis, and methods of aggregation of data from different sources, model generalization and interpretation of results, and their presentation. Within the frame of the GENASIS data browser, we introduce a user-friendly system for the visualization and analysis of contamination of all environmental compartments by persistent organic pollutants and evaluation of actual POPs contamination, its long-term trends and seasonal fluctuations. Project outcomes are useful as information source both for lay public and experts, as well as for the process of the Stockholm Convention implementation.

Keywords: Stockholm Convention, POPs, database, data analysis, online visualization.

1   Background

The current environmental research is producing large quantities of data on all environmental matrices and number of chemical and biological endpoints. Production of a high amount of data has been currently typical for all fields of human activities and concerns also environmental research and monitoring. Thousands of projects have been in progress on the national and international level, which generate data that must

be administered, processed, analyzed, and interpreted. However, there is a paradox in this situation that we suffer from lack of representative data, since their production has not been accordingly supported by development of systems that make them accessible for analysis [1-2]; we currently face the "Data rich but information poor" problem when based on these quantities of data only small part of information is utilized for interpretation or decision support.

The introduced GENASIS data browser [3] aims in its current version on the problem of persistent organic pollutants and the analysis, visualization and interpretation of data from their monitoring networks. The project exists within the frame of the Stockholm Convention [4] tools and its objectives are availability of a user-friendly system for the visualization and analysis of contamination of all environmental compartments by persistent organic pollutants and evaluation of actual Persistent Organic Pollutants (POPs) contamination, its long-term trends and seasonal fluctuations.

The GENASIS project utilizes data from national and international monitoring networks to obtain as-complete-as-possible set of information and a representative picture of environmental contamination by POPs [5]. Project outcomes should be useful as an information source both for lay public and experts, as well as for the process of the Stockholm Convention implementation. The system has been developed while taking into account requirements of the international projects aimed on collection and sharing of environmental information in Europe, such as the Shared Environmental Information System (SEIS) jointly with the INSPIRE Directive, the Single Information Space in Europe for the Environment (SISE) and the Global Monitoring for Environment and Security (GMES) initiative [6-7].

2 Persistent organic pollutants and the Stockholm Convention

Persistent organic pollutants (POPs) remain in the center of scientific attention due to their slow rates of degradation, their toxicity, and potential for both long-range transport and bioaccumulation in living organisms. This group of compounds covers large number of various chemicals from industrial products like PCBs through pesticides like HCH and DDT till by-products of combustion or chemical processes like PAHs. Due to high importance of POPs the Stockholm Convention on Persistent Organic Pollutants (POPs) entered into force on the May 17, 2004, and has currently 179 signatory parties [4]. The main objective of the Stockholm Convention (SC) is to protect human health and the environment from persistent organic pollutants by reducing or eliminating their releases into the environment. Parties to the Stockholm Convention are required to develop National Implementation Plans to demonstrate their implementation of the SC obligations including arrangements necessary for effectiveness evaluation of adopted measures. The data on POPs monitoring and their utilization in decision support is thus a high priority task in the process of the SC implementation.

3   Data sources for GENASIS system

The GENESIS data browser utilizes the available POPs data from the monitoring networks to obtain as-complete-as-possible set of information and a representative picture of environmental contamination by POPs. The system is developed with taking into mind requirements of the international projects aimed on collection and sharing of environmental information in Europe like Single Information Space in Europe for the Environment (SISE) and the Shared Environmental Information System (SEIS) jointly with the INSPIRE Directive and the Global Monitoring for Environment and Security (GMES) initiative [7]. The development of the GENASIS involves problems of the so called information pyramid: *monitoring* (obtaining environmental data), *data validation* (including uncertainty quantification), *data analysis*, and *methods of aggregation of data* from different sources, *model generalization* and *interpretation of results*, and *their presentation*.

The GENASIS browser [5] offers visualization of data from both national and international monitoring projects. Among the most important are Košetice EMEP station, MONET network and others. The database contains data from 5 projects, 4 continents, 38 countries, 98 compounds and over 500,000 records collected during more than 20 years of continuous monitoring. Range of monitored compounds covers most of those listed under the Stockholm Convention.

4   System description

GENASIS (Global ENvironmetal ASessment and Evaluation System) provides data repository capacities optimized for monitoring of important elements of environment (ambient air, surface water, soil, sediments or even human tissues), statistical processing tools directly connected to data repository, reporting tools and information and publication platform in form of web portal. All tools, features and services are wisely integrated together to create its own "ecosystem".

In the core of system is implemented robust data storage capacity. Data storage must be able to handle big amounts of heterogeneous data and provide fast results for requests of analytical tools and application services which provide data to other system components. Data storage is provided by database of primary data. Compute-intensive data transformations and calculations are preprocessed and stored in data warehouse to provide results as fast as possible. This solution meets general speed requirements and is characterized by good scalability.

On the top of the robust data storage infrastructure is built a set of services with Service Oriented Architecture (SOA) design pattern in mind. Individual services are focused on providing specialized functionality and provide its own interfaces so it is possible to combine and interconnect them in applications which are built over this system layer. These architectural design patterns were extended by integration of additional technologies: ArcGIS Server provides extension for spatial operations and map publishing, server side R is used for advanced statistical computations.

All of these core features are further used in interconnected applications of GENASIS system. These applications provide analytical tools directly connected to data repository and perform statistical computations in real time (Fig. 1). All GENASIS applications are publicly available so anyone can access and evaluate data and use the outputs for his needs. In addition to general data browser, which is described in this paper, subsets of data are used also for specialized case studies [8].
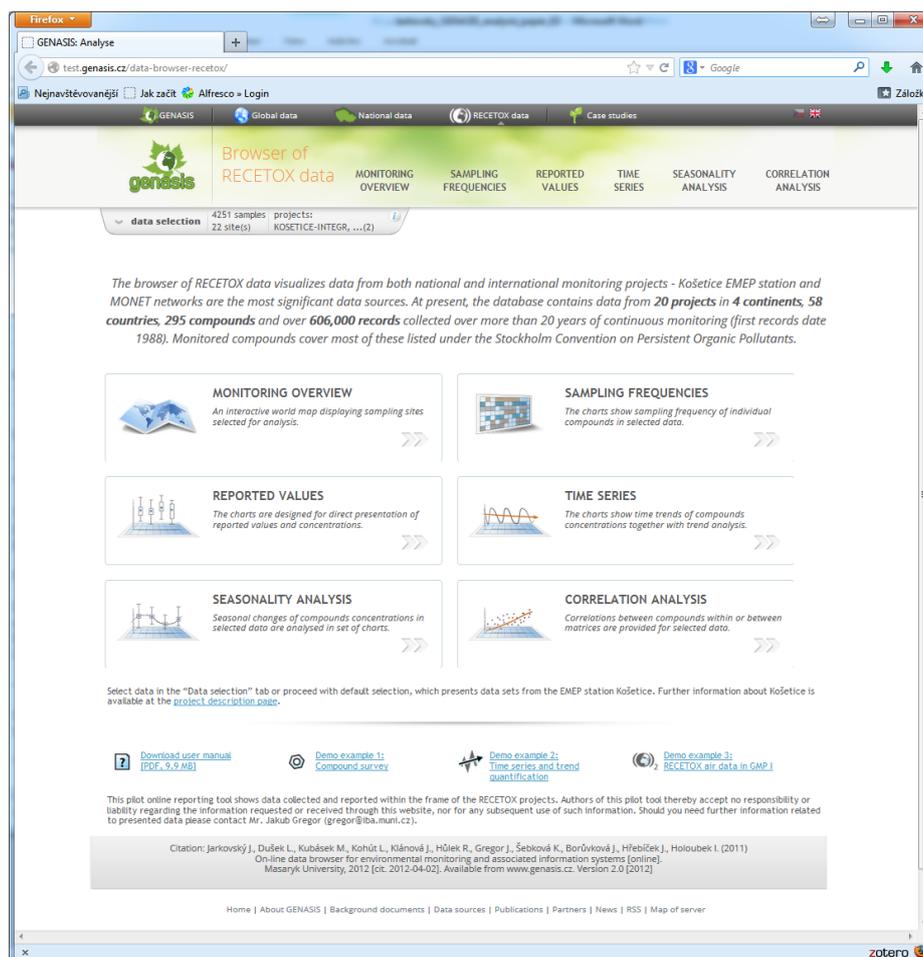

Fig. 1. Entry page of the GENASIS system (Source: www.genasis.cz)

5   Data analysis tools

The primary step of the analysis is the selection of the data using the basic set of characteristics (several predefined selections are also provided): Matrix; Type of monitoring; Project/monitoring network; Time range; Region; Parameter/compound; Individ-

ual sites (Fig. 2). The selected data can be analyzed in the following ways covering all standard requirements for the environmental research in this field.
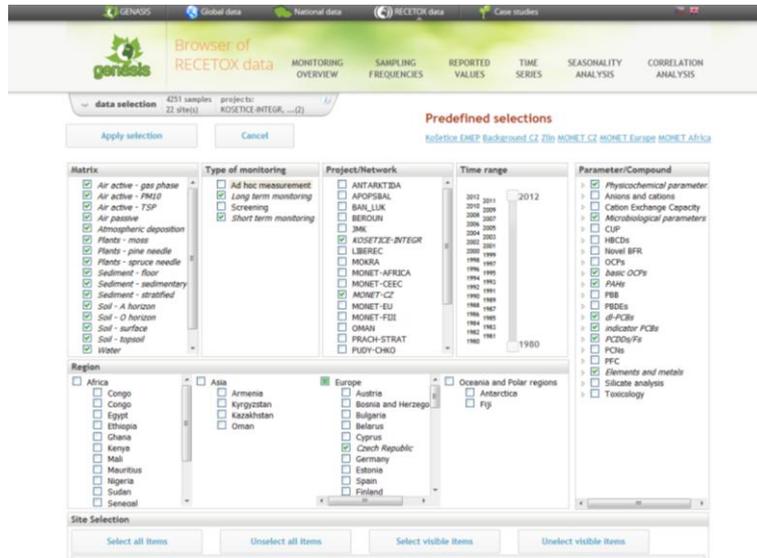

Fig. 2. Selection module of the GENASIS system

Monitoring overview is an interactive world map displaying sampling sites selected for analysis (Fig. 3).
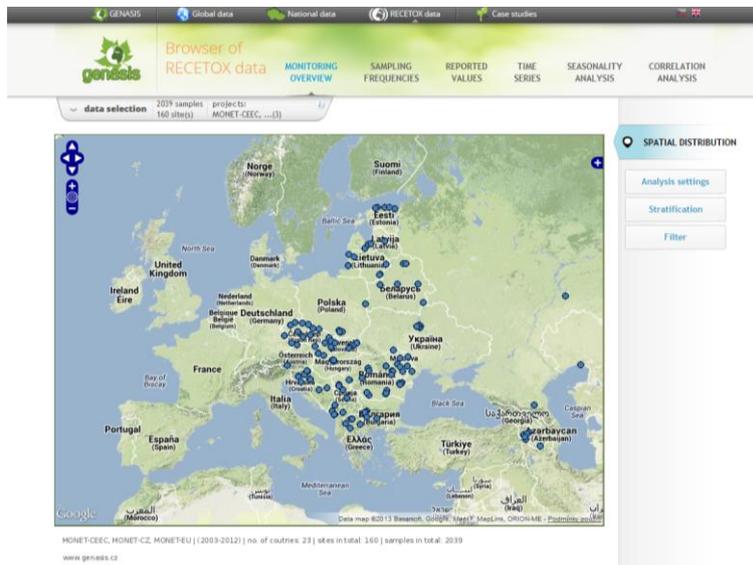

Fig. 3. Overview of the selected sites

Sampling frequencies are charts showing sampling frequency of individual compounds in selected data; the analytical module allows description of single compounds, comparison of compound on several sites or comparison of sampling frequency of multiple compounds (Fig. 4).
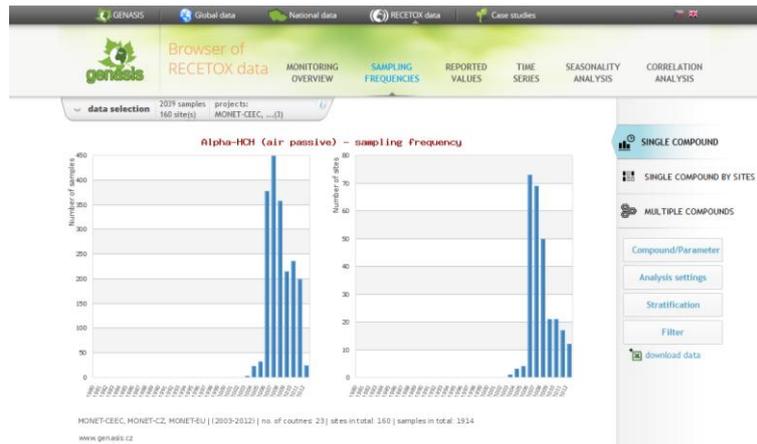

Fig. 4. Overview of sampling frequencies

Reported values charts are designed for direct presentation of reported values and concentrations; several types of charts (histograms, box and whisker plots and charts for individual sites) are provided (Fig. 5).
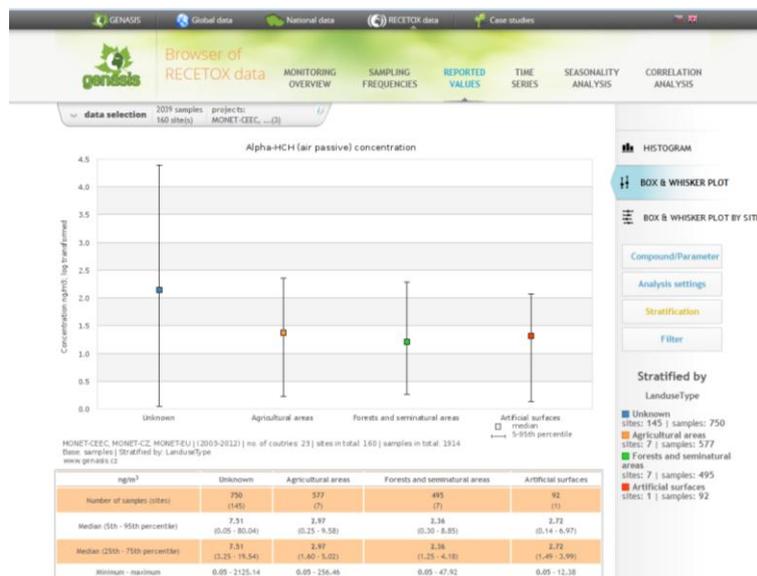

Fig. 5. Visualization of measured concentrations

Time series analysis shows time trends of compounds concentrations together with trend analysis (Fig. 6). The visualization of time series is supported by smoothing and time series data aggregation; trend analysis uses both non parametric (Daniels test, Mann Kendall test) and parametric approach (trend estimate using linear regression model with least square method). Seasonality analysis presents seasonal changes of compounds concentrations in selected data; several levels of time aggregation are possible.

Fig. 6. Time series and trend analysis

Correlation analysis analyses correlations between compounds within or between matrices for selected data; both parametric and nonparametric correlation coefficients are available (Fig. 7).
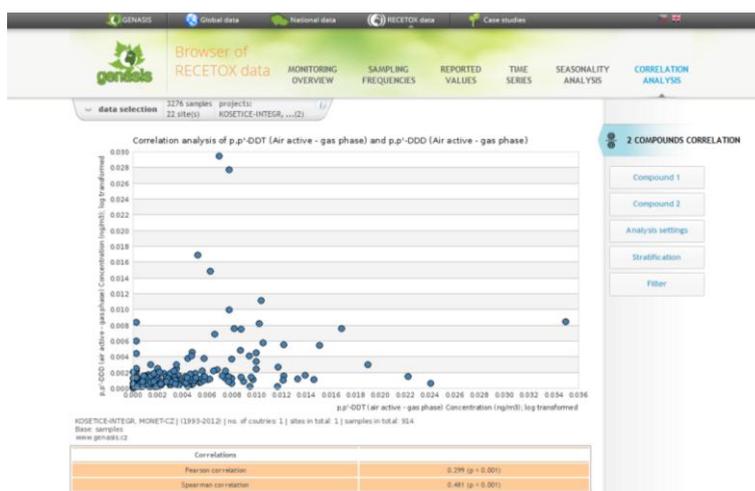

Fig. 7. Correlation analysis

All the analyses share also additional abilities allowing focused analyses:

- Filtering and splitting of selected data according to set of sampling sites characteristics; due to dynamic structure of the database these sampling sites characteristics can be easily enriched by newly available data.
- Several types of data transformation are available allowing user to work with various statistical distribution of data.
- Both parametric and nonparametric descriptive statistics and trend tests are provided.
- Export of the data for charts is available.

6   Conclusion

Within the frame of the GENASIS data browser [9], we introduce a user-friendly system for the visualization and analysis of contamination of all environmental compartments by persistent organic pollutants and evaluation of actual POPs contamination, its long-term trends and seasonal fluctuations. Project outcomes are useful as information source both for lay public and experts, as well as for the process of the Stockholm Convention implementation.

References

1. Luis, G.B.: Principles for evaluating health risks in children associated with exposure to chemicals. Environmental health criteria 237, WHO Library Cataloguing-in-Publication Data. World Health Organization 2006. (2006)
2. JRC ISPRA: 2nd edition of the Technical Guidance Document (TGD) on Risk Assessment of Chemical Substances following European Regulations and Directives. (2007)
3. Holoubek, I., Dušek, L., Klánová, J., Kubásek, M., Jarkovský, J., Baroš, R., Komprdová, K., Bednářová, Z., Hůlek, R., Hřebíček, J.: GENASIS Information System: A Global Environmental Assessment of Persistent Organic Pollutants. In: Hřebíček, J., Schimak, G., Denzer, R. (eds.) 9th IFIP WG 5.11 International Symposium on Environmental Software Systems: Frameworks of eEnvironment, ISESS 2011, pp. 480-485. Springer, Heidelberg (2011)
4. Stockholm Convention on Persistent Organic Pollutants (POPs): Interim Secretariat for the Stockholm Convention, United Nations Environmental Programme (UNEP) Chemicals: Geneva,Switzerland, `http://chm.pops.int`
5. Kubásek, M., Hřebíček, J., Kalina, J., Dušek, L., Urbánek, J., Holoubek, I.: Global Environmental Assessment Requires Global Functional Searching Engines: Robust Application of TaToo Tools. In: Hřebíček, J., Schimak, G., Denzer, R. (eds.) 9th IFIP WG 5.11 International Symposium on Environmental Software Systems: Frameworks of eEnvironment, ISESS 2011, pp. 507-518. Springer, Heidelberg (2011)

6. Hřebíček, J., Pillmann, W.: Shared Environmental Information System and Single Information Space in Europe for the Environment: Antipodes or Associates? In Proceedings of the European conference of the Czech Presidency of the Council of the EU TOWARDS eENVIRONMENT. Opportunities of SEIS and SISE: Integrating Environmental Knowledge in Europe. p. 447-458. Masaryk University, Brno (2009)

7. Hřebíček, J., Pillmann, W.: eEnvironment: Reality and Challenges for eEnvironment Implementation in Europe. In: Hřebíček, J., Schimak, G., Denzer, R. (eds.) 9th IFIP WG 5.11 International Symposium on Environmental Software Systems: Frameworks of eEnvironment, ISESS 2011, pp. 1-14. Springer, Heidelberg (2011)

8. Komprdová, K., Komprda, J., Sáňka, M., Hajek, O.: Spatial Soil Modeling of Organochlorine Pesticides, Their Pools and Volatilization Fluxes. In: Hřebíček, J., Schimak, G., Denzer, R. (eds.) 9th IFIP WG 5.11 International Symposium on Environmental Software Systems: Frameworks of eEnvironment, ISESS 2011, pp. 500-506. Springer, Heidelberg (2011)

9. Kubásek, M., Hřebíček, J., Kalina, J., Dušek, L., Urbánek, J., Holoubek, I.: Global Environmental Assessment Requires Global Functional Searching Engines: Robust Application of TaToo Tools. In: Hřebíček, J., Schimak, G., Denzer, R. (eds.) 9th IFIP WG 5.11 International Symposium on Environmental Software Systems: Frameworks of eEnvironment, ISESS 2011, pp. 507-518. Springer, Heidelberg (2011)