



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

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

www.genasis.cz

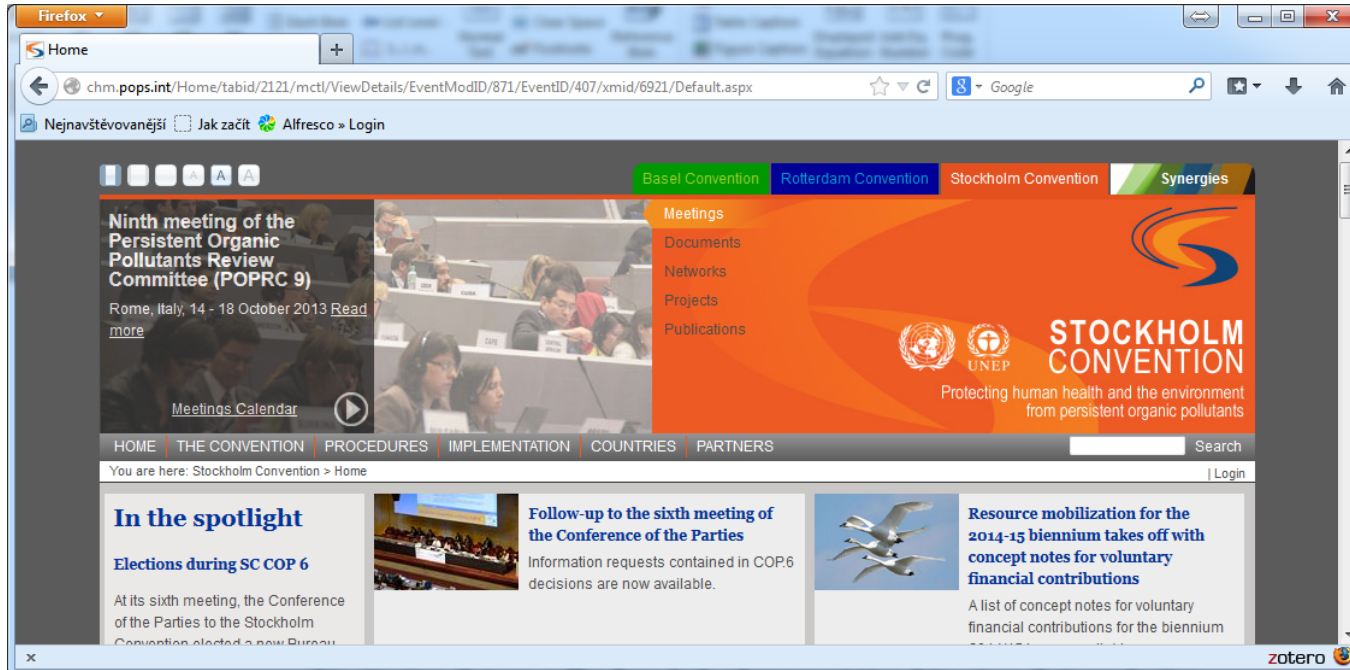
MASARYK UNIVERSITY
Brno, Czech Republic



cejocoen



- Current environmental research is producing large quantities of data from many monitoring programmes
- Decisions based on these data have deep impact on the society – nice example are POPs and measures taken under the Stockholm convention
 - POPs are characterized by their acute and chronic toxicity, carcinogenicity, immunotoxicity, they are persistent and capable of long-range transport and bioaccumulation in animal tissues
- But mostly there is not standardized accessibility of these environmental data !
- **We need comprehensive IT infrastructure which meets these tasks**

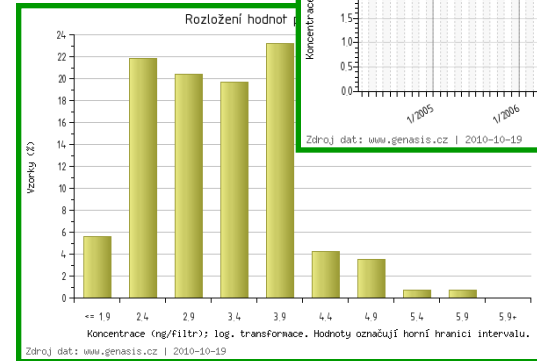
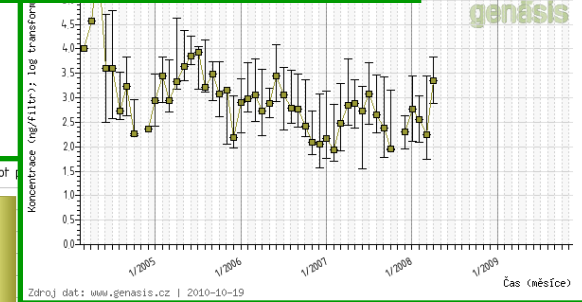
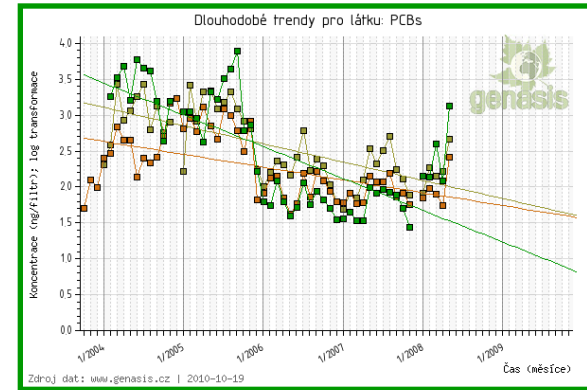
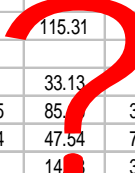




„Data rich – information poor“ never more !

Ukázka uspořádaného datového souboru

číslo	stadium	vek	tran1_3	tran1_4	tran1_5	tran1_6	alb_pbsct	ldh_vstup	sternum	typ_myel
1	3	33	104.36	23.24	104.3	57.77	33	6.02	0.4	IgG
2	3	33	184.88	7.84	105.5	13.82	26	4.01	30	IgG
3	1	34	123.41	9.8	73.3	13.05	32	3.73	45.2	IgG
4	2	43	52.17	6.66	18.03	17.19	42	4.67	40.8	IgG
5	1	45	8.22	2.2	8.22		32	8.25	2	B-J
6	3	46	403.08		115.31		29.7	7.17	38.8	IgA
7	2	49		4.5		1.25	34	4.99	6.4	IgG
8	2	50	33.13	9.64	33.13		35	3.99	14	IgG
9	3	52	257.08	12.05	85.5	3.24	39	12.14	12.2	IgG
10	2	53	78.33	11.34	47.54	7.77	39	5.3	13.6	IgG
11	3	53	61.43	4.67	14.6	3.72	32	4.51	49.2	B-J
12	3	53	135.8	6.7	135.8	59.3	38		26	IgG
13	3	54	129.16	13.33	92.6	38.24	32	4.18	20	IgG
14	3	54	66.89	6.74	33.58	17.3	38	8.44	7.2	B-J
15	3	54	82.86	4.32	18.9	16.4	37	3.6	50	B-J
16	3	55	71.37	6.34	23.91	5.34	43	8.75	27	B-J
17	3	60	14.6	0.9	14.6	11.88	44	5.35	7.5	IgG
18	3	61	94.07	5.62	94.07	1.51	33	4.29	6.4	B-J
19	3	62	86.84	7.53	32.13	2.61	29	4.55	34	IgA



- How to make accessible monitoring outcomes, typically hundreds of separated files, with thousands possible views and statistics ?

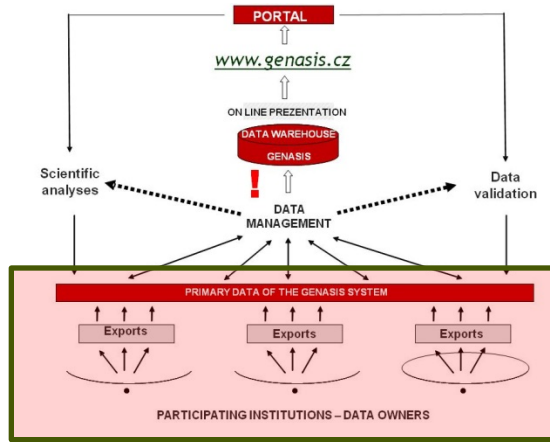


Our proposal and solution:

GENASIS data browser (www.genasis.cz):

Example of monitoring of POPs concentrations

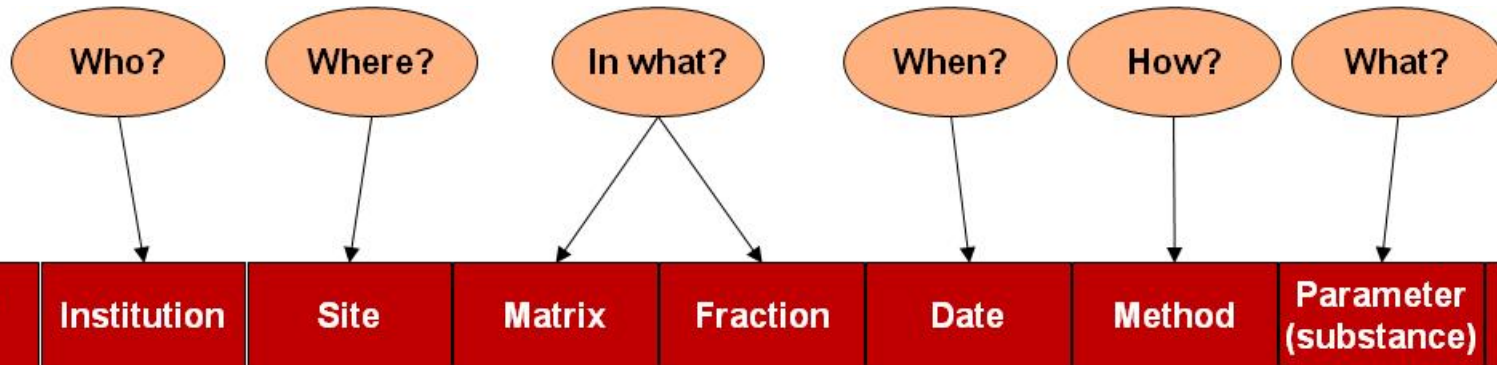




Principle 1. Contract-based participation of partners and safety management



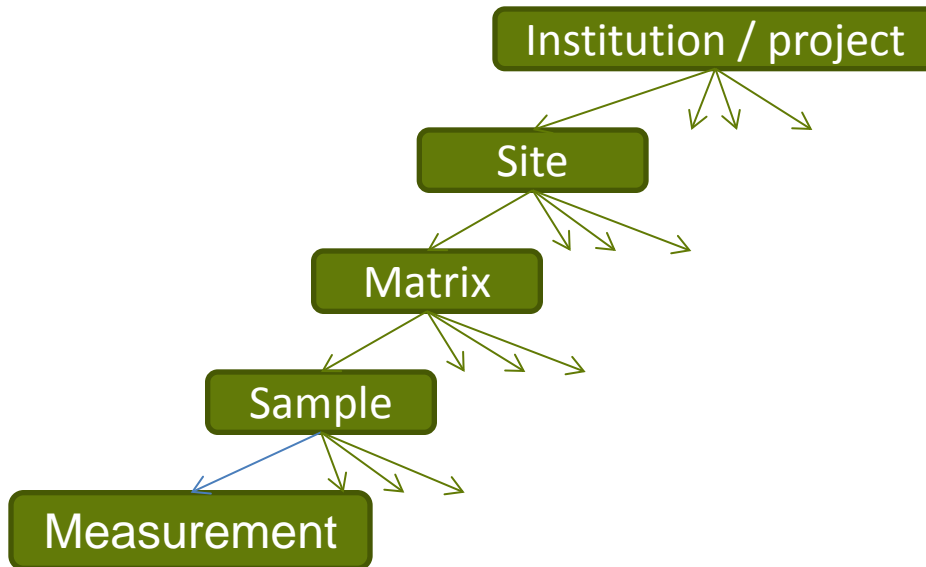
Principle 2. Strictly defined and standardized minimum data record



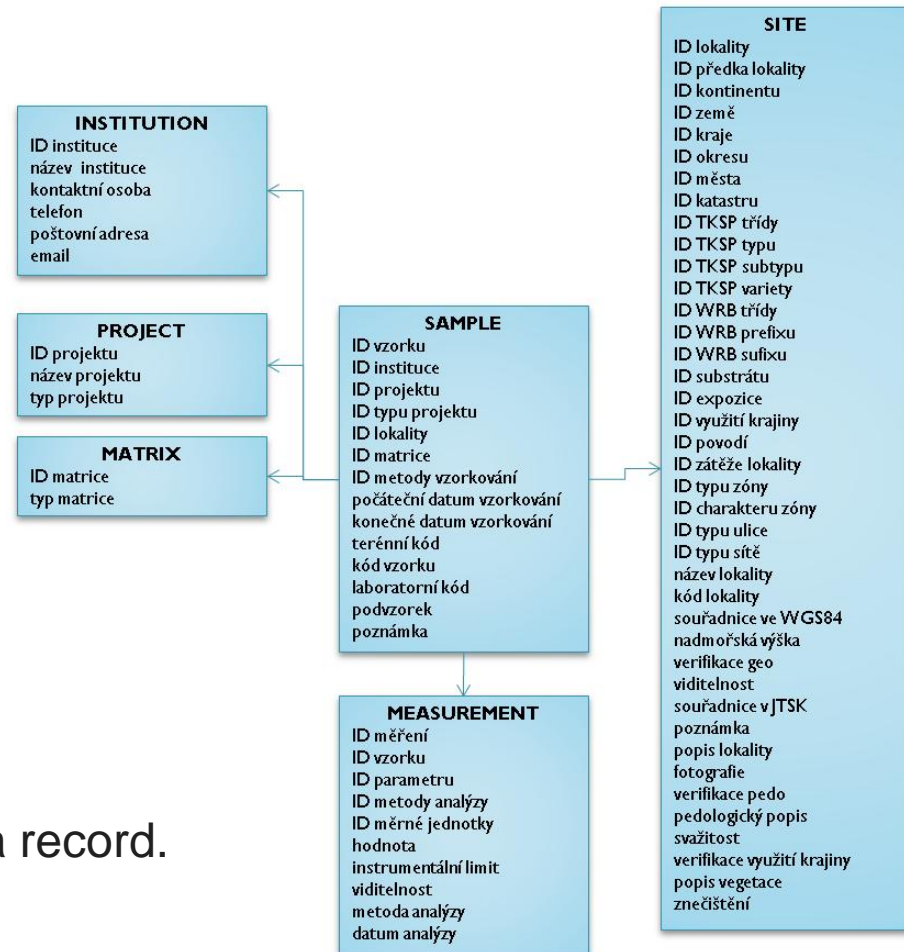


Principle 3.

Multi-layer, hierarchically structured data model



- time
- parameter
- method
- value
- unit

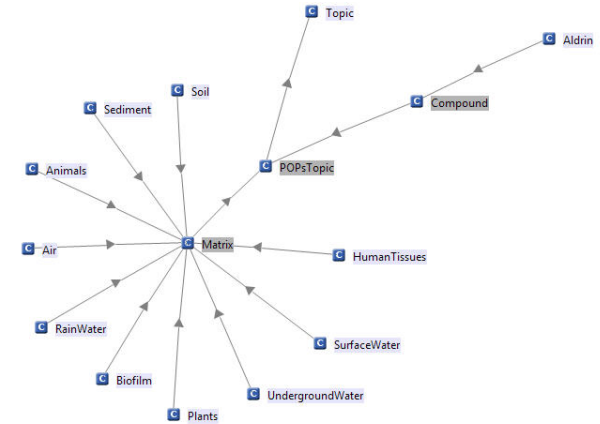
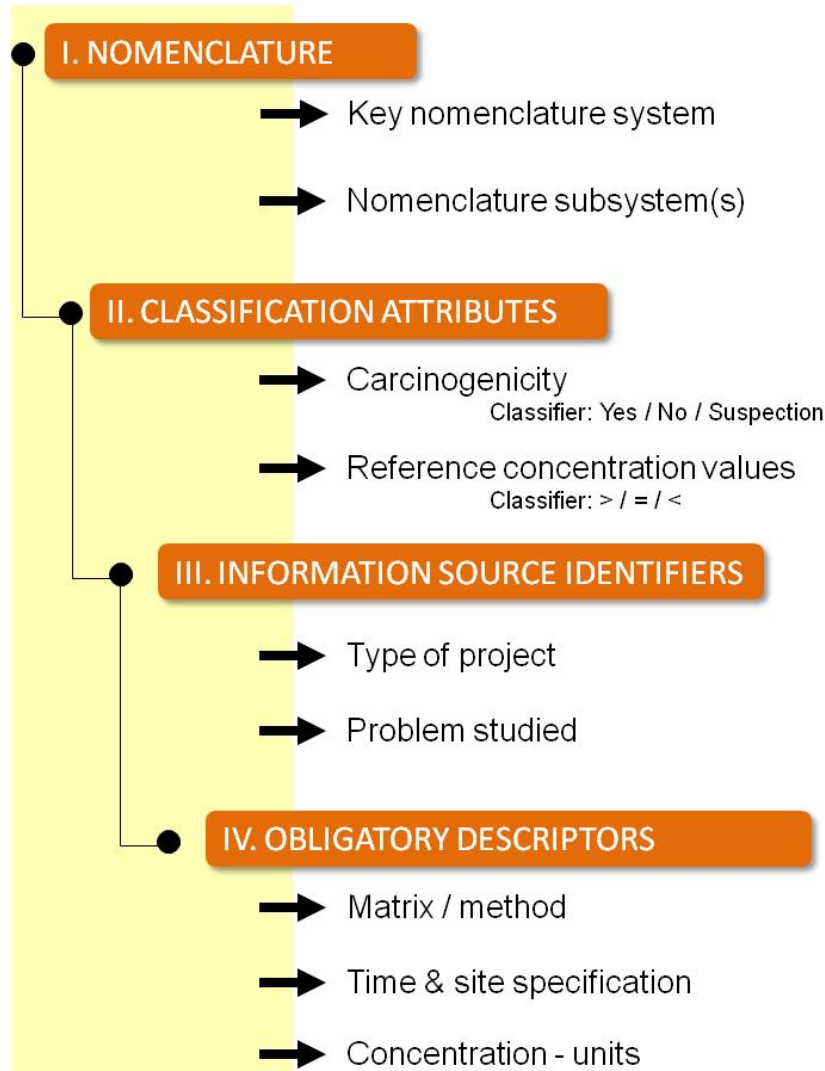


Each level has its own defined minimum data record.



Principle 4.

Well defined ontology and conceptual data model of a repository



Conceptual model enhancing accessibility of data from cancer-related environmental risk assessment studies

Ladislav Dušek^{1,2}, Jiří Hřebíček¹, Miroslav Kubásek¹, Jiří Jarkovský¹, Jiří Kalina¹, Roman Baroš², Zdeňka Bednářová², Jana Klánová², Ivan Holoubek²

¹ Institute of Biostatistics

² Research Centre for To

dusek@iba.muni.cz, hrel
kalina@iba.muni.
klanova

GENASIS Information System: a Global Environmental Assessment of Persistent Organic Pollutants

Ivan Holoubek¹, Ladislav Dušek^{1,2}, Jana Klánová¹, Miroslav Kubásek², Jiří Jarkovský², Roman Baroš¹, Klára Kubošová¹, Zdeňka Bednářová¹, Richard Hůlek², Jiří Hřebíček²

¹ Research Centre for Toxic Compounds in the Environment, Masaryk University, Kamenice 126/3, 625 00 Brno, Czech Republic

² Institute for Biostatistics and Analyses, Masaryk University, Kamenice 126/3, 625 00 Brno, Czech Republic,

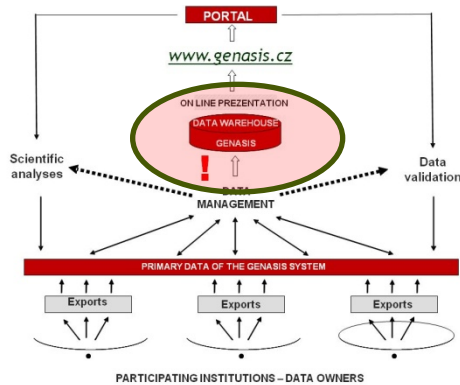
Abstract. This paper describes the discovery, integration and analysis of data from cancer risk studies. Persistent organic pollutants (POPs) are a group of chemicals that do not break down easily in the environment. They can be found in air, water, soil and food. POPs are a major concern because of their potential to cause cancer and other health problems. This paper describes the discovery, integration and analysis of data from cancer risk studies. Persistent organic pollutants (POPs) are a group of chemicals that do not break down easily in the environment. They can be found in air, water, soil and food. POPs are a major concern because of their potential to cause cancer and other health problems. This paper describes the discovery, integration and analysis of data from cancer risk studies. Persistent organic pollutants (POPs) are a group of chemicals that do not break down easily in the environment. They can be found in air, water, soil and food. POPs are a major concern because of their potential to cause cancer and other health problems.

Abstract. Global Environmental Assessment and Information System (GENASIS) is a tool developed by expert teams of the Research Centre for Toxic Compounds in the Environment (RECETOX) and the Institute for Biostatistics and Analyses (IBA) of the Masaryk University in Brno. The aim of GENASIS is to compile validated data on persistent organic pollutants, including their properties, sources, long-term levels, life times, transport mechanisms, effects and risks, scattered throughout various institutions and



Principle 5.

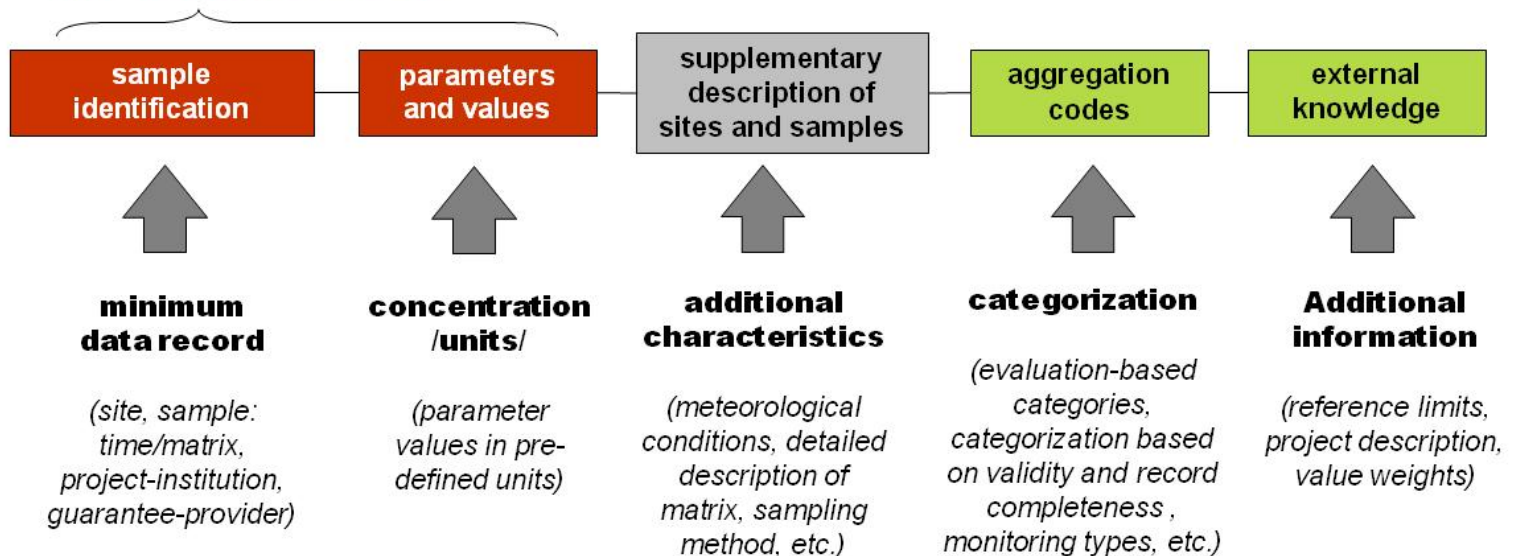
Extended data model in data warehouse and metadata management



Data warehouse operator: pre-processing of entering data:

1. Standardization of items, parameters in records
2. Enrichment of data model by key descriptors (classifiers)
3. Necessary data transformation

Minimum data standard





Statistical processing of POPs data should solve the following tasks:

- 1. Statistical summary = baseline summary statistics** of data and their variability
- 2. Variability / uncertainty analysis**, description of the variability in the data
- 3. Trend analysis** – Identification of variability sources (random processes, seasonality, trend) and their processing prior to trend quantification
 - > **detection**: nonparametric detection of trends
 - > **quantification**: stochastic linear and nonlinear models for trend detection



Simplicity and robustness are the main principles when processing the POPs records.

Non-parametric tests and summary statistics with no or negligible assumptions for the distribution patterns are highly recommended:

- Median estimates supplied with 5th-95th percentile range and geometric mean estimated on the basis of log-transformed data with corresponding 95% confidence interval are recommended for the summary statistics.
- Spearman's rank correlation coefficient is recommended for correlation analysis.
- Mann-Kendall and Daniel's test are recommended for trend detection.



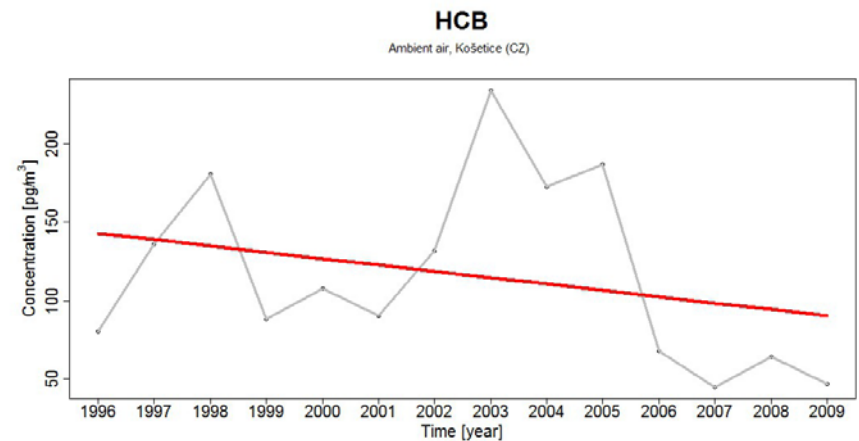
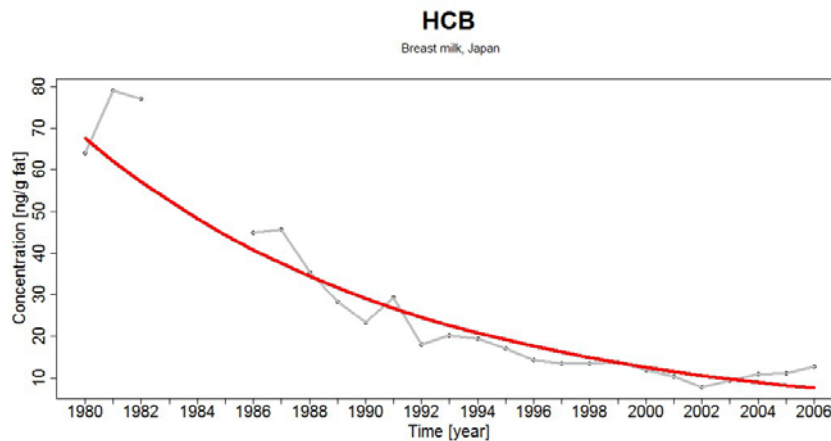
Trend detection and quantification

Short time series

Delta difference between start-end of time series

Long time series

- Nonparametric trend identification (Daniel, Mann-Kendall test)
- Trend quantification using regression models (linear or exponential)
- Delta difference between start-end of time series as supplementary descriptive statistics





The R statistical environment was chosen as a tool for GMP data analysis and visualisation.

- R could be run as a desktop or server application (as a part of a web system), extensive library of statistical functions is used for environmental data evaluation on IBA;
- the environment disposes of a huge pool of ready-made packages for almost all applications in the field of environmental statistics;
- an open source character and recurrent form of R environment allows an easy creation of new functions and packages for sharing experiences;
- R allows to use proven functions and procedures avoiding ambiguities in the data analysis and save a substantial part of the development capacity used for validation and correction processes;
- R algorithms are optimized for quick performance in the case of the web implementation.

GENASIS v. 3.9 – system functions



Firefox

GENASIS: Analyse

www.genasis.cz/data-browser/

Nejnavštěvovanější Jak začít Alfresco » Login

GENASIS Global data National data RECETOX data Case studies

genesis Browser of RECETOX data

MONITORING OVERVIEW SAMPLING FREQUENCIES REPORTED VALUES TIME SERIES SEASONAL AGGREGATION CORRELATION ANALYSIS

data selection 1963 samples 1 site(s) projects: KOSETICE-INTEGR, ...(2)

The browser of RECETOX data visualizes data from both national and international monitoring projects - Košetice EMEP station and MONET networks are the most significant data sources. At present, the database contains data from **6 projects in 4 continents, 55 countries, 204 compounds** and over **412,000 records** collected over more than 20 years of continuous monitoring (first records date 1988). Monitored compounds cover most of these listed under the Stockholm Convention on Persistent Organic Pollutants.

MONITORING OVERVIEW
An interactive world map displaying sampling sites selected for analysis.

SAMPLING FREQUENCIES
The charts show sampling frequency of individual compounds in selected data.

REPORTED VALUES
The charts are designed for direct presentation of reported values and concentrations.

TIME SERIES
The charts show time trends of compounds concentrations together with trend analysis.

SEASONAL AGGREGATION
Seasonal changes of compounds concentrations in selected data are analysed in set of charts.

CORRELATION ANALYSIS
Correlations between compounds within or between matrices are provided for selected data.

Select data in the "Data selection" tab or proceed with default selection, which presents data sets from the EMEP station Košetice. Further information about Košetice is available at the [project description page](#).

zotero

Data selection



GENASIS Global data National data RECETOX data Case studies

genasis Browser of RECETOX data MONITORING OVERVIEW SAMPLING FREQUENCIES REPORTED VALUES TIME SERIES SEASONALITY ANALYSIS CORRELATION ANALYSIS

data selection 4251 samples 22 site(s) projects: KOSETICE-INTEGR, ...(2)

Apply selection Cancel

Predefined selections

[Košetice EMEP](#) [Background CZ](#) [Zlin](#) [MONET CZ](#) [MONET Europe](#) [MONET Africa](#)

Matrix	Type of monitoring	Project/Network	Time range	Parameter/Compound
<input checked="" type="checkbox"/> Air active - gas phase <input checked="" type="checkbox"/> Air active - PM10 <input checked="" type="checkbox"/> Air active - TSP <input checked="" type="checkbox"/> Air passive <input checked="" type="checkbox"/> Atmospheric deposition <input checked="" type="checkbox"/> Plants - moss <input checked="" type="checkbox"/> Plants - pine needle <input checked="" type="checkbox"/> Plants - spruce needle <input checked="" type="checkbox"/> Sediment - floor <input checked="" type="checkbox"/> Sediment - sedimentary <input checked="" type="checkbox"/> Sediment - stratified <input checked="" type="checkbox"/> Soil - A horizon <input checked="" type="checkbox"/> Soil - O horizon <input checked="" type="checkbox"/> Soil - surface <input checked="" type="checkbox"/> Soil - topsoil <input checked="" type="checkbox"/> Water	<input type="checkbox"/> Ad hoc measurement <input checked="" type="checkbox"/> Long term monitoring <input type="checkbox"/> Screening <input checked="" type="checkbox"/> Short term monitoring	<input type="checkbox"/> ANTARKTIDA <input type="checkbox"/> APOPSBAL <input type="checkbox"/> BAN_LUK <input type="checkbox"/> BEROUN <input type="checkbox"/> JMK <input checked="" type="checkbox"/> KOSETICE-INTEGR <input type="checkbox"/> LIBEREC <input type="checkbox"/> MOKRA <input type="checkbox"/> MONET-AFRICA <input type="checkbox"/> MONET-CEEC <input checked="" type="checkbox"/> MONET-CZ <input type="checkbox"/> MONET-EU <input type="checkbox"/> MONET-FJII <input type="checkbox"/> OMAN <input type="checkbox"/> PRACH-STRAT <input type="checkbox"/> PUDY-CHKO	2012 2011 2010 2009 2008 2007 2006 2005 2004 2003 2002 2001 2000 1999 1998 1997 1996 1995 1994 1993 1992 1991 1990 1989 1988 1987 1986 1985 1984 1983 1982 1981 1980	<input checked="" type="checkbox"/> Physicochemical parameter. <input type="checkbox"/> Anions and cations <input type="checkbox"/> Cation Exchange Capacity <input checked="" type="checkbox"/> Microbiological parameters <input type="checkbox"/> CUP <input type="checkbox"/> HBCDs <input type="checkbox"/> Novel BFR <input type="checkbox"/> OCPs <input checked="" type="checkbox"/> basic OCPs <input checked="" type="checkbox"/> PAHs <input type="checkbox"/> PBB <input type="checkbox"/> PBDEs <input checked="" type="checkbox"/> dl-PCBs <input checked="" type="checkbox"/> indicator PCBs <input checked="" type="checkbox"/> PCDDs/Fs <input type="checkbox"/> PCNs <input type="checkbox"/> PFC <input checked="" type="checkbox"/> Elements and metals <input type="checkbox"/> Silicate analysis <input type="checkbox"/> Toxicology

Region

Africa	Asia	Europe	Oceania and Polar regions
<input type="checkbox"/> Congo <input type="checkbox"/> Congo <input type="checkbox"/> Egypt <input type="checkbox"/> Ethiopia <input type="checkbox"/> Ghana <input type="checkbox"/> Kenya <input type="checkbox"/> Mali <input type="checkbox"/> Mauritius <input type="checkbox"/> Nigeria <input type="checkbox"/> Sudan <input type="checkbox"/> Senegal	<input type="checkbox"/> Armenia <input type="checkbox"/> Kyrgyzstan <input type="checkbox"/> Kazakhstan <input type="checkbox"/> Oman	<input type="checkbox"/> Austria <input type="checkbox"/> Bosnia and Herzego <input type="checkbox"/> Bulgaria <input type="checkbox"/> Belarus <input type="checkbox"/> Cyprus <input checked="" type="checkbox"/> Czech Republic <input type="checkbox"/> Germany <input type="checkbox"/> Estonia <input type="checkbox"/> Spain <input type="checkbox"/> Finland	<input type="checkbox"/> Antarctica <input type="checkbox"/> Fiji

Site Selection

Select all items Unselect all items Select visible items Unelect visible items

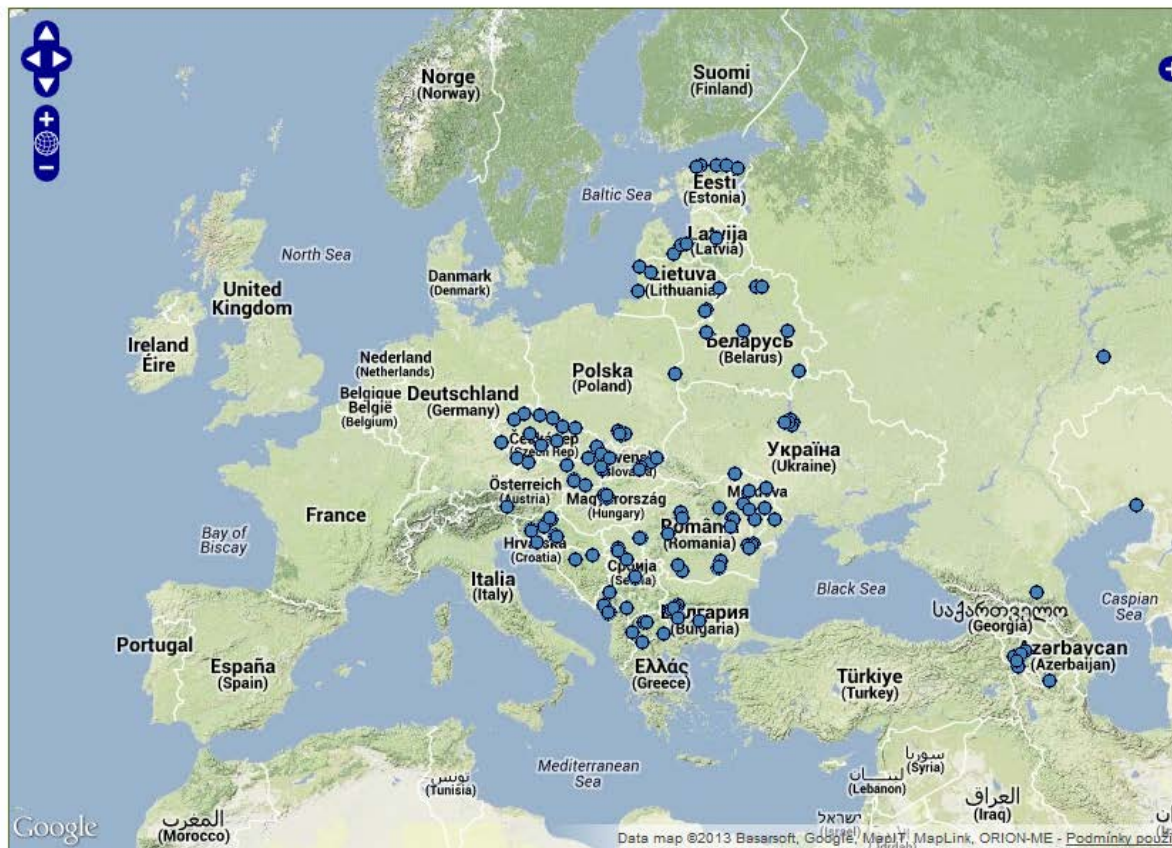
Overview of selected sites



GENASIS Global data National data RECETOX data Case studies

genasis Browser of RECETOX data **MONITORING OVERVIEW** SAMPLING FREQUENCIES REPORTED VALUES TIME SERIES SEASONALITY ANALYSIS CORRELATION ANALYSIS

data selection 2039 samples 160 site(s) projects: MONET-CEEC, ... (3)



SPATIAL DISTRIBUTION

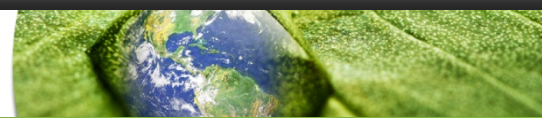
Analysis settings

Stratification

Filter

MONET-CEEC, MONET-CZ, MONET-EU | (2003-2012) | no. of countries: 23 | sites in total: 160 | samples in total: 2039

Overview of sampling frequencies

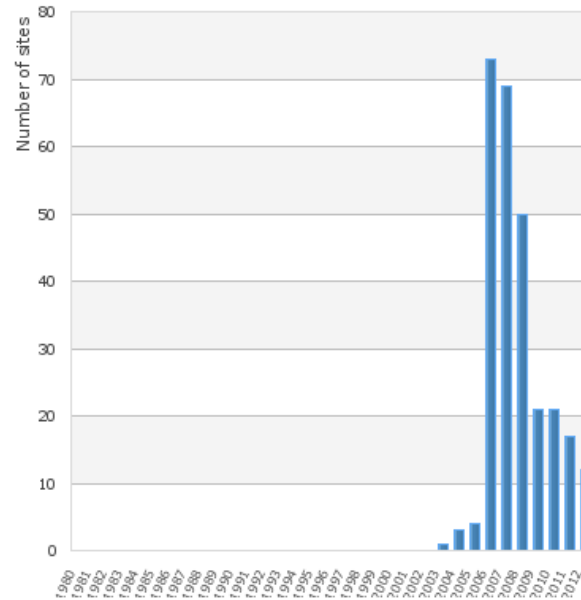
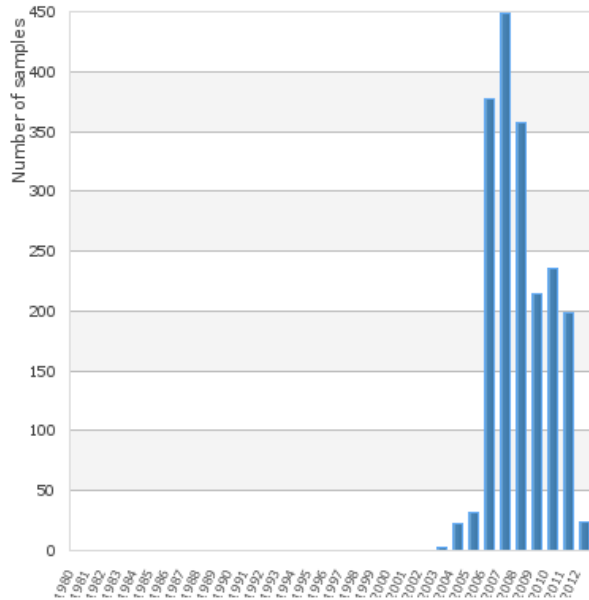


GENESIS Global data National data RECETOX data Case studies

genasis Browser of RECETOX data MONITORING OVERVIEW **SAMPLING FREQUENCIES** REPORTED VALUES TIME SERIES SEASONALITY ANALYSIS CORRELATION ANALYSIS

data selection 2039 samples 160 site(s) projects: MONET-CEEC, ... (3)

Alpha-HCH (air passive) – sampling frequency



- SINGLE COMPOUND**
- SINGLE COMPOUND BY SITES
- MULTIPLE COMPOUNDS

Compound/Parameter

Analysis settings

Stratification

Filter

download data

MONET-CEEC, MONET-CZ, MONET-EU | (2003-2012) | no. of countries: 23 | sites in total: 160 | samples in total: 1914

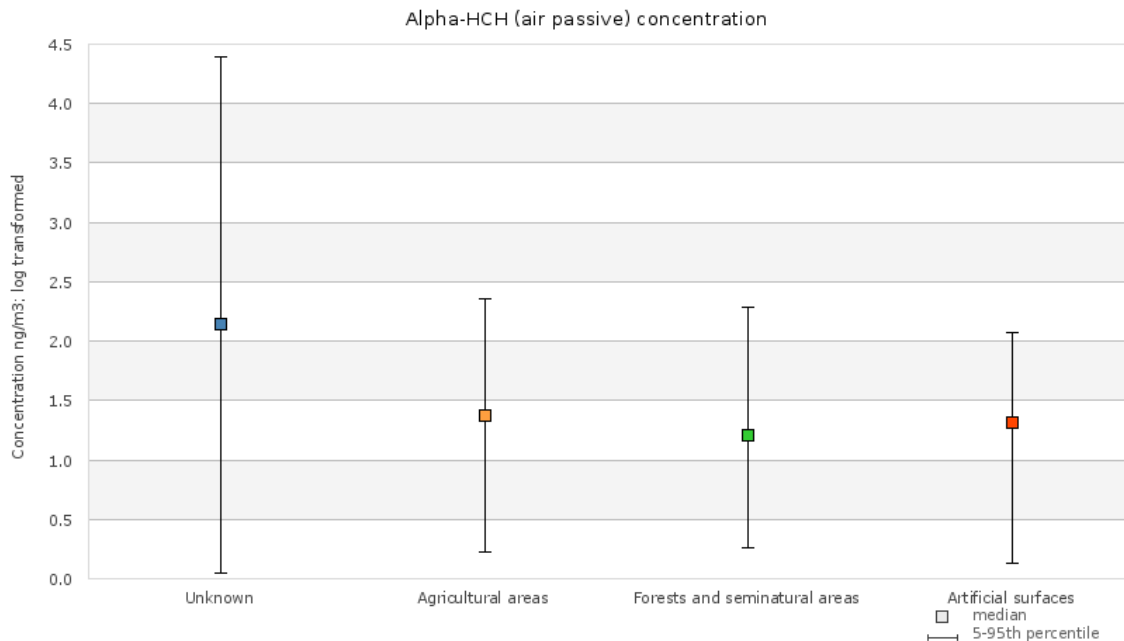
Visualisation of measured concentrations



GENESIS Global data National data RECETOX data Case studies

genesis **Browser of RECETOX data** MONITORING OVERVIEW SAMPLING FREQUENCIES **REPORTED VALUES** TIME SERIES SEASONALITY ANALYSIS CORRELATION ANALYSIS

data selection 2039 samples 160 site(s) projects: MONET-CEEC, ... (3)



MONET-CEEC, MONET-CZ, MONET-EU | (2003-2012) | no. of countries: 23 | sites in total: 160 | samples in total: 1914
Base: samples | Stratified by: LanduseType
www.genesis.cz

ng/m ³	Unknown	Agricultural areas	Forests and seminatural areas	Artificial surfaces
Number of samples (sites)	750 (145)	577 (7)	495 (7)	92 (1)
Median (5th - 95th percentile)	7.51 (0.05 - 80.04)	2.97 (0.25 - 9.58)	2.36 (0.30 - 8.85)	2.72 (0.14 - 6.97)
Median (25th - 75th percentile)	7.51 (3.25 - 19.54)	2.97 (1.60 - 5.02)	2.36 (1.25 - 4.18)	2.72 (1.49 - 3.99)
Minimum - maximum	0.05 - 2125.14	0.05 - 256.46	0.05 - 47.92	0.05 - 12.38

HISTOGRAM

BOX & WHISKER PLOT

BOX & WHISKER PLOT BY SITES

Compound/Parameter

Analysis settings

Stratification

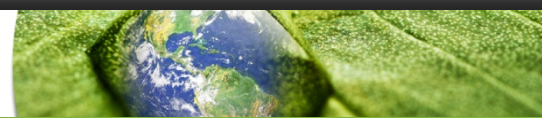
Filter

Stratified by

LanduseType

- Unknown sites: 145 | samples: 750
- Agricultural areas sites: 7 | samples: 577
- Forests and seminatural areas sites: 7 | samples: 495
- Artificial surfaces sites: 1 | samples: 92

Time series and trend analysis

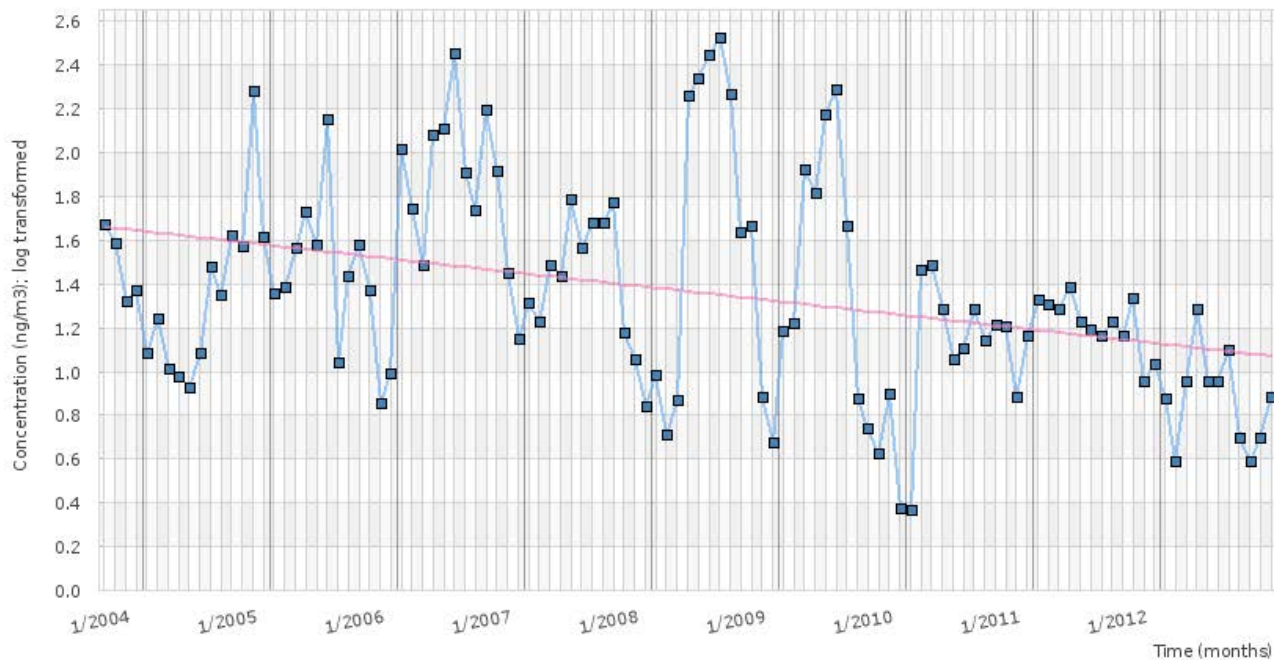


GENASIS Global data National data RECETOX data Case studies

genesis Browser of RECETOX data MONITORING OVERVIEW SAMPLING FREQUENCIES REPORTED VALUES **TIME SERIES** SEASONALITY ANALYSIS CORRELATION ANALYSIS

data selection 2039 samples 160 site(s) projects: MONET-CEEC, ... (3)

Alpha-HCH (air passive) time series trend



Σ SUMMARY

→ TRENDS

Compound/Parameter

Analysis settings

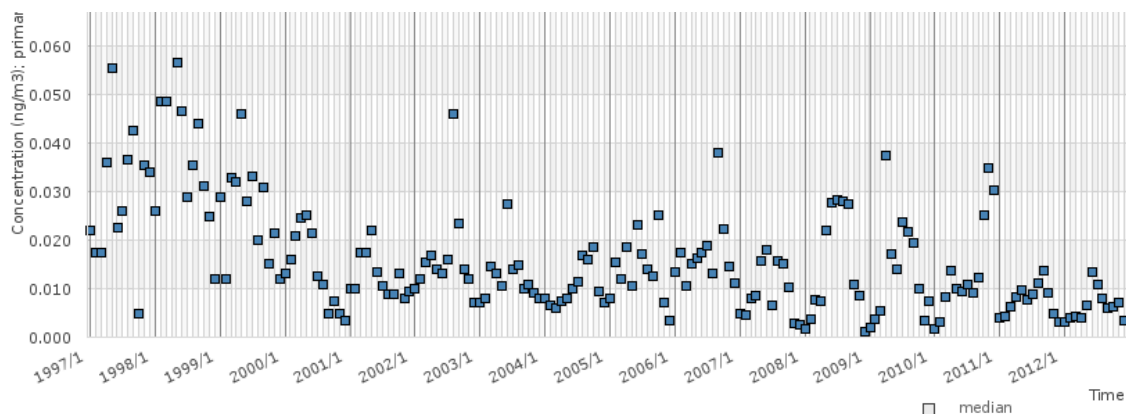
Stratification

Filter

MONET-CEEC, MONET-CZ, MONET-EU | (2003-2012) | no. of countries: 23 | sites in total: 160 | samples in total: 1914
Base: samples
www.genasis.cz

□ median

Time series and trend analysis



KOSETICE-INTEGR | (1997-2012) | no. of countries: 1 | sites in total: 1 | samples in total: 829
 Base: sites
 www.genasis.cz

Median (5th - 95th percentile)	0.0122 (0.0021 - 0.0510) ng/m ³
Median (25th - 75th percentile)	0.0122 (0.0070 - 0.0220) ng/m ³
Minimum - maximum	0.0003 - 0.2120 ng/m ³
Geometric mean (95% CI)	0.0117 (0.0016 - 0.0880) ng/m ³
Arithmetic mean (95% CI)	0.0183 (0.0000 - 0.0574) ng/m ³
Trend tests	
Daniels test ⁱ	-0.511 (p < 0.001)
Mann - Kendall test ⁱ	-0.356 (p ≥ 0.05)
Trend estimates	
Difference between beginning/end of time series 1/1997 - 12/2012 ⁱ	-0.0185 ng/m ³ (-84.1045%)
Annual change based on difference between beginning/end of time series ⁱ	-0.0012 ng/m ³ (-5.2841%)
Annual change estimated using least square method ⁱ	-0.0013 ng/m ³ (-6.0911%)

Citation: Jarkovský J., Dušek L., Kubásek M., Kohút L., Klánová J., Hůlek R., Gregor J., Šebková K., Borůvková J., Hřebíček J., Holoubek I. (2011) On-line data browser for environmental monitoring and associated information systems [online]. Masaryk University, 2012 [cit. 2012-04-02]. Available from www.genasis.cz. Version 3.9. (only air monitoring) - data export June 2013 [2013]

[Home](#) | [About GENASIS](#) | [Background documents](#) | [Data sources](#) | [Publications](#) | [Partners](#) | [News](#) | [RSS](#) | [Map of server](#)

© GENASIS ISSN 1805-3181 | Research Centre for Toxic Compounds in the Environment MU & Institute of Biostatistics and Analyses MU

Project under the auspices of CETOCOEN, OPVaVPI, Ministry of the Environment of the Czech Republic

valid XHTML 1.0 Strict

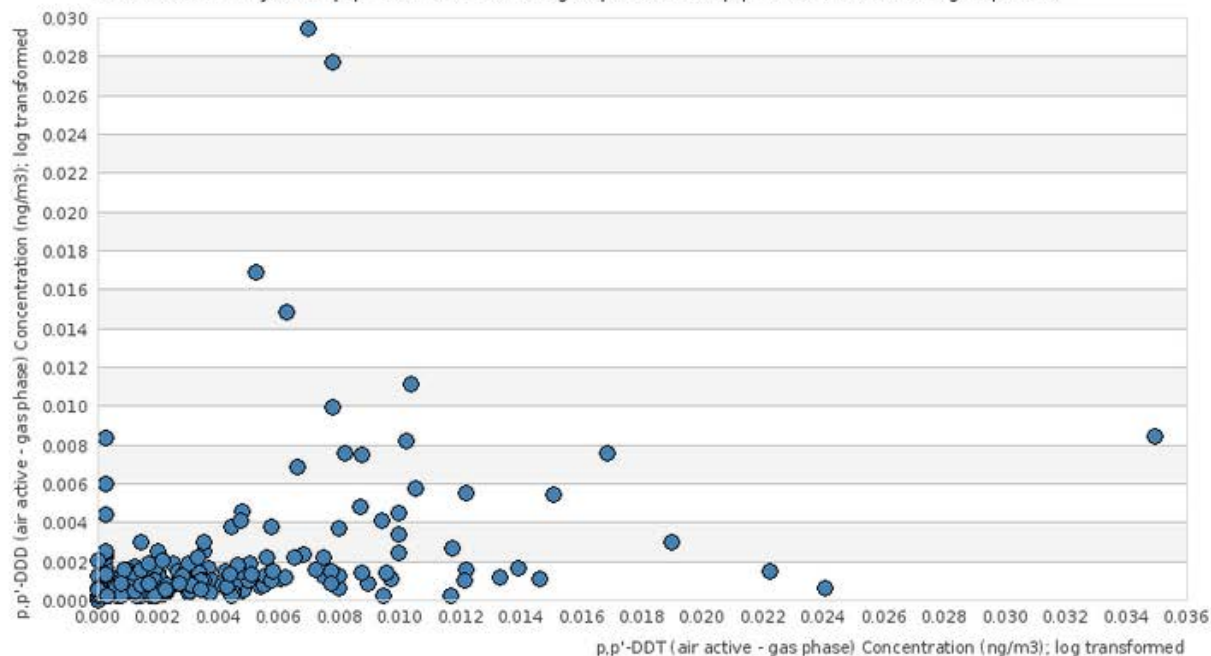
Correlation analysis



Browser of RECETOX data

[MONITORING OVERVIEW](#)[SAMPLING FREQUENCIES](#)[REPORTED VALUES](#)[TIME SERIES](#)[SEASONALITY ANALYSIS](#)[CORRELATION ANALYSIS](#)**data selection**3276 samples
22 site(s)projects:
KOSETICE-INTEGR, ... (2)

Correlation analysis of p,p'-DDT (Air active - gas phase) and p,p'-DDD (Air active - gas phase)



KOSETICE-INTEGR, MONET-CZ | (1993-2012) | no. of countries: 1 | sites in total: 1 | samples in total: 914
Base: samples
www.genesis.cz

2 COMPOUNDS CORRELATION

Correlations

Pearson correlation

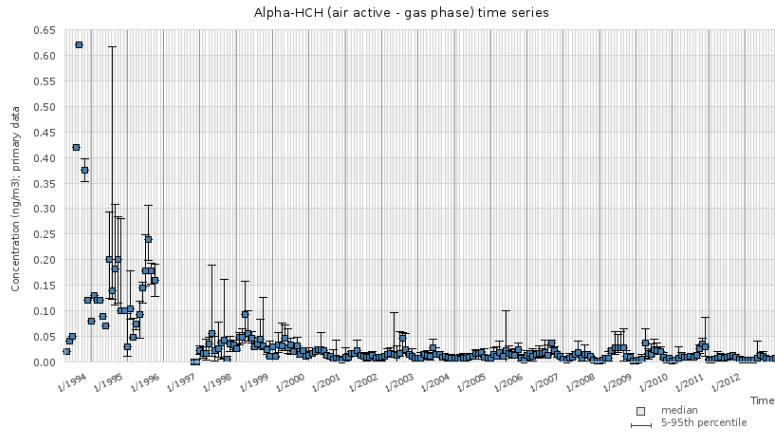
0.299 (p < 0.001)

Spearman correlation

0.481 (p < 0.001)



Graphical and tabular outputs



KOSETICE-INTEGR | [1993-2012] | no. of countries: 1 | sites in total: 1 | samples in total: 914
 Base: samples
 www.genasis.cz

Show Summary table

Median (5th - 95th percentile)	0.0122 (0.0021 - 0.0510) ng/m ³
Median (25th - 75th percentile)	0.0122 (0.0070 - 0.0220) ng/m ³
Minimum - maximum	0.0003 - 0.2120 ng/m ³
Geometric mean (95% CI)	0.0117 (0.0016 - 0.0880) ng/m ³
Arithmetic mean (95% CI)	0.0183 (0.0000 - 0.0574) ng/m ³
Trend tests	
Daniels test ⁱ	-0.511 (p < 0.001)
Mann - Kendall test ⁱ	-0.356 (p ≥ 0.05)
Trend estimates	
Difference between beginning/end of time series 1/1997 - 12/2012 ⁱ	-0.0185 ng/m ³ (-84.1045%)
Annual change based on difference between beginning/end of time series ⁱ	-0.0012 ng/m ³ (-5.2841%)
Annual change estimated using least square method ⁱ	-0.0013 ng/m ³ (-6.0911%)

Export of outputs

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

Analysis specific settings

data transformation

parametric and nonparametric statistics and trend tests are provided



- We introduce a user-friendly system for the visualization and analysis of contamination of all environmental compartments by persistent organic pollutants
- 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
- Important IT tool for the process of the Stockholm Convention implementation.



CONCLUSION: Standardized IT services can help to skip the gap between heterogeneous primary data and on-line widely accessible reporting

**Thank you very much
for your kind attention**

