



Midterm regional workshop for GMP2 project in Africa 2018, Lusaka, Zambia, 23-25 July 2018

GMP Data Warehouse: current data content and functionalities

Kateřina Šebková

on behalf of the GMP DWH development team

Research Centre for Toxic Compounds in the Environment & Institute of Biostatistics and Analyses Masaryk University, Brno, Czech Republic



Outline

- Why do we have GMP data warehouse (GMP DWH)?
- Principles beyond GMP DWH
- How to access GMP DWH?
- What is its current content global information on POP levels
- OBJECTIVE: provide additional information on a tool for global POPs data access and on its functionalities and use to countries in Africa region



capacity building project on POPs monitoring (GMP2)





Research Centre for Toxic Compounas in the Environment

SCENE SETTING

- POPs monitoring embedded in the Stockholm Convention text (art 16 and 15)
- regional/global POPs monitoring reports serve as one of the key inputs for assessing effectiveness of the Stockholm Convention (follow up of the BRS presentation)
- Two sets of the **Global Monitoring Plan regional reports** (2008 and 2014) approved by COP4 (2009) and COP6 (2015) available online: http://chm.pops.int/Programmes/GlobalMonitoringPlan/MonitoringReports/tabid/525/Default.aspx
- but no harmonization of reports (content) and more user friendly tool suitable for 21st century existed in 2011 - BRS requested support to develop an electronic tool to SCRC Czech Republic RECETOX in 2011 and the first proposal was prepared in 2012 and refined further to be inaugurated in 2013 at the COP6 in Geneva.







Objective of the GMP Data Warehouse:

- Collection, storage, analysis, and visualization of GMP data on POP levels as required by the Stockholm Convention on POPs
- Put global data on POPs together
- Support to ROGs and GCG in compilation of regional/global monitoring reports
- Online presentation of GMP data publicly available 24/7
- Useful to countries
- allow for data collection from heterogeneous sources
- also allow input from existing international and national monitoring programmes











Substances to be monitored: GMP Guidance, chapter 2.2. table 2.2.

	Compounds to be Monitored			
	Air	Human Milk	Human Blood	Water
Initial POPs				
Aldrin	Aldrin	Aldrin	Aldrin	
Chlordane	<i>cis-</i> and <i>trans-</i> chlordane; and <i>cis-</i> and <i>trans-</i> nonachlor, oxychlordane	<i>cis</i> - and <i>trans</i> -chlordane; and <i>cis</i> - and <i>trans</i> -nonachlor, oxychlordane	<i>cis-</i> and <i>trans-</i> chlordane; and <i>cis-</i> and <i>trans-</i> nonachlor, oxychlordane	_
DDT	4,4'-DDT, 2,4'-DDT and 4,4'-DDE, 2,4'-DDE, 4,4'- DDD, 2,4'-DDD	4,4'-DDT, 2,4'-DDT and 4,4'-DDE, 2,4'-DDE, 4,4'- DDD, 2,4'-DDD	4,4'-DDT, 2,4'-DDT and 4,4'-DDE, 2,4'-DDE, 4,4'- DDD, 2,4'-DDD	
Dieldrin	Dieldrin	Dieldrin	Dieldrin	
Endrin	Endrin	Endrin	Endrin	
HCB	НСВ	HCB	HCB	Water has not been recommended as a core
Heptachlor	Heptachlor and heptachlorepoxide	Heptachlor and heptachlorepoxide	Heptachlor and heptachlorepoxide	matrix for the lipophilic and nonpolar initial twelve POPs;
Mirex	Mirex	Mirex	Mirex	therefore, analysis of surface waters is not recommended
РСВ	ΣPCB ₆ (6 congeners): 28, 52, 101, 138, 153, and 180	ΣPCB ₆ (6 congeners): 28, 52, 101, 138, 153, and 180	ΣPCB ₆ (6 congeners): 28, 52, 101, 138, 153, and 180	
	PCB with TEFs* (12 congeners): 77, 81, 105, 114, 118, 123, 126, 156, 157, 167, 169, and 189	PCB with TEFs* (12 congeners): 77, 81, 105, 114, 118, 123, 126, 156, 157, 167, 169, and 189	PCB with TEFs* (12 congeners): 77, 81, 105, 114, 118, 123, 126, 156, 157, 167, 169, and 189	
PCDD/PCDF	2,3,7,8-substituted PCD/PCDF (17 congeners)	2,3,7,8-substituted PCD/PCDF (17 congeners)	2,3,7,8-substituted PCD/PCDF (17 congeners)	-
Toxaphene	Congeners P26, P50, P62	Congeners P26, P50, P62	Congeners P26, P50, P62	





Substances to be monitored: GMP Guidance, chapter 2.2. table 2.2.

POPs listed at COP-4				
	Air	Human Milk	Human Blood	Water
Chlordecone	econe Chlordecone Chlordecone		Chlordecone	
α-НСН	α-НСН	α-НСН	α-НСН	
β-НСН	β-НСН	β-НСН	β-НСН	
γ-НСН	ү-НСН	γ-НСН	γ-НСН	
Hexabromobiphenyl	PBB 153	PBB 153	PBB 153	
Pentachlorobenzene	PeCBz	PeCBz	PeCBz	
c-penta BDE c-octa BDE	BDE 47, 99, 153, 154, 175/183 (co-eluting) Optional: BDE 17, 28, 100	BDE 47, 99, 153, 154, 175/183 (co-eluting) Optional: BDE 100	BDE 47, 99, 153, 154, 175/183 (co-eluting) Optional: BDE 100	
PFOS ⁶ PFOS, NMeFOSA, P		PFOS (linear and sum of PFOS)	PFOS (linear and sum of PFOS)	PFOS (linear and sum of PFOS)
POPs listed at COP-5				
Endosulfan	α-, β-endosulfan; and endosulfan sulfate	α-, β-endosulfan; and endosulfan sulfate	α-, β-endosulfan; and endosulfan sulfate	
POPs listed at COP-6				
HBCD	α-HBCD, β-HBCD, γ-HBCD	α-HBCD, β-HBCD, γ-HBCD	α-HBCD, β-HBCD, γ-HBCD	α-HBCD, β-HBCD, γ-HBCI

+ POPs listed at COP7: Hexachlorobutadiene, PCN, Pentachlorophenol

COP8: SCCPs, deca-BDE

+ candidate chemicals considered - dicofol, PFOA, PFHxS





Principles:

- Electronic data collection
 - Data providers
 - identified by ROGs
 - Data reported as:
 - Annually aggregated
 - Primary (and later aggregated in GMP DWH)
- Standardized parametric data structure





PCB 153 was monitored at the Košetice station in 2013. 26 values were collected by means of air-active sampling. Median value was 1.457 pg/m³, mean value 1.633 pg/m³, maximum concentration reached 4.382 pg/m³.

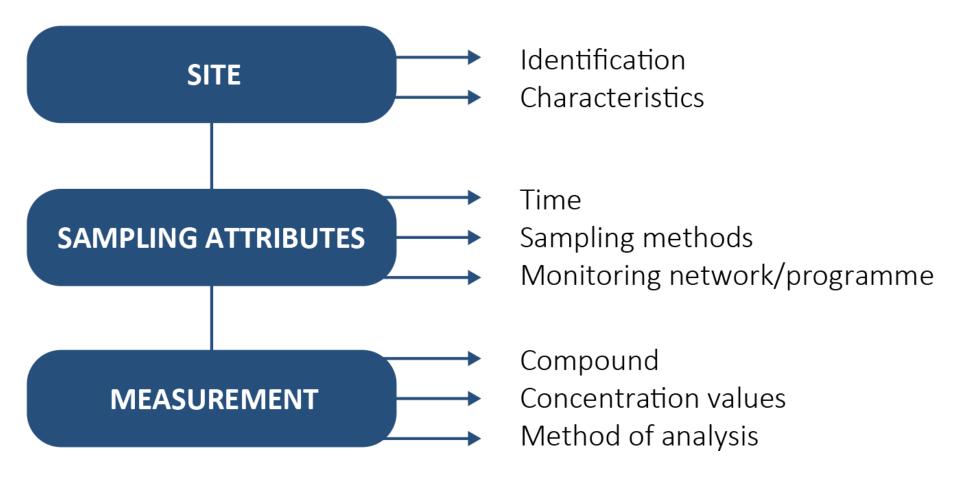
Air-active monitoring of PCB 153 at the Košetice station was performed in 2013. Median of the total 26 values was 1.457 pg/m³. Mean value was higher and reached 1.633 pg/m³.

Site	Year	Parameter	Ν	Mean	Median	Minimum	Maximum	Unit
Košetice	2013	PCB 153	26	1.633	1.457	0.506	4.382	pg/m ³





• All items grouped into 3 hierarchical sections:







Site

- Site ID (number)
- Site name (text)
- Longitude (number)
- Latitude (number)
- Region (code list)
- Country (code list)
- Site type (code list)
- Potential source type (code list)

Sampling attributes

- Year (number)
- Start of sampling (number)
- End of sampling (number)
- Type of sampling (code list)
- Type of passive sampling (code list)
- Recalculation (code list)
- Calibration description (text)
- Monitoring programme/network
 Value (median) (number)^A (text)

Measurement

- Chemical group (code list)
- Parameter (code list)
- Method (code list)
- LOQ (number)
- No. of values (number)^A
- No. under LoQ (number)^A
- Value (number)^P
- Value (mean) (number)^A
- Minimum (number)^A
- Maximum (number)^A
- 5th percentile (number)^A
- 95th percentile (number)^A
- SD (number)^A
- Laboratory (text)

A – the item is valid for aggregated data reporting only

^P – the item is valid for primary data reporting only





GMP DWH is implemented fully on-line and accessible via Internet

- Standard web browsers (Internet Explorer, Mozilla Firefox, Google Chrome)
 - recent versions highly recommended!







Web portal

- Information on GMP
- System documentation

http://www.pops-gmp.org/

Data repository

- Authorised access
- Data import, processing, validation
- Data approval

Data visualization

- Public access (since COP-7, 2015)
- Descriptive statistic analysis
- Time trends assessment

http://visualization.pops-gmp.org/2014/

GMP DWH – web portal







GMP DWH – web portal



Stockholm Convention Regional Centre for Capacity Building and the Transfer of Technology



POPs GMP data visualization and analysis



http://www.pops-gmp.org/dwh

IBA



Home

Background GMP DWH

Global Monitoring Plan Data Warehouse: online tool to store and visualize data on levels of POPs in core matrices

Contact

Introduction

The POPs Global Monitoring Plan Data Warehouse (GMP DWH) has been developed by the Stockholm Convention Regional Centre in the Czech Republic through the Research Centre for Toxic Compounds in the Environment and the Institute of Biostatistics and Analyses, Masaryk University, Brno, Czech Republic, under the guidance of the GMP Global Coordination Group, and based on Chapter 6 of the Guidance on the Global Monitoring Plan for Persistent Organic Pollutants relevant to data handling (UNEP/POPS/COP.6/INF/31).

The data reporting model involves compiling and archiving primary GMP data within a "regional data repository" in the GMP DWH for each of the five regional organization groups. In addition, the GMP DWH compiles and archives aggregated data, including supplementary data, in cases where no primary data is made available.

Attributes of the GMP DWH

The following attributes are embedded in the multi-modular GMP DWH:

- > Fully parametric data sheets harmonized data and information structure to improve the quality of information reported from particular monitoring activities, supporting their broader comparability;
- > Standardized data structure, handling and outputs the GMP DWH is designed to work with data from a wide range of heterogeneous sources, such as national monitoring programmes or large international monitoring networks, without compromising incoming information;
- Compatibility check GMP DWH contains only completed and validated data records;
- Regional data repositories- automatic tools for storage, archiving of both primary and aggregated data;
- Multilayer data validation procedure compatible data records stored in the GMP DWH are considered by members of the respective regional organization group and validated for further use in the publication;
- Data visualization- presentation of data in a uniform format;
- > Public access to the data once the validation process and preparation of regional reports is completed.

Structure and content of the GMP DWH

The GMP DWH is composed of three layers:

- > Data layer for data import, online data collection, data standards (code lists) and archiving.
- > Core layer for data management, validation, recoding, transformation, and background for



System documents

- GMP DWH Overview
- Tools in the GMP DWH
- Les User Guide for DATA PROVIDERS and USERS
- User Guide for DECISION MAKERS and USERS

Factsheets

- Factsheet 1 Overview
- Factsheet 2 Tools
- Factsheet 3 Data Import
- Factsheet 4 Data Management

Data structure

- Air
- 📙 Human milk

Human blood Water

Analytical data reporting spreadsheets

Air - primary data

X: Air - aggregated data

- Human blood primary data
- Human blood aggregated data





http://visualization.pops-gmp.org/2014/

- Data collection phase
 - Access for authorised users only (ROGs members, data providers, administrators)
 - Support for data validation and approval/rejection by ROG members
- After publication of monitoring reports
 - Public access
 - Analysis of POPs data from global view





http://visualization.pops-gmp.org/2014/



GMP Data Warehouse – Data Visualization

SPATIAL DATA SUMMARY TIME DATA DISTRIBUTION AVAILABILITY STATISTICS SERIES EXPORTS

GMP Data Warehouse - Data Visualization

Global Monitoring Plan (GMP) for Persistent Organic Pollutants (POPs) under the Stockholm Convention

The GMP Data Warehouse (GMP DWH) is an online tool developed for handling persistent organic pollutants (POPs) monitoring data generated in the frame of the Global Monitoring Plan (GMP) under the Stockholm Convention on POPs.

Available Tools

GMP Data Visualization 2014 is a tool for consideration of GMP data per region, it allows the following visualizations of available data:

- Map overview
 Data Availability
- Summary statistics
- Time Series
- Available data Parameters
- Available data Time
- Trend Map
- o Time Series Analysis
 o Time Series Bar Charts Map
- Data Exports
 Sites Summary
 Data Sources S
 - Data Sources Summary
 Analytical Methods Summary
 - Export of All Data Selected

Available tools:

- Spatial distribution
- Data availability
- Summary statistics
- Time series
- Data exports

Open data selectio

urpose of the GMP DWH is to:

as regional node for electronic data collection, storage, processing and presentation in regions with limited capacity; t the development of regional monitoring reports and the global report in the frame of the GMP;

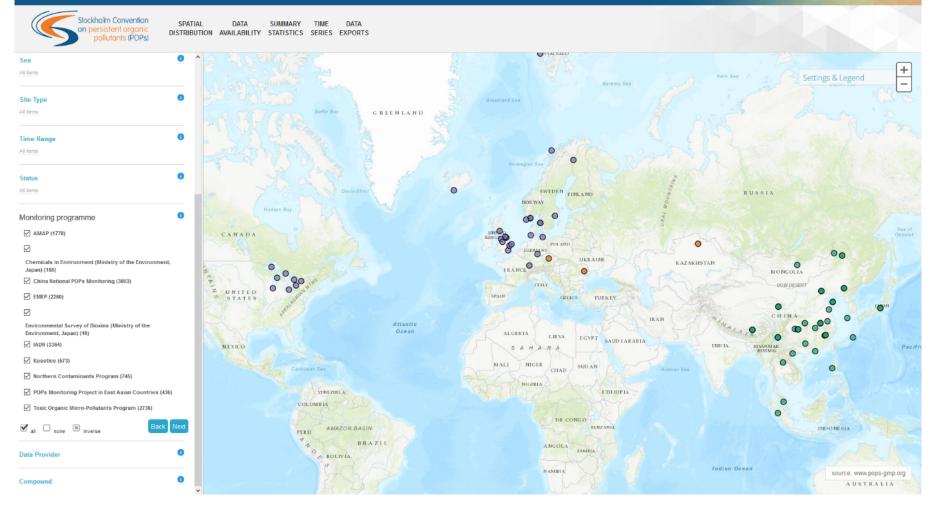
t the effectiveness evaluation of the Stockholm Convention by compiling and visualizing results of global POPs monitoring activities; ng user-friendly access to the POPs monitoring data to all stakeholders and the broad public.

s information on **POPs** concentrations in ambient air, human tissues (breast milk and maternal blood) and surface water for water-OPs (perfluorooctane sulfonic acid, its salts and perfluorooctane sulfonyl fluoride) collected in the frame of the GMP and validated by the regional organization groups of the five UN regions. These data are presented also in the regional monitoring reports.





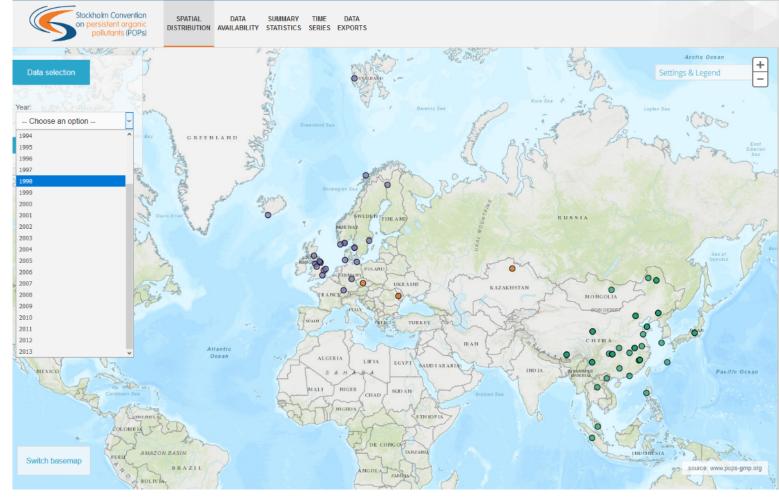
GMP Data Warehouse - Data Visualization







GMP Data Warehouse – Data Visualization



Site Detail

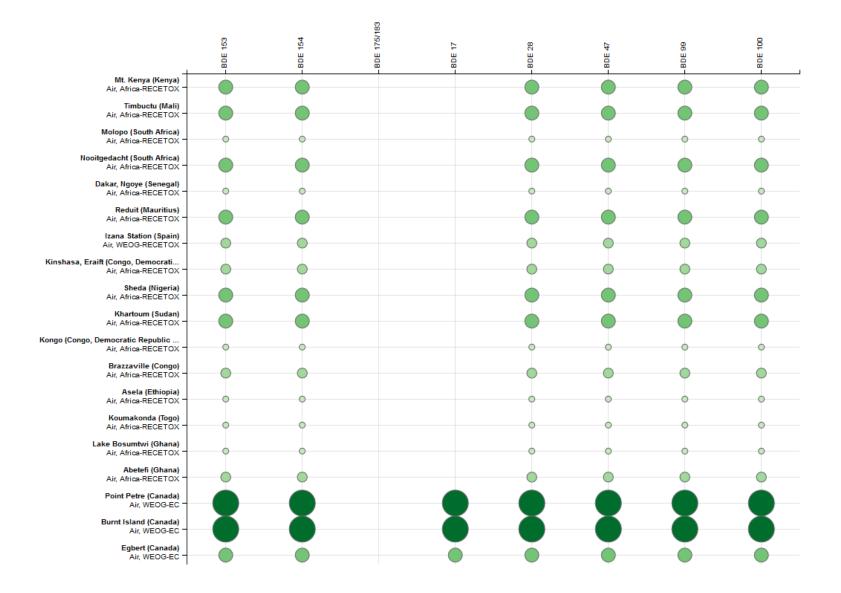
User can view additional information related to a sampling site. Click on a site (coloured point) to view: site name, latitude, longitude, UN region, country, sea, site type, and sources of pollution (Source type-air, Discharges-water). To leave the Site Detail view, click on the viewed site again.

Description

The Map Overview displays map with selected sampling sites in a given time interval. Sampling sites are marked by coloured points distinguishing the individual UN regions. Map can be scaled down or up by using the zooming tool. The map is also interlinked with data selection module and displays only data included in the selection.



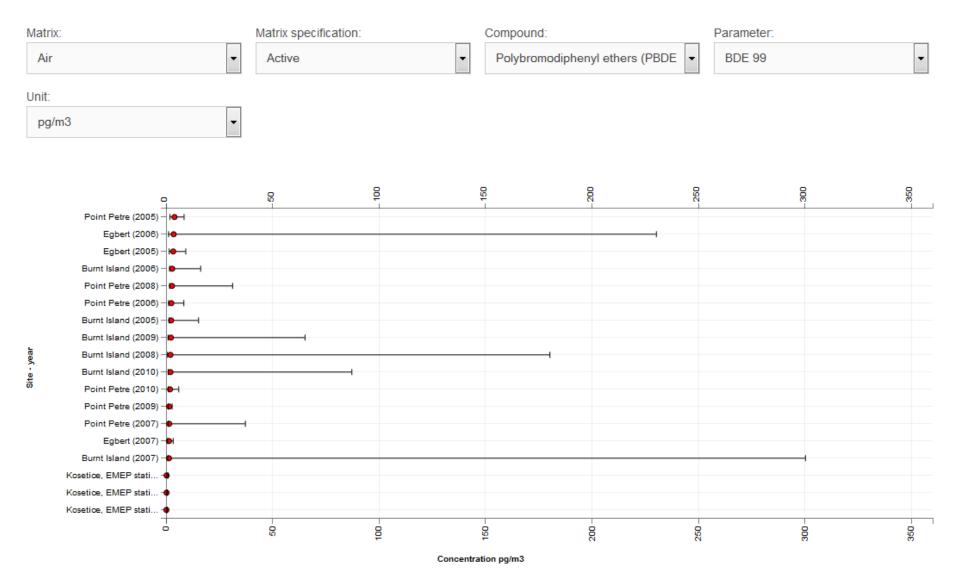




Summary statistics



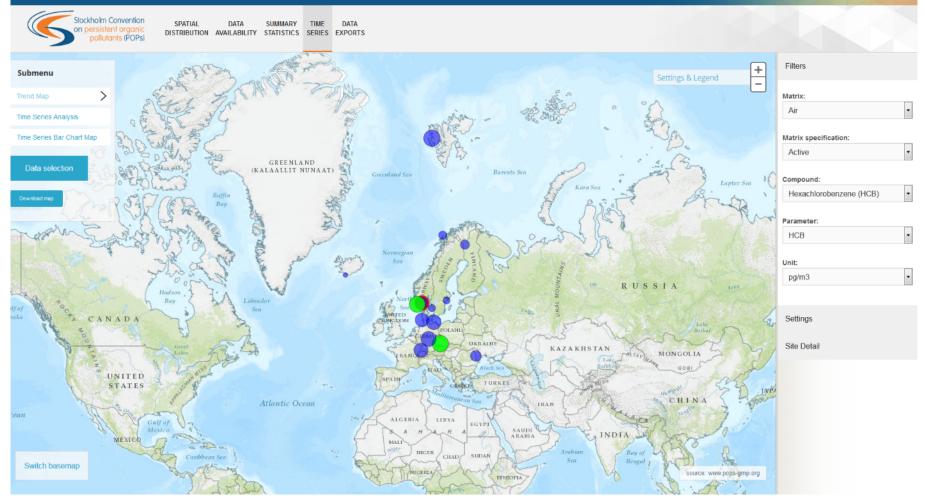








GMP Data Warehouse – Data Visualization

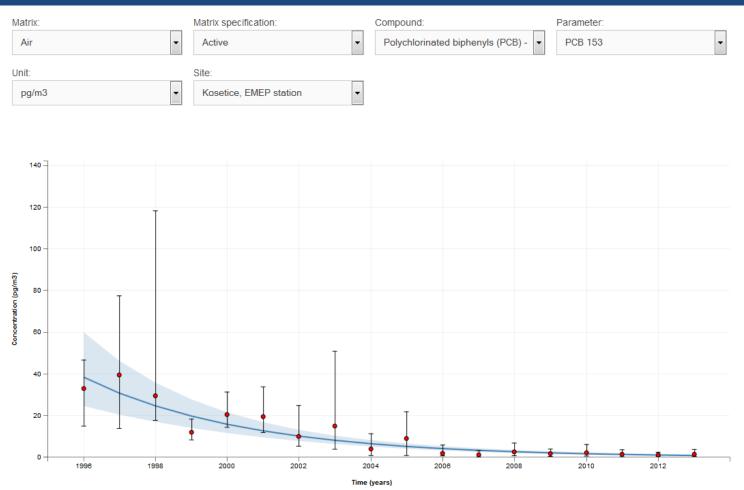


Time trends



Stockholm Convention Regional Centre for Capacity Building and the Transfer of Technology





Summary

Trend description

Mean	11.428 pg/m3	Delta	-31.543 pg/m3
Median	6.5 pg/m3	Mann-Kendal test	-0.79085 (p = 2.2949E-7)
Minimum / maximum	1.15 – 39.5 pg/m3	Daniels test	-0.9257 (p = 0)
5th percentile / 95 percentile	1.1925 – 33.975 pg/m3		

Dat	a ex	ports			(* 1) R	ockholm Convention egional Centre for Capa nd the Transfer of Techr				
Expo	rt Site	S								
Show 25	•				Search:					
entries										
Site ID	 Site Nam 	e 🔺 Latitude 🍦	Longitude 💠 UN Region 🗘 Country	Site Type	Data Provider [♦]					
GMP-A- 0000358	Leova II	46,488330N	28,283330E CEE Moldova, Republic of		CEE-NILU					
GMP-A- 0000360	Sevan, Tsovagyug	Analyti	cal Methods Sum	mary						
GMP-A- 0000361	Amberd	Show		5			Search:			
GMP-A- 0000362	Artashat	25	T							
GMP-A- 0000363	Berezinsky nature reserve	entries Matrix	Compound Analytical Method		ources S	ummar	у			
GMP-A- 0000364	Visokoie	Air	Aldrin GC-MS-NCI	-Show 25	•					Search:
GMP-A- 0000365	Lazaropole		Aldrin GC-MS/MS	entries						
		Air	Aldrin GC-MS/MS	Matrix	Matrix Specification	Monitoring Programme	First Year	🔶 Last Year 🗍	Num. of Compounds	Num. of Sites
GMP-A-	Bujkovci	Air	hexachlorocyclohexaneGC-MS (α-HCH)	Air	Active	AMAP	2008	2009	12	1
			Alpha-	Air	Active	EMEP	2009	2010	7	2
		Air	hexachlorocyclohexaneGC-MS/MS (α-HCH)	Air	Active	Kosetice	1996	2011	20	1
		Air	Alpha- hexachlorocyclohexaneGC-MS	Air	Active	Kosetice - active air sampling	2012	2013	21	1
			(α-HCH) Alpha-	Air	Active	NOT_CLASSIFIED	2012	2012	8	1
			Aluna-	Air	Passive	APOPSBAL	2004	2004	7	1
				Air	Passive	GAPS	2004	2009	12	3
				Air	Passive	MONET - CEEC	2006	2008	7	54
				Air	Passive	MONET - CZ	2003	2014	7	15
				Air	Passive	MONET - EU	2009	2013	7	21
				Air	Passive	NOT_CLASSIFIED	2011	2012	7	4

Stockholm Convention Regional Centre for Capacity Building





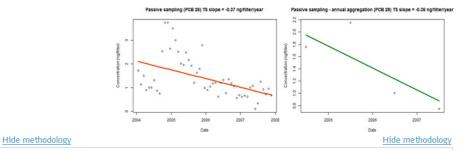
- Figures and maps exports to PNG
- Data exports to CSV (for further processing in MS Excel and other software)





Statistical and analytical tools

- R package
- Data aggregation
- Trend assessment



Annual aggregation

The process of annual aggregation represents nothing else than searching for one representative value to substitute 1-year segment of the time series. On one hand this process omits and hides certain characteristics of the time series (especially the within-year fluctuations), but on the other hand it also allows to compute some more extensive statistics not to be affected by details - especially the long-term time trends.

As in the previous examples, the selection of the best function used for aggregation of more measurements into one number depends on the distribution of the values. If we search for the central tendency, the most often used mean (intended also by the GMP methodology):

$$mean = \bar{y} = \sum_{i=1}^{n} y_i/n$$

is suitable for normal (symmetric in general) distributions, while geometric mean:

geometric mean =
$$\prod_{i=1}^{n} \sqrt[n]{y_i} = e^{\sum_{i=1}^{n} \log(y_i)/n}$$

is better in case of lognormal distribution. The nonparametric central tendency is represented by median:

$$median = \tilde{y} = y_j : P(y_i \le y_j) = P(y_i \ge y_j).$$

Nevertheless, not only the central tendency is important for characterization of the one year period of measurements. We can use other functions to get the information about spread and distribution of the values. E.g., maximum tell us about the highest concentration through the year, which could be of high importance, or the number of individual measurement describes the design of pollutant monitoring. Therefore, there is an option to input your own function into the R example.

Find more at: http://www.genasis.cz/time-series/

Current GMP DWH content

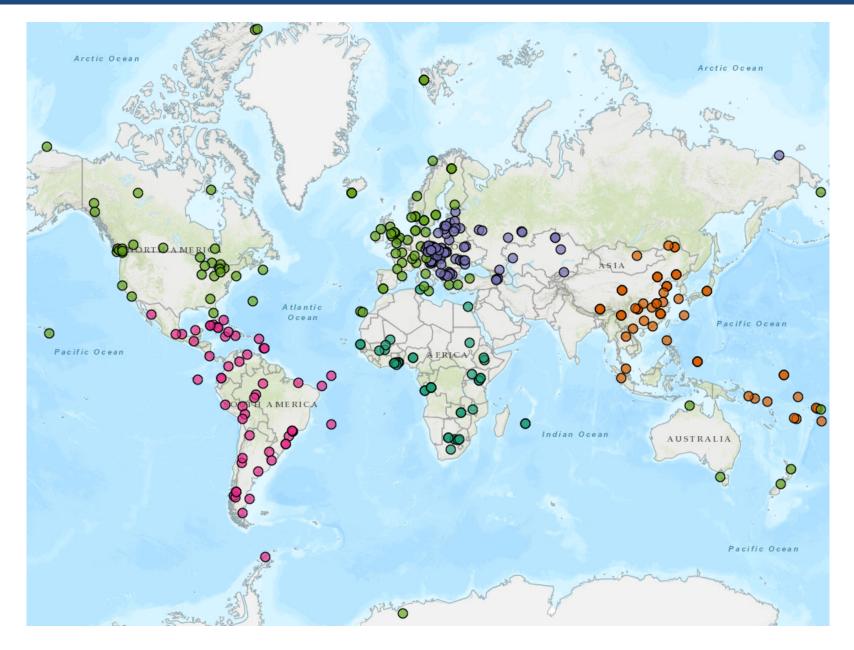
- mirrors POPs data shown in the most recent GMP regional reports (published for COP7 in 2015)
- global data in GMP core matrices
- until 2014 where possible
- REMINDER: data are **annually aggregated!**



Sites – air







Sites – air in GMP DWH



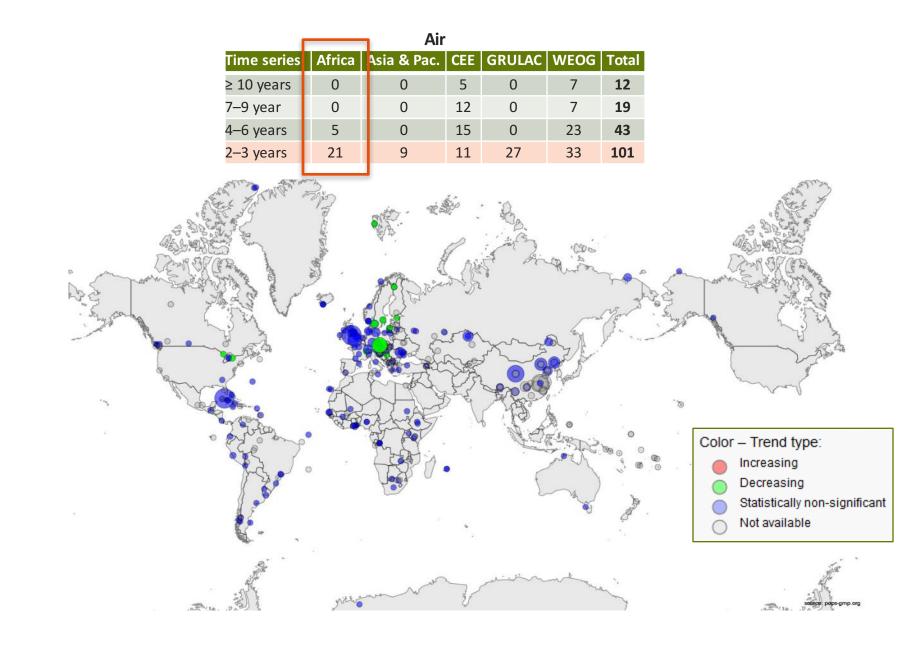
Stockholm Convention Regional Centre for Capacity Building and the Transfer of Technology



Monitoring programme	Africa	Asia & Pac.	CEE	GRULAC	WEOG	Total	Year	Africa	Asia & Pac.	CEE	GRULAC	WEOG	Total
AMAP			1		4	5	1990					1	1
EMEP			2		16	18	1991					4	4
Dioxins Survey (Japan)		1				1	1992					7	7
GAPS	7		3	14	42	66	1993					8	8
GAPS-GRULAC				11		11	1994					9	9
GMP UNEP	11	14		16		41	1995					12	12
Chemicals Environ. (Japan)		1				1	1996			1		13	14
China Nat. POPs Monitor.		37				37	1997		1	1		14	16
IADN					5	5	1998		2	1		14	17
Košetice			1			1	1999		2	1		13	16
LAPAN				13		13	2000		2	1		14	17
MONET	23		76		23	122	2001		2	1		14	17
Northern Contam. Program					1	1	2002		2	1		14	17
POPs East Asia		8				8	2003		2	1		16	19
TOMPS					8	8	2004		2	5		23	30
Total	41	61	83	54	99	338	2005	4	2	7	9	45	67
							2006	4	2	37	6	43	92
							2007	2	12	42	5	52	113
							2008	18	21	29		25	93
							2009	3	5	39	7	79	133
							2010	24	22	36	24	46	152

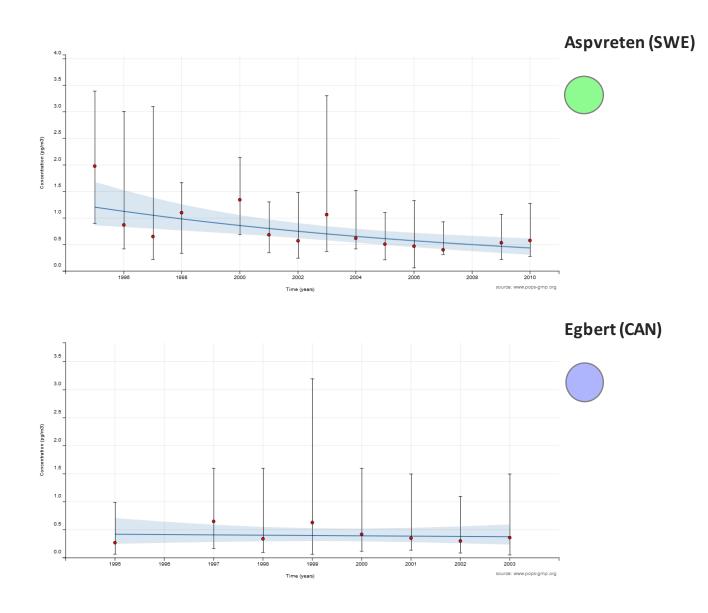








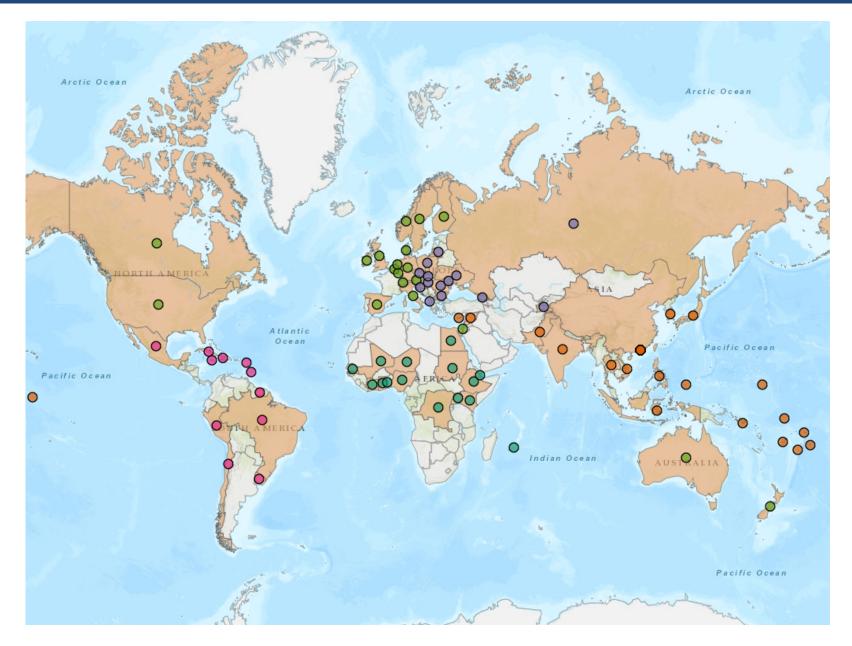




Countries – human milk







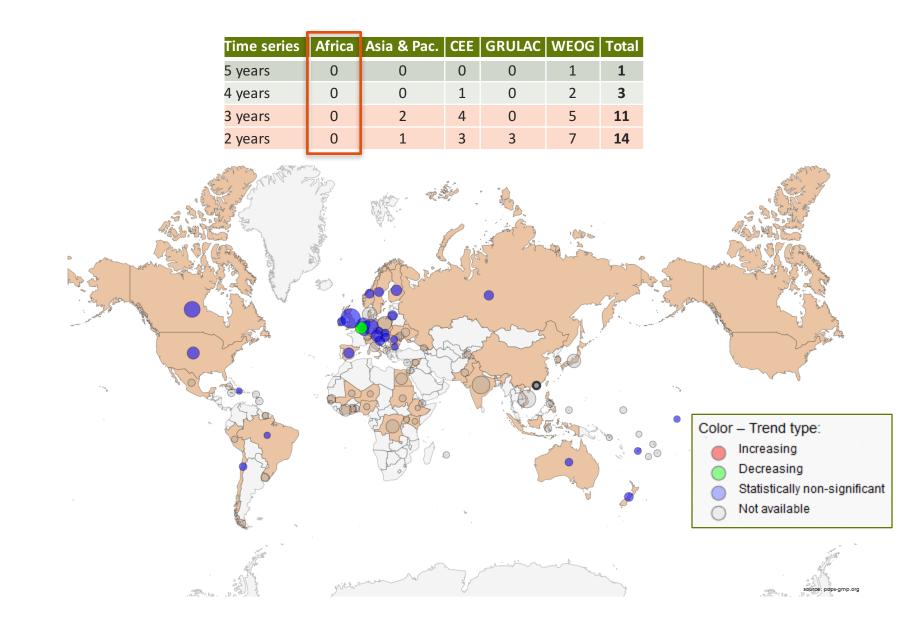




Monitoring programme	Africa	Asia & Pac.	CEE	GRULAC	WEOG	Total	Year	Africa	Asia & Pac.	CEE	GRULAC	WEOG	Total
GMP UNEP	15	20	14	11	19	79	1987		4	3		12	19
China Nat. POPs Monitor.		27				27	1992		1	8		8	17
Гotal	15	47	14	11	19	106	2000					1	1
							2001	1		8	1	7	17
							2002	1	3	1		6	11
							2003					1	1
							2004				1		1
							2006	1	3	3		3	10
							2007		23			2	25
							2008	1	2		2		5
							2009	7	3	4	1	1	16
							2010	2			1	3	6
							2011	2	36		6	1	45
							2012	1			2	1	4
							2014			2		1	3

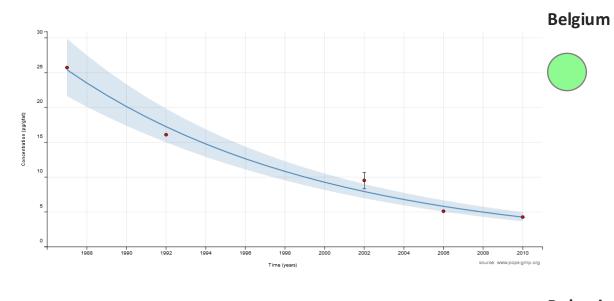


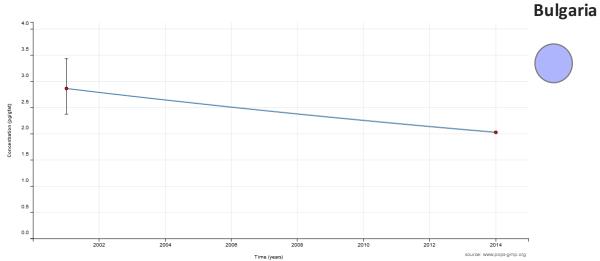








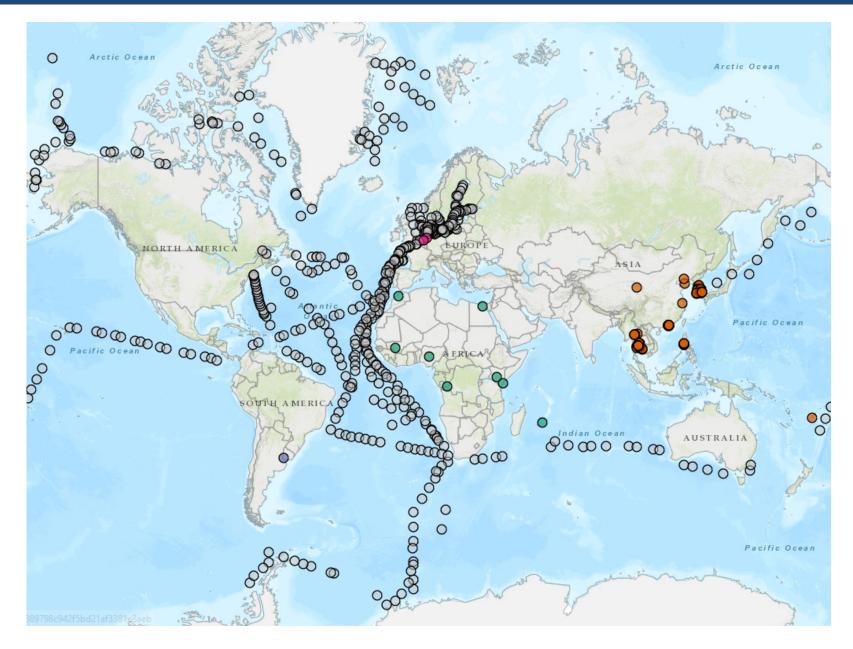




Sites – water

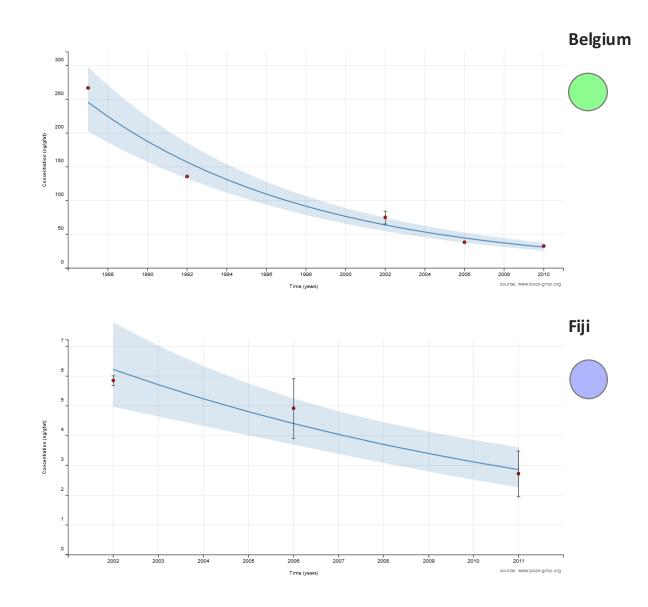












GMP DWH HELP DESK: gregor@iba.muni.cz

www.pops-gmp.org

THANK YOU FOR YOUR KIND ATTENTION



