# Executive summary

This report is prepared as a review of available Global Monitoring Plan (GMP) data on persistent organic pollutant (POPs) concentrations in three environmental matrices: ambient air, breast milk and human blood that serve as core media to monitor and examine effectiveness of the measures adopted by the Stockholm Convention on POPs.

A wide team of authors performed a critical review of the GMP reports on qualitative attributes, i.e. reported chemicals and sampling frequency of reported concentration levels. The main part of the review summarizes data available for 12 chemicals (including their recommended congeners, isomers and degradation products) that were mandatory to report in the GMP in 2008 and identifies other reported data on POPs that could be used in the future data collection campaigns.

The outcome of the review consists of three parts: review of available GMP data, methodical proposal how to improve statistical processing of UNEP-GMP data in the future and finally, the proposal of a new e-data capture system suitable for the next GMP campaigns. Each of these chapters respected principal characteristics of examined environmental matrices. Key conclusions are highlighted in this executive summary.

**Review of available GMP data**

Compounds and parameters found in the GMP reports were sorted by their relation to the Stockholm Convention, by reported data content and its appropriateness for further trend analyses. The classification strictly followed scope of the Stockholm Convention and the GMP Guidance document and their amendments in time. Reported compounds were therefore classified into four groups: 1) original 12 POPs included in the Stockholm Convention in 2001, their congeners, isomers and degradation products determined in the GMP Guidance (2007); 2) additional 10 POPs included into the SC in 2009 and 2011 and specified in the updated GMP Guidance; 3) all other compounds, their sums and toxic equivalents related to the Stockholm Convention but not specified in any of the GMP Guidance documents; 4) compounds found in the GMP reports but not related to the SC.

All regional GMP reports contain 171 variables (including concentration data on congeners, isomers, transformation products, various summations and toxic equivalents – TEQs). Analysing the primary pool of reported parameters, 149 of them were related to 12 original Stockholm Convention POPs and 10 additional POPs. Out of this number, 58 +7 were listed as compounds highly recommended for monitoring and evaluation in the GMP Guidelines documents for 12 original and 10 additional POPs, respectively. Remaining parameters (84) consisted of non-recommended congeners, various sums and toxic equivalents, often not correctly identified. There were also 22 other reported parameters with no relation to the SC.

The qualitative part of the review concluded that most of the original 12 POPs were reported in all matrices and existing data would thus enable a time-related comparison with data from future sampling campaigns.

Although the GMP Guidelines from 2007 would not cover the current Stockholm Convention scope as amended, some of the additional 10 POPs were also identified in the reviewed reports (8 in ambient air and 4 in both breast milk and human blood, respectively).

In addition, reported sampling frequency was evaluated to determine content of the GMP data usable in time trend analyses. Chemicals with concentration reported over time span of at least 4 consecutive years were marked as suitable for baseline time series analyses („historical time series available"), while other records belong to the family of „point estimates" or „occasional reports". Moreover, historical time series may be very useful in investigation of background time changes and trends in POPs concentrations before adoption of the Stockholm Convention.

Availability of time series differed in the examined matrices, both within the group of original 12 POPs and additional 10 POPs. Altogether, a sufficient number of 12+10 POPs were reported for ambient air monitoring data in the category "historical time series available" for at least one country from three UN regions with "point estimates" being available in all UN regions. Records spread over 1990 – 2008, however majority of the time series originate from recent measurements (after 2004). Ambient air monitoring data thus represent a relevant basis for pair-wise comparison with the future GMP data collection campaigns.

POPs detected in human tissues were mostly reported as lipid-adjusted concentrations and mostly as annual point estimates. Available data are suitable for future statistical processing. The spectrum of chemicals and their spatial and temporal coverage reported for breast milk and human blood was a significantly lower than that for ambient air. Nevertheless, all of the original 12 POPs in the Stockholm Convention were reported.

Although reported data in analyzed reports were, in general, rated suitable for planned statistical processing, the audit of the content revealed serious challenges related to data standardization in the GMP report. The reports suffer from the lack of standardized taxonomy for POPs, their isomers, transformation products and summations which were frequently used. Heterogeneity of the data is further enhanced by reporting various toxic equivalents (TEQ) (based on WHO TEF values from various years) rather than concentrations of the individual PCDDs, PCDFs and PCBs congeners. Unclear identification of units, time and spatial scales of the reported concentrations as well as insufficient specification of aggregated data belong to other frequently identified drawbacks.

Another source of variation is associated with different structure of reports provided by different UN regions; some of the national records included detailed primary data including even rarely measured compounds while others were based only on sums of key groups of POPs. Low standardization of the regional GMP reports also contributed to a remarkable variation in reported concentration of examined chemicals.

Nevertheless, a significant part of available data allowed development and testing of a pilot database. We conclude that data processing can continue (after expert review and approval) in further analyses including assessment of the contamination baselines.

Data from all GMP reports were re-coded into an electronic database template, standardized in structure and in content of data fields and validated with respect to their spatial and temporal variability. This analysis definitely confirmed a reliable coverage for all original 12 POPs included in the Stockholm Convention, in both ambient air and human tissues monitoring.

The review and pilot processing of data resulted in conclusion that available GMP data can be used for baseline statistical processing. Data can be analyzed as annually aggregated time series or at least as relevant point estimates, prepared for comparison with the next data collection campaign. However, detailed standardization and relevant quality scoring of individual records are needed prior any future processing. The methodical solution of this problem is proposed in a section dealing with the structure of the future GMP database in chapter 6. Although available data were inserted into e-data capture forms, some uncertainties could not be solved at this stage (i.e. missing values). The information relevant to that particular parameter is stored in the database but such record could not be used in the statistical processing at the moment. Fully standardized re-coding of available data prepared for the direct linking with the next GMP data collection campaigns in the same database appears to be effective way how to solve all limits in data standards.

**Methodical proposal how to improve statistical processing of UNEP-GMP data**

The content of the GMP reports was analyzed and summary overview of the data structure and content suitable to run statistical summaries were prepared (see Annex 6.1). Based on this analysis, several methodical concepts were proposed to improve future GMP data collection campaigns, associated data management and processing:

- proposal on how to solve main problems with statistical processing of unclear inputs, i.e. reconstruction of the sample distribution variability in a heterogeneous field of used variability measures, estimation of minimum effect size for future comparison of GMP campaigns and a robust methodology of trend quantification and assessment;
- proposal on how to solve data heterogeneity and how to implement data collection standards.

The main added value of this part is the successful pilot quantification of effect size, relevant for statistical comparison of POPs concentrations. Using the model data from the Košetice station (Czech Republic), statistically significant quantified difference could be determined for POPs concentration data, provided that they were detected in the same matrix and by same methodology. This approach might be helpful in comparison of data between the former and future GMP records. It will be further verified using a larger international data set

Based on the test data, it has been shown that although annually aggregated concentration data sets definitely loose the seasonal variability, they are still suitable for comparative statistical processing. The overall trends can be assessed on annually-adjusted data provided that the statistically significant and detectable difference (effect size) is known/determined. Therefore, annually aggregated data were recommended as a minimum for the relevant statistical comparison of former and future GMP reports.

Finally, a robust statistical methodology was applied for pair-wise comparison of consecutive time measurements or for trend estimation in short time series and tools allowing for processing of heterogeneous data were built in order to keep maximum information for future comparison.

**Proposal for a new electronic data capture system, suitable for the next GMP reports**

The required IT support necessarily includes centralized data management, database customized for POPs concentration data and its standardized superstructure allowing comparison of former and future GMP data collection campaigns, and thereby quantification of time trends. These tasks represent a challenge for several disciplines including environmental informatics, statistics and software development.

A comprehensive ICT background for planned processing of the UNEP-GMP data was proposed and preliminary designed to fulfil required functionality: standardization of already obtained GMP data, database suitable for archiving and processing of POPs data and on-line data visualization system. The development followed criteria and standards adopted for the information system GENASIS (www.genasis.cz), which was also used for the pilot visualization of available GMP data (www.genasis.cz/unep). The pilot version of the on-line data visualization was prepared and tested on the re-coded GMP data. Prepared graphical and classificatory functions are password protected and allow authorized users to sort data, study their structure and completeness. However, we recognize that further validation procedures are strongly needed. That is why the on-line visualization is opened only to authorized data administrators at the moment. Finally validated POPs data should be further re-coded to a robust GMP database that would potentially serve as a centralized storage and visualisation system for the next GMP campaign.

In addition to its functionality, security management of the proposed solution is highly emphasized. Only password protected data administrators are approved to use the pilot on-line data visualization tools. The system was developed using common and advanced standards adopted for information systems of high quality, e.g. ISO 9001:2001/2009, ISO/IEC 20000-1:2006, ISO/IEC 27001:2006.

Proposal for the standardized GMP database based on fully parametric data sheets that could be used in the next GMP data collection campaign is one of the outputs of our work. The whole system is proposed to improve quality of collected global data set on POPs occurrence

in the environment and to strengthen the ownership and responsibility of the local data administrators.

In the pilot database proposal, the set of parameters was arranged hierarchically, corresponding to the ontology of the planned information system. Three main levels are prioritized in the ontology: a) identification of the sampling site, b) spatial data description, c) measurement – value items. Besides primary data (measurement level), the data collection system will be equipped with classifiers facilitating data validation and processing: coded information on spatial aggregation (if used/applicable), coded information on time aggregation (if used/applicable). Only annually aggregated values will be accepted and additional classifier linking new GMP reports with formerly reported variables is proposed. We are of the view that next GMP data collection campaign could contribute to the retrospect identification of missing records such as incomplete time span of reported data, location, missing records and to the verification of uncertain records and thereby increase the number of validated and standardized parameters collected in the first set of GMP reports.