

# Partnership for the Assessment of Risks from Chemicals

Additional Deliverable AD8.3

Conceptualization of a computational system for the EWS

WP8 – T8.2



D8.1: Report on the conceptual design of the SSbD toolbox	Dissemination level: PU
WP8: Concepts and toolboxes	Version: 1.0
Main Authors: Denis Sarigiannis	Page: 2

## Technical reference

Work package	<b>WP8 - Concepts and toolboxes</b>
Task	<b>T 8.2 - Scientific and technical basis for an Early warning system (EWS) on chemical risks</b>
Dissemination level <sup>1</sup>	PU
Lead Beneficiary/ Responsible AE	<b>AUTH</b>
Contributing Participants	<b>UMU (SE), NKUA (EL), AUTH (EL), INSERM (FR), IRFM (IT), UG-PL (PL), IISPV (ES), AU (DK)</b>
Responsible author(s)	<b>Ioana Chelcea (UMU) and Patrik L. Andersson (UMU)</b>
Co-authors	<b>Karine Audouze/Inserm/karine.audouze@u-paris.fr Olivier Taboureau/Inserm/olivier.taboureau@u-paris.fr Nikiforos Alygizakis/NKUA/nalygizakis@chem.uoa.gr Achilleas Karakoltzidis/AUTH/karakoltzidis.achilleas@gmail.com Spyros Karakitsios/AUTH/spyrosk@auth.gr S Kumar/IISPV/saurav.kumar@iispv.cat D. Deepika/IISPV/deepika@iispv.cat V. Kumar/IISPV/vikas.kumar@urv.cat Ewelina Wyrzykowska/UG-PL/e.wyrzykowska@qsarlab.com Tomasz Puzyn/UG-PL/t.puzyn@qsarlab.com Katrin Vorkamp/AU/kvo@envs.au.dk Farina Tariq/UMU/farina.tariq@umu.se</b>
Reviewers external to WP	<b>Magnus Lofstedt, European Environment Agency, magnus.lofstedt@eea.europa.eu Uko Maran, Tartu University, uko@theor.chem.ut.ee</b>
Due date of deliverable	31/10/2023
Actual submission date	25/01/2024

D8.1: Report on the conceptual design of the SSbD toolbox	Dissemination level: PU
WP8: Concepts and toolboxes	Version: 1.0
Main Authors: Denis Sarigiannis	Page: 3

## Document history

Version	Date	Reviewer name/Institutions	Short description of changes
1	19/09/2023	Ioana Chelcea (UMU) and Patrik L. Andersson (UMU)	Draft version of AD8.3 for review
2	13/10/2023	Magnus Lofstedt (EEA) Uko Maran (TU) Pascal Sanders (ANSES)	Draft version with comments from the reviewers and PARC Coordinator
3	22/01/2024	Denis Sarigiannis (AUPh) Spyros Karakitsios (AUPh)	Final version taking into account the comments from the reviewers and PARC CT

“Funded by the European Union. Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or the Health and Digital Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.”

D8.1: Report on the conceptual design of the SSbD toolbox	Dissemination level: PU
WP8: Concepts and toolboxes	Version: 1.0
Main Authors: Denis Sarigiannis	Page: 4

## Abstract

Computational tools will be critical in an effective early warning system (EWS) to enable identification of new emerging risk chemicals (NERCs). In several areas of chemical risk assessment, tools are well developed and mature to be applied in an EWS whereas in emerging fields including non-target screening, omics, and natural language processing of social media, tools are under rapid development. A majority of chemicals in commerce lack data to complete hazard screening for early warning and in silico-based methods can serve to fill data gaps and increase throughput of an EWS. Current report reviews existing computational tools with a focus on those that are available and maintained, and that could be implemented and applied in a first version of an in silico driven EWS. The first step in a systematic search for NERCs is automatic data mining and curation of any source and inventory aiming at delivering molecular structure information for further processing and if available experimental data. Next step covers parallel assessment of exposure and effects that will feed information for a weighing of an overall score of hazard and ultimately identification of potential NERCs. Several challenges are identified and discussed including integration and scoring of hazard data of various kinds from fate and distribution of chemicals towards subtle effects in certain species and tissues. A large number of computational tools exist, and it is already today possible to apply these as a starting point in an integrated EWS workflow for automatic identification of NERCs.

## Key Words

Computational, Modelling, EWS, NERCs, in silico, QSAR, PBK, retrospective suspect screening, bioinformatics, systems biology

D8.1: Report on the conceptual design of the SSbD toolbox	Dissemination level: PU
WP8: Concepts and toolboxes	Version: 1.0
Main Authors: Denis Sarigiannis	Page: 5

## Table of Contents

<b>1. Introduction</b>	<b>6</b>
<b>1.1 Computational tools for EWS</b>	<b>6</b>
<b>1.2 Workflow for the integration of computational tools into an EWS</b>	<b>7</b>
<b>1.3 Aims</b>	<b>9</b>
<b>2. Data collection and curation</b>	<b>10</b>
<b>2.1 Data sources</b>	<b>10</b>
<b>2.2 Data extraction</b>	<b>11</b>
<b>2.3 Use of non-target analytical data, digital archiving and retrospective suspect screening</b>	<b>12</b>
<b>2.4 Data curation</b>	<b>12</b>
<b>3. Exposure-based warning systems</b>	<b>13</b>
<b>3.1 External exposure models</b>	<b>14</b>
3.1.1 Human exposure models	14
3.1.2 Environmental fate models	15
<b>3.2 Internal exposure models</b>	<b>16</b>
3.2.1 Bioconcentration and bioaccumulation models	16
3.2.2 Compartmental models	17
3.2.3 Physiologically-based kinetic models	19
<b>3.3 Transformation models</b>	<b>21</b>
<b>4. Effect-based warning systems</b>	<b>23</b>
<b>4.1 Adverse outcome pathways (AOPs)</b>	<b>23</b>
<b>4.2 QSAR models</b>	<b>24</b>
<b>4.3 Bioinformatic tools</b>	<b>24</b>
<b>4.4 Other computational tools</b>	<b>26</b>
<b>5. Integration of data for identification of potential NERCs</b>	<b>28</b>
<b>5.1 Scoring of chemicals</b>	<b>30</b>
<b>5.2 Grouping and clustering</b>	<b>30</b>
<b>5.3 Computational platforms for data integration</b>	<b>31</b>
<b>6. Conclusions and Future Aspects</b>	<b>32</b>
<b>7. References</b>	<b>33</b>

D8.1: Report on the conceptual design of the SSbD toolbox	Dissemination level: PU
WP8: Concepts and toolboxes	Version: 1.0
Main Authors: Denis Sarigiannis	Page: 6

## 1. Introduction

With an increasing number of chemicals being produced and released into the environment, the need for early warning systems (EWS) has become apparent, to flag chemicals of potential concern at an early stage. Previous deliverables of WP8 in PARC have highlighted the need for an EWS and identified its potential components (D8.1 and D8.2). The deliverable D8.2 included information on EWS developed by various national agencies. This includes projects such as those presented by the Swedish toxicological council for identifying new emerging risk chemicals (NERCs) in products (Bruks et al., 2021, 2022), by the NORMAN network (NORMAN, 2016) for identifying hazardous chemicals in the environment or for identifying NERCs in human serum (Weiss et al., 2022). The European Commission recently reviewed this field and highlighted the need for EWS to detect NERCs (European Commission, 2021). EWS have also been discussed in the field of occupational exposure aiming at a healthy and safe working environment e.g., clinical watch systems, or systems relying on epidemiological databases and biomarkers (Palmen, 2016) EWS for food safety have been developed that apply computational tools such as machine learning (Geng et al., 2017; N. Liu et al., 2022). Although existing EWS are very different in their approaches and aims, several utilize computational tools in the process for either data storage, collection, analysis, or property predictions. This highlights the key role of computational tools in developing a robust EWS for detection of NERCs.

### 1.1 Computational tools for EWS

Computational tools have become an important cornerstone in toxicology research with new approach methodologies (NAMs) and next generation risk assessment (NGRA) focusing on their continued development. *In silico* approaches have proven to be useful for a multitude of purposes, which makes them crucial for the development of early warning systems. Supervised machine learning approaches have been used for information gathering including cases of text mining, predictions of environmental fate and hazard properties using quantitative structure-property relationship (QSPR) or quantitative structure-activity relationship (QSAR) models and predictions of transformation products. Furthermore, they have supported the interpretation of complex multivariate biological effect data in the field of bioinformatics and systems biology. Additionally, unsupervised methods such as clustering have been employed to infer chemical and toxicological similarities between emerging chemicals and those with known properties and risks. Furthermore, mathematical models have been developed to simulate internal exposure and dose at target using physiologically based kinetic (PBK) models. It is important to note that many of these computational tools are developed for very specific purposes and these may thus have narrow applicability domains. In a first phase of the development of a computational EWS, models with large applicability domains will be in focus that would cover a variety of potential NERCs. Therefore, this report

D8.1: Report on the conceptual design of the SSbD toolbox	Dissemination level: PU
WP8: Concepts and toolboxes	Version: 1.0
Main Authors: Denis Sarigiannis	Page: 7

mainly describes more generic models as well as freely available tools that can be easily incorporated into an automated computational workflow. However, it will be critical over time to include novel and specific tools and thus also emerging technologies and related activities ongoing in PARC are being briefly presented, that have the potential to be incorporated in EWS. Combinations of global and local models would also enable refinement in predictions and identification of NERCs with confidence. Another emerging field is natural language processing which can be applied to identify early signals of chemical hazards discussed in social media and related platforms.

## 1.2 Workflow for the integration of computational tools into an EWS

The computational EWS is conceptualized in Figure 1 including data collection and curation signals towards scoring exposure and effect potential to an integration of signals and notification of potential NERC. The EWS starts with 1) an *ad hoc* expert based reporting of findings from omics, non-target screening or alike, or 2) from a wide scope horizon scanning undertaken on regular basis of e.g., grey literature including social media, patent information, environmental occurrence and human exposure data bases covering non-target screening, and effect directed analysis data aiming at identifying novel entities (Figure 1, I). Critical in scanning data systematic and automatic will be using natural language processing methods and thorough curation approaches aiming at reaching chemical structure information for use in next steps. This step will also gather existing experimental data, if available, on all hazard aspects. If experimental data is available for hazard scoring, this entry could go directly to data integration and scoring NERC. However, as data for NERCs are sparse, computational predictive tools are needed to estimate various fate and effect properties and thus a curated chemical structure is the essential output of step I. The next steps in the EWS workflow are to predict data on exposure potential (Figure 1, II) and effect potential (Figure 1, III). Exposure can be assessed *in silico* by predicting the potential environmental fate of a compound, including, for example, its accumulation in biota or specific tissues. It also aims at predicting dose at target, i.e., the internal exposure. Predicting exposure potential will overall require scenario settings on emissions related to use and production information, and potential transformation reactions and their kinetics. Predictive models for assessing effects can be used within an adverse outcome framework. Models will be required covering various types of effects, species, biological complexity and including for example omics data. Overall, several computational tools exist that can be incorporated into the EWS for assessing exposure and effect potential.

D8.1: Report on the conceptual design of the SSbD toolbox	Dissemination level: PU
WP8: Concepts and toolboxes	Version: 1.0
Main Authors: Denis Sarigiannis	Page: 8

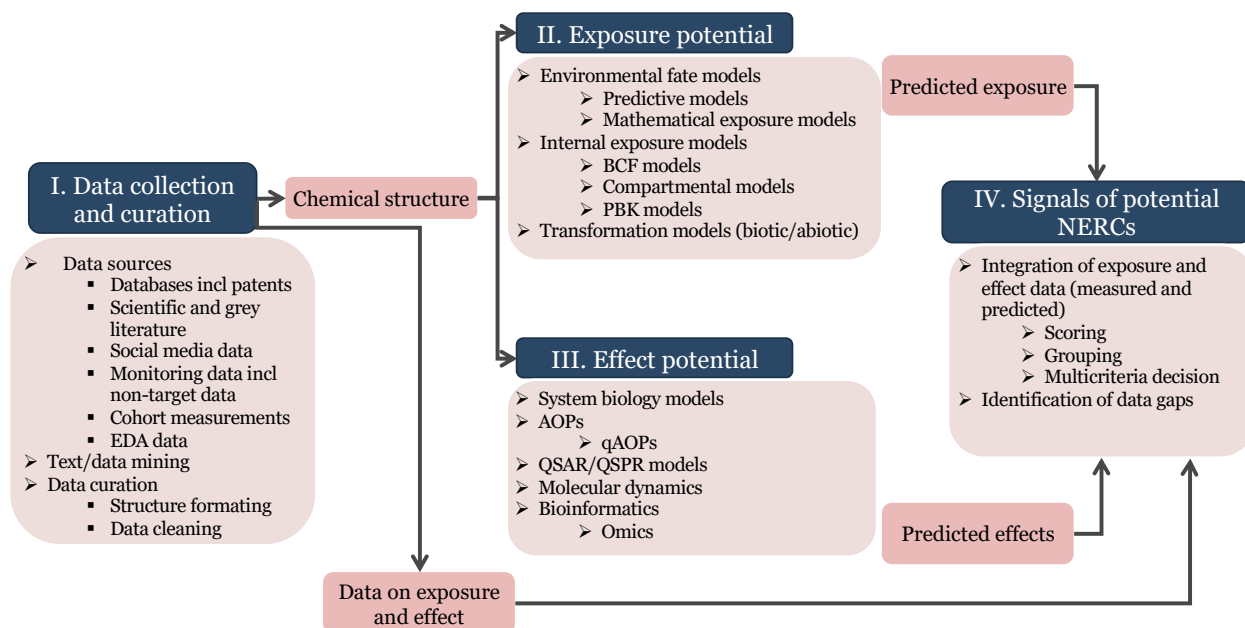


Figure 1. Conceptual workflow for a model assisted early warning system from I) Data collection and curation using text mining and automatic data curation, II) external and internal exposure modelling, III) effect modelling from system biology to bioinformatics, to IV) integration of signals for identification of potential NERC.

Signals from both experimental data and computational predictions can be combined in a matrix of exposure and effect indicators. Here, each compound can be scored in terms of their potential hazard properties or other criteria (Figure 1, IV). Complementary grouping methods can be used applying known hazardous chemicals as anchors and positive controls for identification of NERCs. An example of a scoring matrix is given in Figure 2. A scoring system could be based on specific cut-off values in existing regulations, for example a bioconcentration factor of 2000, which defines bioaccumulation in REACH. Such a scoring matrix would allow for a variety of data sources to be integrated. It would also allow to merge multiple models targeting the same endpoint in a consensus approach. Additionally, weighting factors can be introduced to account for data uncertainty, i.e., experimental data could be assigned a higher weight than predicted data. Lastly, dependency weights can be assigned to account for the added risk of two concerning properties being additive. For example, a compound with predicted high liver accumulation is more concerning if it is also predicted as hepatotoxic and thus this dependency needs to be programmed into the final score. The EWS workflow proposed in this report is suggested as a screening tool especially in the case of data scarcity and could be complementary to other EWS that have been suggested (e.g. European Commission, 2021).

D8.1: Report on the conceptual design of the SSbD toolbox	Dissemination level: PU
WP8: Concepts and toolboxes	Version: 1.0
Main Authors: Denis Sarigiannis	Page: 9

		Tool							
		Exposure models			Effect models				
		BCF model (VEGA)	BCF model (EpiSuite)	Persistence (VEGA)	Liver accumulation (httk)	EDC (VEGA)	ER (VEGA)	Hepatotox (VEGA)	....
Compounds	Compound 1	Green	Green	Green	Green	Red	Red	Orange	
	Compound 2	Red	Red	Red	Green	Green	Green	Green	
	Compound 3	Green	Green	Green	Orange	Red	Red	Green	
	Compound 4	Red	Red	Red	Green	Green	Green	Orange	
	Compound 5	Orange	Orange	Green	Red	Green	Green	Red	
⋮									

Figure 2. Example of a scoring matrix to identify NERCs within the EWS where red indicates that a set threshold has been reached or the value indicates an alarming property, orange indicates that a hazardous property is likely, and green no indication of hazard. In such an approach, fields can be left empty to indicate data gap or compound outside applicability domain. The matrix is exemplified with a few predicted hazard properties with indicated potential model platforms.

### 1.3 Aims

This report builds on the PARC deliverable D8.2 entitled “Report on the conceptual design and development guidelines of the PARC EWS”. D8.2 highlights the need to develop EWS that move away from expert judgements towards an automated process, which can scan literature for existing data and make use of computational tools, to fill data gaps but also to analyze and incorporate data from effect-direct analysis and non-target screening. Here we aim to describe the concept of a computer driven EWS for identification of NERCs in various matrices and data sources including databases, grey and scientific literature, and social media. The specific aims are to (i) present the role of computational tools in the conceptual structure of an EWS including strengths and weaknesses, (ii) describe computational tools that are generic, practical to implement and suitable for an automated EWS, (iii) discuss how an automated EWS, handling multiple data sources and types, will identify NERCs, and (iv) present emerging techniques that could be implemented in an EWS as they mature including local and specific models. Note that this report is not focused on user perspectives, long term governance and maintenance of the EWS.

D8.1: Report on the conceptual design of the SSbD toolbox	Dissemination level: PU
WP8: Concepts and toolboxes	Version: 1.0
Main Authors: Denis Sarigiannis	Page: 10

## 2. Data collection and curation

The initiation of the process of identification of NERCs could be as described above a broad scanning on a regular basis or data provided by any party indicating a potential risk of hazard. The first step is data collection and with data we refer to any information about the compound such as molecular structure, physical-chemical properties, measured biological effects, production volume, use patterns in products or detection in the environment or humans.

### 2.1 Data sources

There are over 900 different databases which can be used to extract chemical information, build computational tools, or parameterize models (Pawar et al., 2019). Pawar et al., (2019) categorized these into 13 different classes according to the information each database provides. Databases of relevance include e.g., those including information on physicochemical properties, toxicological information, omics data, product and material use and characteristics, patents, environmental and human monitoring data, and data on adsorption, distribution, metabolism, and excretion (ADME). In addition, social media platforms including X (twitter), Facebook, and alike could provide valuable information where potential new chemical threats are being discussed on new use of chemicals, new products, effects, and occurrence in various human and environmental samples. Data from social media has great potential as it could provide real-time information with geospatial and demographic resolution. In addition, social media could contribute to identifying trends or emerging risks, by providing information not only on actual use, but also of misuse of, and exposure to, chemicals. However, such data requires elaborate and careful treatment using natural language processing to handle issues of bias, reliability, etc.

Large chemical databases or chemical registry databases such as PubChem ([pubchem.ncbi.nlm.nih.gov](http://pubchem.ncbi.nlm.nih.gov)), ChEMBL ([ebi.ac.uk/chembl](http://ebi.ac.uk/chembl)), ChemSpider ([chemspider.com](http://chemspider.com)) or CAS SciFinder ([cas.org](http://cas.org)) are the most suitable databases to screen for new compounds. Other chemical databases are summarized by Pawar et al., (2019) and Wang et al., (2020). The EPA CompTox Dashboard is one of the databases providing toxicity data ([comptox.epa.gov/dashboard](http://comptox.epa.gov/dashboard)) and includes over 1.2 million entries with structure information and experimental data on properties and toxic effects. Environmental toxicity data and reference values for a variety of species are included in databases such as the ECOTOX database ([cfpub.epa.gov/ecotox](http://cfpub.epa.gov/ecotox)) or IRIS ([epa.gov/iris](http://epa.gov/iris)) of the U.S EPA.

Many ADME databases such as Pharmapendium ([pharmapendium.com](http://pharmapendium.com)) are commercial and focus on pharmaceuticals, which limits their applicability to identifications of a broad spectrum of NERCs. However, this information may still be useful for training or parameterizing models or identifying emerging compounds that may have ADME properties similar to a compound in the database. Omics databases, such as ArrayExpress ([ebi.ac.uk/biostudies/arrayexpress](http://ebi.ac.uk/biostudies/arrayexpress)) or BiGG ([bigg.ucsd.edu](http://bigg.ucsd.edu)) contain data on metabolomics,

D8.1: Report on the conceptual design of the SSbD toolbox	Dissemination level: PU
WP8: Concepts and toolboxes	Version: 1.0
Main Authors: Denis Sarigiannis	Page: 11

genomics or proteomics for a variety of compounds which may be of use when investigating and comparing chemical responses. How this type of data can be used in an EWS will be discussed later in this report.

Patent information is relevant for identifying chemicals that may be of concern even before their commercialization and to predict potential sources and exposure routes. Some national registry databases contain patent information (Wang et al., 2020). However, it may be more efficient to make use of larger databases such as those of the Derwent World Patents Index (DWPI) ([clarivate.com](https://clarivate.com)). Additionally, the EPA CompTox Dashboard also provides information categorizing usage or function of chemicals in consumer products. Patent or usage data may have to be processed and divided into larger classes for exposure predictions. Such classes could then be used to calculate an exposure index (Swedish Chemicals Agency, 2020, 2022; Weiss et al., 2022).

Human biomonitoring data is another type of information that may be useful for identifying chemicals of risk, although it often focuses on well-studied compounds rather than NERCs. In a context of identification of NERCs, suspect and non-target screening data will be more useful as digital archiving systems are being developed. However, biomonitoring data may indicate unexpected exposure patterns or trends and might also allow extrapolations to lesser-known emerging chemicals. The EU HBM Dashboard ([hbm4eu.eu](https://hbm4eu.eu)) provides summarized statistics of biomonitoring data of eight substances, which is a limited data amount for uses in EWS. The IPChem Portal ([ipchem.jrc.ec.europa.eu](https://ipchem.jrc.ec.europa.eu)) offers the possibility to access a variety of monitoring data including environmental and human matrices, food, and products for a variety of compounds. Outside of Europe, the National Center for Health Statistics ([cdc.gov/nchs/nhanes](https://cdc.gov/nchs/nhanes)) provides US human biomonitoring data for a limited number of compounds. Regarding environmental monitoring databases, the NORMAN Network ([norman-network.com](https://norman-network.com)) provides monitoring data on emerging substances in a variety of matrices, such as water, sediments, biota, SPM, soil, sewage sludge and air.

## 2.2 Data extraction

Automation of data collection from databases, scientific literature or gray literature, is a crucial component in building an automated EWS. Natural language processing (NLP) and machine learning have been used for this purpose in applications such as systematic reviewing (Khalil et al., 2022). The tools ExaCT (Kiritchenko et al., 2010), EPPI Reviewer (Park & Thomas, 2018) and Robot Reviewer ([robotreviewer.net](https://robotreviewer.net)) have been developed to extract information from scientific literature in an automated way and have been suggested as suitable tools for systematic reviewing.

Although these tools are useful for data extraction in systematic reviews, they may not be applicable for extracting information from databases. In principle, text mining tools could be developed for the specific purpose of data extraction from literature. While NLP is generally

D8.1: Report on the conceptual design of the SSbD toolbox	Dissemination level: PU
WP8: Concepts and toolboxes	Version: 1.0
Main Authors: Denis Sarigiannis	Page: 12

considered most suitable for processing text, searches in databases may require less complex language models, since the information is highly structured and not part of sentences. In this case, a simple script to automatically retrieve data from a database can be written.

Within the PARC project, a tool (AOP-helpFinder) has been developed in a collaboration of WPs 5, 6 and 8, to build AOPs. This tool combines text mining (Natural Language ToolKit) and graph theory (Dijkstra algorithm) to identify, extract and score knowledge published in scientific articles (i.e., the PubMed database), such as chemical-biological event relationships. A new version of the AOP-helpFinder was released in June 2023 with a new module to search for event-event linkages, and with visualization options to interpret the results. It is planned to adapt AOP-helpFinder 2.0 to characterize emergence in the EWS. AOP-helpFinder 2.0 can be accessed via [aop-helpfinder.u-paris-sciences.fr](http://aop-helpfinder.u-paris-sciences.fr) (Jaylet et al., 2023).

### 2.3 Use of non-target analytical data, digital archiving and retrospective suspect screening

High resolution mass spectrometry (HRMS), combined with liquid chromatography (LC-HRMS) and gas chromatography (GC-HRMS), has improved our ability to create precise mass spectra of organic chemicals across various matrices. These advanced methods generate vast amounts of data, capturing almost all the chemicals contained in a sample subject to the inherent limitations in analytical procedures (e.g., sampling techniques, enrichment methods, and solvents used) and instrumental analysis (e.g., ionizability, selectivity, sensitivity, and resolution). The datasets obtained from these techniques can include thousands of chemicals. Analyzing these extensive data can be challenging and call for automated approaches to search for the occurrence of chemicals. In addition, re-visits of these chromatograms should be possible for retrospective screening as science progresses and for example, new standards become available. The NORMAN network has undertaken significant efforts in this regard, systematically archiving data from environmental monitoring campaigns in the NORMAN Database System ([link.springer.com](http://link.springer.com)). Especially, a digital sample freezing platform (DSFP; [dsfp.norman-data.eu](http://dsfp.norman-data.eu)) for LC- and GC-HRMS data features an Application Programming Interface (API) that enables the automated retrieval of exposure information for any chemical, along with their semi-quantitative concentration levels ([sciencedirect.com](http://sciencedirect.com)). This retrospective suspect screening of any chemical within the digitally archived data opens up possibilities for utilizing these tools in an EWS. This tool has already been used by regulators (e.g., ECHA, UBA, HELCOM) to collect evidence for potentially hazardous chemicals ([echa.europa.eu](http://echa.europa.eu)).

### 2.4 Data curation

To enable generation of high-quality computational data, according to the EWS workflow in Figure 1, correct and readable structural information is critical. After data collection a methodology is required to unambiguously identify chemicals and relate a chemical structure to the collected information. If only names are available for a new compound, CAS number

D8.1: Report on the conceptual design of the SSbD toolbox	Dissemination level: PU
WP8: Concepts and toolboxes	Version: 1.0
Main Authors: Denis Sarigiannis	Page: 13

verification is required. This can be achieved with various computational tools, such as the `stdnum` library available in various scripting environments including Python ([arthurdejong.org/python-stdnum](http://arthurdejong.org/python-stdnum)) or Java ([metamolecular.com](http://metamolecular.com)) as well as other tools (Wang et al., 2020).

Although CAS numbers are unique identifiers, they do not describe the chemical structure and the same compound may be related to multiple CAS numbers. Structural information can be obtained from CAS numbers using the USEPA Chemistry Dashboard; the NCI/CADD Chemical Identifier Resolver ([cactus.nci.nih.gov/chemical/structure](http://cactus.nci.nih.gov/chemical/structure)) implemented for Python ([cirpy.readthedocs.io](http://cirpy.readthedocs.io)) or KNIME which are both suitable interfaces for an EWS. This structural information can be presented in a variety of formats depending on the computational tool to be applied for these structures. Simplified Molecular Input Line Entry System (SMILES) is the most commonly used format for representing the chemical structure of a compound using only a line of letters and symbols. Most *in silico* tools require SMILES to make predictions. Although they can accurately describe 2D molecular structures, this format does not possess 3D information. Additionally, SMILES can be formatted in different ways and still refer to the same structure, causing a risk of duplication. InChIKeys on the other hand are not commonly used as input for models but they provide precise information on chemical structure and can only be presented in a specific way, thus are useful for identifying duplicates. After identifying chemical structures, the data have to be curated to enable comparable model predictions. Structure curation may include “washing” of ions, i.e., representing explicit or implicit hydrogens or merging duplicates (Fourches et al., 2010). It should also be able to identify chemical mixtures, polymers, chemicals with rare elements or substances of unknown or variable composition, complex reaction products or biological materials (UVCBs), all of which may not be compatible with computational hazard tools. Both obtaining, formatting and curating of chemical structures could be part of an automated workflows in an EWS context (Bruks et al., 2021, 2022; Chelcea et al., 2020; Gadaleta et al., 2018).

### 3. Exposure-based warning systems

Besides, effect-based indications, environmental and/or human exposure is an important piece of information in an EWS. Here we present approaches to determine external and internal exposure based on human and environmental models. Overall external exposure can be assessed by evaluating the potential of chemicals to distribute, accumulate and persist in various environmental compartments and matrices from which organisms can take up chemicals, such as air, water or soil. For human exposure, food is an important exposure pathway, but others exist as well, e.g., indoor dust and air, dermal contact to consumer products etc. Internal exposure modeling aims at determining dose at target and can take on different levels of complexity of modeling. In order to determine internal or external exposure of emerging contaminants, many computational tools are available for various matrices,

D8.1: Report on the conceptual design of the SSbD toolbox	Dissemination level: PU
WP8: Concepts and toolboxes	Version: 1.0
Main Authors: Denis Sarigiannis	Page: 14

including both empirical models based on machine learning and existing data as well as mechanistic models.

### 3.1 External exposure models

Understanding the potential of emerging chemicals to distribute into various environmental matrices is relevant for identifying NERCs as it gives an indication of potentially exposed species as well as of possible exposure routes. This information can then be used to parameterize the exposure routes of these models, select relevant species for internal exposure modeling, and increase our understanding on sensitive species and targets of effects.

#### 3.1.1 Human exposure models

Humans can be exposed to NERCs via food, water, air, or through reproduction or dermal uptake. In addition to exposure routes, the extent of exposure can vary depending on whether it occurs to workers, consumers or the general population. Thus, simple screening models for dietary intake, inhalation and skin exposure are discussed below while more complex toxicokinetic models with multiple exposure routes and different populations are detailed under internal exposure models. Exposure measurements and assessment via monitoring data are the subject of other work packages in PARC.

High-throughput exposure (HTE) models may be suitable for screening for human exposure to NERCs as they are generic, cover multiple exposure routes and allow for integration in an EWS workflow (Wambaugh et al., 2019a,b). The SHEDS-HT model (Isaacs et al., 2014) for indoor environment is an example that has been implemented in R and offers chemical screening capability with limited need for parameters. It includes various routes of exposure and is maintained by the USEPA. RAIDAR-ICE is another indoor exposure model adapted for use in Excel and actively maintained, which would be suitable for screening, including a physiologically based kinetic (PBK) model for various exposure routes (Li et al., 2018). The EUSES tool ([echa.europa.eu](http://echa.europa.eu)) includes several exposure scenarios including ConsExpo for consumer exposure or SimpleBox for environmental multimedia fate modeling.

Models have also been developed for specific exposure pathways such as exposure from food contact material or cosmetics (Biryol et al., 2017; Csiszar et al., 2016). However, these approaches may conflict with the overall idea of a broad approach to the identification of NERCs, to avoid overlooking compounds of significance. A few models aiming to predict occupational exposure have been developed (Daniels et al., 2003; U.S. Occupational Safety and Health Administration, n.d.), but are still limited in their applicability to screen many compounds simultaneously, making them not yet suitable for incorporation in an EWS.

EFSA has developed the Dietary EXposure (DietEX) tool to estimate dietary exposure to substances present in food (e.g. intentionally added or naturally present chemicals, contaminants, proteins, novel food ingredients). Individual consumption data from the EFSA

D8.1: Report on the conceptual design of the SSbD toolbox	Dissemination level: PU
WP8: Concepts and toolboxes	Version: 1.0
Main Authors: Denis Sarigiannis	Page: 15

Comprehensive European Food Consumption Database are used to estimate the mean and the 95th percentile exposure for different age classes and specific population groups in several EU countries. Consumption records are codified according to the FoodEx2 food classification and description system. Macronutrients and micronutrients, including many chemical compounds and constituents that are added to food for flavor, color, taste, texture and aroma can be identified in the FoodDB database ([foodb.ca](http://foodb.ca)).

The Rapid Assessment of Contaminant Exposure (RACE) tool provides exposure estimates (mean, median and 95th percentile) of different population groups to chemical contaminants from single food items and compares the result to the health-based guidance value or other relevant toxicological reference values. RACE evaluates and characterizes the risk to consumers for the substance under evaluation, based on the assessment of toxicological properties and dietary exposure.

The main differences between the two EFSA's tools are the following: (i) in DietEX several food/food categories can be considered in the exposure estimation while in RACE, dietary exposure estimates are only possible for a single food item, and (ii) the RACE tool is limited to chemical contaminants that EFSA has assessed while DietEX does not present that limitation, and (iii) RACE evaluates and classifies risk for evaluation of contaminants in food while DietEX ends at exposure estimation.

### 3.1.2 Environmental fate models

The UN Stockholm Convention for POPs classifies POPs according to persistence, bioaccumulation, long-range transport and toxicity (pops.int). Similarly, but with slightly different quantitative criteria, REACH includes a PBT assessment, reflecting persistence, bioaccumulation and toxicity. Recently, an analogous concept of persistence, mobility and toxicity (PMT) has been discussed, which is now being implemented in the EU classification, labeling and packaging (CLP) legislation ([echa.europa.eu](http://echa.europa.eu)).

Assessing the environmental fate of compounds using computational tools is an important step in the EWS since it gives an indication of risk of exposure. Fate models aim at determining the distribution of a chemical between defined compartments and can be used to estimate a predicted environmental concentration (PEC). However, specific emission scenarios for compounds present a challenge for an automated tool since expert judgment is needed. Generally, emission data are sparse and usually non-existing for NERCs. Spatially resolved fate models have been developed to describe atmospheric or multi-media transport (Ciffroy et al., 2016; Falakdin et al., 2022; M. Armitage et al., 2007). Software platforms exist for fate modeling including the BETR North America (MacLeod et al., 2001), EpiSuite (usepa.gov), SimpleBox (Hollander et al., 2016), CoZMo-POP (Wania et al., 2006), SoilPCA (Kim et al., 2018), NEM (Breivik et al., 2021), USEtox (Rosenbaum et al., 2008), and the PiFs model (Fantke et al., 2016). The Merlin-Expo tool (Ciffroy et al., 2016) also provides several multi-media models which can be used to predict environmental fate. The latest version of the SimpleBox tool

D8.1: Report on the conceptual design of the SSbD toolbox	Dissemination level: PU
WP8: Concepts and toolboxes	Version: 1.0
Main Authors: Denis Sarigiannis	Page: 16

(Hollander et al., 2016) shows good potential for use in an EWS, since it is designed to be generic in chemical applicability, has been implemented in R, making it easy to integrate in a larger workflow and is actively being upgraded and maintained. Some of the challenges with incorporating such models into an EWS include the specific simulation scenarios required for some models, the limited amount of data available to validate these models and the lack of applicability domain (AD) assessment.

Empirical persistence models have been used in previous attempts to develop EWS or identify possible chemicals of concern (Arp et al., 2017; Bruks et al., 2021, 2022; Holmberg et al., 2021). One frequently applied model is the EpiSuite BioWIN3 model that can be used to estimate aerobic and anaerobic biodegradability and thus indicate potential for persistence in the environment ([epa.gov](https://www.epa.gov)). This tool has been widely used in environmental research and by agencies to prioritize chemicals. The model does however not distinguish between persistence in various environmental matrices and does not include an AD assessment. In contrast, the VEGAHUB offers a selection of qualitative and quantitative persistence models for sediment, soil, water and air. In addition, this tool provides reliability estimates based on AD which can be used to adjust the weights and influence the scores of each compound in the EWS.

Intrinsic physico-chemical properties determine the fate of chemicals including their environmental transport and distribution. Mobility is related to water solubility as well as distribution between water and organic matter (Neumann & Schliebner, 2019). Other fate related key properties include vapor pressure, partitioning coefficients, half-lives in various media etc. Both EpiSuite and VEGAHUB provide models for predicting these properties.

## 3.2 Internal exposure models

Models that predict internal concentrations in organisms are important tools for identifying potential risks of chemicals since estimated concentrations in an organism or target tissue indicate exposure at targets of toxicity in contrast to environmental fate models predicting environmental concentrations or levels in exposure media. There are various approaches to estimate internal exposure including predictive models for bioconcentration or bioaccumulation as well as compartmental models.

### 3.2.1 Bioconcentration and bioaccumulation models

Bioaccumulation is commonly quantified using metrics of bioconcentration factors (BCFs), bioaccumulation factors (BAFs) and n-octanol-water partition coefficients (Kow). The BCF is the concentration of the chemical in the organism's tissue relative to the concentration in the surrounding water or soil, while the BAF measures the concentration of the chemical in the organism's tissue relative to the concentration in the organism's food or other external sources. BCF, BAF and Kow are suitable for characterizing bioconcentration and bioaccumulation in aquatic organisms, but may not be as relevant for assessing bioaccumulation in some terrestrial organisms where biota-to-soil accumulation factors and

D8.1: Report on the conceptual design of the SSbD toolbox	Dissemination level: PU
WP8: Concepts and toolboxes	Version: 1.0
Main Authors: Denis Sarigiannis	Page: 17

octanol-air partition coefficients may be important. (Gobas et al., 2016; Gobas et al., 2003; Kelly & Gobas, 2001).

BCFs or BAFs are typically estimated using empirical data or models that take into account various physicochemical properties of the chemical and biological information on the organism. Several BCF/BAF models have been published for various organisms, such as fish, birds, and invertebrates. Bioaccumulation models for terrestrial species were reviewed by Gobas et al., (2016). Few empirical models exist for predicting BCF in other species than fish, which is in part due to lack of experimental data to build such models. The few existing empirical BCF models for terrestrial species generally display limited AD focusing on mainly non-ionic organic compounds, which makes them less suitable for an integration into an EWS.

A widely used model for BCF and BAF predictions in the field of environmental toxicology, is the BCFBAF model in EpiSuite developed by the US EPA. The BCFBAF model uses two approaches to estimate these properties. The first method employs a linear regression model based on log Kow as well as correction factors for various functional groups. It includes empirical data for the bioconcentration and bioaccumulation of chemicals in fish and other aquatic organisms. The second approach estimates BCF and BAF at several trophic levels using predicted metabolic half-lives in fish with the Arnot-Gobas method. The VEGAHUB platform offers four BCF models which include CAESAR, Meylan, Arnot-Gobas and KNN-Read-across. The platform includes a variety of exposure (and effect) models. Some of these models have been used to predict properties of potential risk chemicals under the Annex III criteria of REACH. The advantages of the VEGAHUB platform compared to many other software or published models include the selection of many different models and the assessment of an AD. A reliability score indicates whether the predicted compound falls within the AD of the model or not. This kind of reliability assessment can be considered in a potential weighting of results as described in section 1.2.

Both EpiSuite and VEGAHUB BCF models have been used in EWS (Bruks et al., 2021) and for prioritization of NERCs. They both allow a relatively rapid calculation of data on large numbers of chemicals. In addition, VEGAHUB is available as an implementation in KNIME, a software which has been previously used to build automated workflows for data collection, curation and analysis. Empirical BCF and BAF models are useful tools for quickly identifying the bioaccumulation potential of a compound, primarily for fish. However, they can be inaccurate as they do not consider species-specific physiology or ADME properties. In order to predict internal exposure to emerging compounds, empirical models may not be sufficient and mechanistic models may be required for the EWS.

### 3.2.2 Compartmental models

Simple one-compartment models can be used to predict internal concentrations in organisms. These models assume that the organism is represented by a single compartment and that the chemical is uniformly distributed throughout the compartment. The concentration of the

D8.1: Report on the conceptual design of the SSbD toolbox	Dissemination level: PU
WP8: Concepts and toolboxes	Version: 1.0
Main Authors: Denis Sarigiannis	Page: 18

chemical in the compartment is predicted using equations that take into account the chemical's physicochemical properties, the organism's metabolic rate, and the rate of chemical elimination. One-compartment models are relatively simple and easy to use, but they have limitations in predicting the distribution of chemicals in more complex organisms with multiple compartments.

The high throughput toxicokinetic (HTTK) package maintained by the USEPA has shown great potential as a screening tool for identifying compounds of concern. HTTK offers a one-compartment and a three-compartment model for hundreds of compounds, thus providing a way to quickly simulate internal exposure in humans, rats, mice, rabbits and dogs. However, parameters such as oral uptake rate and intrinsic clearance rate are needed from *in vitro* measurements, limiting the applicability of the model for less-studied compounds. These models are available through R script and can therefore be incorporated both in a script-based computational workflow or in a software such as KNIME where a scripting node can be part of the workflow.

Wiecek et al., (2019) presents a generic human one-compartment model and PBK model with the aim of performing forward dosimetry for human health risk assessment of chemicals (contaminants) in food. It is presented as case study XIII in the OECD guidance for PBK modeling, Annex 4 ([one.oecd.org](http://one.oecd.org)) together with the code which is available via the EFSA knowledge junction ([zenodo.org](http://zenodo.org)). As for HTTK, the main uncertainty regards data availability for metabolic parameters which need to be measured *in vitro*. Furthermore, these presented models focus on humans and a few laboratory animals, thus they do not provide information in a broader environmental context.

A compartmental model presented by Hendriks et al., (2001) was developed to simulate accumulation of chemicals for different trophic levels based on Kow as well as a few species-specific parameters. This model can be incorporated in an EWS. Although the model aims to be generic, the AD does not include extremely hydrophobic or hydrophilic compounds, nor exposure through air. Nonetheless, the model showed good performance even when compared to more recent and more complex PBTK models for fish (Stadnicka et al., 2012) and could be useful for a large variety of species in the EWS.

Another type of compartmental models are *in vitro* kinetic models aiming at describing kinetics of compounds in *in vitro* systems to better understand effective dose and therefore improve the accuracy of *in vitro* to *in vivo* extrapolation (IVIVE). The Armitage model (Armitage et al., 2014) and later the Honda model (Honda et al., 2019) have been incorporated as a function in R thus allowing for easy incorporation of these models into the designed EWS.

Overall compartmental models are promising screening tools, requiring only few parameters to predict internal concentrations. However, they do need *in vitro* data for biotransformation parameterization. Furthermore, only few models exist for species other than humans and a

D8.1: Report on the conceptual design of the SSbD toolbox	Dissemination level: PU
WP8: Concepts and toolboxes	Version: 1.0
Main Authors: Denis Sarigiannis	Page: 19

few laboratory animal species. Additionally, these models can be used to predict internal concentrations in the whole body but the limited number of compartments does not allow for assessment of compound accumulation in targets of toxicity.

### 3.2.3 Physiologically-based kinetic models

Physiologically-based kinetic (PBK) models have been developed with the purpose to predict organ concentration of various compounds and understand their ADME properties. PBK models reconstructing exposure through quantitative *in-vitro* to *in-vivo* extrapolation are considered New Approach Methodologies (NAM) for improving human risk assessment (Cronin et al., 2022; Deepika & Kumar, 2023; Spinu et al., 2020). These approaches can reduce the need for animal testing and also predict risks at an early level. As discussed in Deliverable 8.2, the use of PBK models shows great potential in identifying chemicals of risk in an EWS context.

Many compound-specific PBK models have been developed which can be employed for better understanding dose at target. A recent overview of these chemical-specific PBK models highlighted that they had mainly been developed for low molecular weight compounds which generally follow Lipinski's rule of 5 (Thompson et al., 2021). Consequently, there are large knowledge gaps in the chemical applicability domain of these specific models. Some of these models require large amounts of data including *in vivo* parametrization data for their development. As such, chemical-specific PBK models may not currently be suitable for application in EWS. Applying read-across from one chemical to another could be a way to use such models for new compounds. However, this would require similar chemical structure and properties as already parametrized chemicals.

Generic PBK models would allow for the simulation of toxicokinetics of a wider range of emerging compounds and thus support the identification of NERCs based on dose at potential targets of toxicity. Additionally, developing PBK without the use of *in vivo* data is a goal for the next-generation PBK models which has been set by various governmental agencies (Paini et al., 2019). Generic models employing NAMs for parameterization could therefore be both a more suitable and more ethical choice for incorporation in EWS. It is also motivated by the lower use of data to predict ADME properties. Although some data is still required, much of the parameterization can be achieved using *in silico* and *in vitro* data when *in vivo* data are unavailable. Various generic models have been developed and employed in recent years with different applications in the field of risk assessment (Fragki et al., 2022; Jongeneelen & Berge, 2011; Lautz et al., 2020; Paini et al., 2019; Tebby et al., 2020).

The European Human Biomonitoring Initiative (HBM4EU) has collected human biomonitoring data for assessments of chemical exposure of the European population. It has also derived health-based guidance values. The HBM4EU report on PBTK models (Sarigiannis et al., 2018)

D8.1: Report on the conceptual design of the SSbD toolbox	Dissemination level: PU
WP8: Concepts and toolboxes	Version: 1.0
Main Authors: Denis Sarigiannis	Page: 20

presents a variety of compound-specific and generic models including INTEGRA (Sarigiannis et al., 2014; Sarigiannis et al., 2019), IndusChemFate, MENTOR-3P (Georgopoulos & Liou, 2006), and MERLIN-Expo tool (Brochot & Quindroit, 2018). This report concludes that efforts should be directed at generic, multi-compartmental and multi-route models with parameterization using *in silico* and *in vitro* data. The QSAR models for parameterization should be expanded to model chemico-biological interactions such as intestinal absorption or renal clearance. Additionally, further development of mixture PBK models would be an improvement for chemical prioritization and identification of NERCs.

The OECD guidance document on the characterization, validation and reporting of PBK models for regulatory purposes highlight several case studies of PBK models with the potential to be used for identifying chemical risks (OECD, 2021). Among these, two generic PBK models are presented, including a model focusing on farm animals (Lautz et al., 2020) and one for fish species (Grech et al., 2017). Since each animal in the farm and the fish model has multiple compartments, it may be computationally demanding to predict and incorporate all of these in an EWS. Instead, the EWS should be simplified and focus on key compartments of interest such as muscle for all animals used in the meat and fishing industry, milk for all the species used in the dairy industry and eggs for chicken. Both models use a combination of *in silico*, *in vitro* and *in vivo* data for model calibration and validation and face the same challenges in parametrization of absorption rates, clearance rates and plasma fraction unbound.

Interestingly, Tebby et al., (2020) investigated the performance of this generic model using three different parametrization scenarios, based on QSAR models, additional *in vitro* data and further additions of *in vivo* data, respectively. These authors found that the first two scenarios performed similarly for predicting dose at targets while parameterization using *in vivo* data showed much higher accuracy. Nonetheless, they concluded that models that rely upon existing databases or simple QSAR models are more feasible when providing PBK models for a large number of substances in lower-tier calculations. Considering that EWS aim at identifying emerging issues at an early stage, higher uncertainty in model calculations can likely be accepted, possibly including refinements at later stages.

The models presented in this section with provided scripts, such as HTK can be easily incorporated in a computational workflow in an EWS context (Lautz et al 2020; Grech et al., 2017). The on-line tools could potentially be automatically accessed provided the output files are consistently formatted in machine readable formats. Although generic PBK models are promising tools in the development of EWS, they still pose some challenges related to parameter estimates. Specifically, the estimations of 1) metabolic clearance rates in a quick high-throughput manner, and 2) oral and intestinal absorption as well as excretion rates have been identified as critical gaps and uncertainties in the identification of NERCs (Paini et al., 2019; Punt et al., 2022).

D8.1: Report on the conceptual design of the SSbD toolbox	Dissemination level: PU
WP8: Concepts and toolboxes	Version: 1.0
Main Authors: Denis Sarigiannis	Page: 21

### 3.2.3.1 Population models

PBK models have also been used to simulate entire populations. Combined with calculated safety limits based on effect data, population models can be used to assess what proportion of the population and which sub-groups are exposed to levels above safety limits (Breen et al., 2022; Paini et al., 2019; Ring et al., 2017; Testai et al., 2021). This offers indications of how the predicted or observed exposure and effect data may affect populations. One example is a lifetime PBK model developed for effects of PFAS compounds on humans where all age groups were included and which contained the flexibility of adding other specific population groups as well as variability within a group (Deepika et al. 2021). The previously discussed HTKK package offers the possibility to quickly simulate population kinetics using HTKK-Pop with already provided physiological parameters distributions (Breen et al., 2022; Ring et al., 2017). Ring et al., (2017) identified chemicals of risk based on hazard from *in vitro* screening assays and exposure via HTKK-Pop simulations, and used HTKK-Pop to identify demographic groups at higher exposure risk. The HTKK-Pop has also been used to determine age-refined toxicological points of departure in combination with effect data from *in vitro* assays (Silva & Kwok, 2023). Although HTKK-Pop can be widely applied in the risk assessment of chemicals, it may be more suitable for detailed investigation of compounds rather than NERC identification, for several reasons. Firstly, the chemical data variability and distribution that is required for population modeling is usually not available. Secondly, specific exposure scenarios need to be chosen for the simulation which may differ between compounds and thus may be challenging to automate within an EWS. Finally, the model outputs can be complex including interactions between physiological and chemical parameters and thus may require manual analysis for each compound. In summary, population models are an important computational tool for next generation risk assessment, but may not be practical for EWS.

## 3.3 Transformation models

Transformation products are of importance to consider in an EWS as they can be of higher toxicological concern than the parent compound. They can be the result of abiotic and biotic transformation processes. Biotransformation rates in the form of intrinsic clearance have been discussed in previous sections as a crucial and challenging parameter in predictions of internal exposure and bioaccumulation. Thus, *in silico* tools for predicting biotransformation rates and transformation products play an important role in several places of the EWS. Predicting transformation reactions *in silico* has proven to be challenging with existing models showing a great need for improvement in both applicability domain and accuracy. One of the challenges related to biotic transformation is the large inter-individual and inter-species variability in metabolic enzymes resulting in different products and rates. Existing tools focus on specific species or metabolic enzyme groups.

Most models predicting intrinsic clearance rates are tailored for pharmaceuticals. These often

D8.1: Report on the conceptual design of the SSbD toolbox	Dissemination level: PU
WP8: Concepts and toolboxes	Version: 1.0
Main Authors: Denis Sarigiannis	Page: 22

have a limited applicability domain and are thus likely not applicable for a large variety of compounds (Chang et al., 2009; Paixão et al., 2010). An exception is the kM/Half-life model incorporated in the VEGAHUB, which can predict primary biotransformation half-lives and rate constants in fish based on the work by Arnot et al., (2008a,b). However, models for other species are lacking which highlights a major data gap and the need for developing *in silico* tools for this purpose. Being able to predict clearance rates would increase the use of compartmental and PBK models and their application to emerging chemicals for which *in vitro* and *in vivo* data are unavailable.

Several open-source software aims at predicting transformation products (Zheng et al., 2021) including the OECD QSAR toolbox, CTS, BioTransformer and EAWAG-BBD/PPS as well as commercial software, such as Meteor Nexus. Among these, CTS, EAWAG-BBD/PPS and Meteor Nexus provide likelihoods of formation of a given transformation product while the others only predict products. These tools which do not provide likelihood of formation of specific metabolites are less suited for incorporation in an EWS, since they may predict unlikely metabolites that would increase computational efforts without identifying NERCs. This should be considered depending on the purpose and context of the EWS. Note also that here we will not focus on commercial software due to issues with their incorporation in an open EWS.

For the purpose of predicting abiotic transformation products, the CTS platform can be used to predict likely transformation products for hydrolysis and reduction reactions. Additionally, microbial transformation products can be predicted using EAWAG-BBD/PPS. Degradability in the environment can also be assessed using the two models in VEGAHUB which predict ready biodegradability. One of these models has been applied for a national EWS (Bruks et al., 2021, 2022). The JANUS tool automatically provides environmental degradation products (using > 200 degradation pathways) and it generates predictions for degradation products as well.

Biotransformation products in organisms are relevant for assessment of toxicity or ADME properties in addition to those of parent compounds. CTS can be used but it only includes phase I metabolism in humans. Although the QSAR Toolbox, BioTransformer and Meteor Nexus can predict phase II metabolites, the first two do not provide a likelihood score while the last one is a commercial software. Thus, there is a need for more comprehensive biological transformation models.

Overall, the existing transformation models mainly focus on transformation products rather than rates. Structures of predicted products can be run through the EWS to estimate concerning properties. These results can be accounted for in the assessment of the parent compound, for example via a scoring system. However, one of the main challenges observed with exposure models remains, i.e., the determination of biotransformation rates. This is therefore a major data gap which needs to be filled in order to improve the accuracy of these models as well as their usability for EWS.

D8.1: Report on the conceptual design of the SSbD toolbox	Dissemination level: PU
WP8: Concepts and toolboxes	Version: 1.0
Main Authors: Denis Sarigiannis	Page: 23

## 4. Effect-based warning systems

NERCs can be identified either from the exposure or the effect side. Understanding the capacity of a chemical to elicit an effect which leads to an adverse outcome is a crucial part of a hazard identification. Additionally, the mechanism of action may provide information about sensitive target species or tissues. Predicting possible effects of compounds using computational tools, such as QSAR models, is a common procedure in research-, industry- and regulatory-based hazard identification. In addition, novel bioinformatic tools and their potential use in EWS are discussed in this section, along with systems biology models.

### 4.1 Adverse outcome pathways (AOPs)

Adverse outcome pathways (AOPs) have become an important part of next generation RA and are addressed explicitly in WP5 of PARC. AOPs provide a framework for characterizing effects and improving our understanding, starting from a molecular initiating event (MIE), via key events (KE) which can lead to an adverse outcome (AO) for individuals and populations. AOPs have a great potential for structuring effect data but they present a simplification of complex biological processes. This includes a better understanding of potential risks by connecting a predicted MIE to a potential AO. Furthermore, AOPs can combine experimental data from *in vitro* systems with *in silico* data. Ram et al., (2022) described how AOPs can be used as a structural framework to integrate multiple data sources. A recent EFSA report described a need for automation to integrate heterogeneous mechanistic data for the purposes of risk assessment in an AOP-like knowledge network (Blümmel et al., 2023). Similarly, the EWS proposed in this report could use knowledge of AOPs to identify NERCs. Predictive models of MIEs could be linked with those that can predict KEs in order to identify NERCs.

Additionally, the exposure models discussed in section 3 could be incorporated into an AOP framework. This would increase our understanding of which combinations of exposure and effects could lead to an added risk. For example, in an AOP framework, a compound can be predicted as antagonistic to ER which is an MIE leading to reproductive dysfunction in fish ([aopwiki.org/aops/30](http://aopwiki.org/aops/30)). If the exposure modeling also predicts an environmental distribution of this compound with presence in water together with high liver accumulation, the main site of ER expression in fish, then this compound may be considered a potential risk. This approach could be taken a step further with the incorporation of quantitative predictions from PBK models in order to build quantitative AOPs (qAOPs). qAOPs have been discussed in the context of NAMS and RA as potential future *in silico* tools. Spinu et al., (2020) identified five probabilistic qAOPs and ten mechanistic ones developed for the purpose of ecotoxicological RA, four for human RA and one for screening and prioritization. Although more qAOPs have been developed in recent years, they remain complex, in particular regarding their requirement for calibration and validation data, and therefore limited in their applicability (Perkins et al., 2019; Spinu et al., 2020). Foran et al., (2019), however, proposed a modular approach for qAOP in order to facilitate the development of these networks for the purpose

D8.1: Report on the conceptual design of the SSbD toolbox	Dissemination level: PU
WP8: Concepts and toolboxes	Version: 1.0
Main Authors: Denis Sarigiannis	Page: 24

of screening and prioritization of environmental pollutants. Thus, qAOP may be a valuable tool for EWS but may prove too complex and limited in their AD. Nonetheless, their potential for the future for animal-free RA as well as for computational EWS is evident (Cronin et al., 2022; Deepika & Kumar, 2023; Spinu et al., 2020). Existing AOPs could be incorporated in the EWS (Figure 1), possibly as part of the scoring step.

## 4.2 QSAR models

*In silico* models including QSARs can be used for a fast screening of substances and provide property values for a large number of endpoints and substances. Endpoints can be chosen to represent MIEs or KEs, which may give an indication of potential adverse outcomes. Although many promising models have been published in the scientific literature, this section will focus on openly available tools that are widely used to assess chemical hazards, including regulatory applications, and that are kept updated.

It is an advantage if a variety of platforms and systems can be used in the EWS, since each has different collections of models, which can be complementary. Inventories exist where models have been collected, for example for predictions of properties of relevance for REACH ([life-concertreach.eu](http://life-concertreach.eu)). There are several free platforms that could be useful including QsarDB ([qsar.db.org](http://qsar.db.org)), VEGAHUB ([vegahub.eu](http://vegahub.eu)), EPISuite ([epi-suite](http://epi-suite)), QSAR TOOLBOX ([qsartoolbox.org](http://qsartoolbox.org)) and OPERA ([ntp.niehs.nih.gov](http://ntp.niehs.nih.gov)). In addition, the Danish (Q)SAR Database provides predictions from a large range of models ([qsar.food.dtu.dk](http://qsar.food.dtu.dk)).

Some tools integrate multiple models and predict a large range of properties simultaneously. JANUS (available from VEGAHUB and mainly used for prioritization) is an example that provides experimental data (when available) and predicted values for a series of substances processed in a batch. Property values can be given for carcinogenicity, mutagenicity and reprotoxicity (CMR), persistence, bioaccumulation and toxicity (PBT) and endocrine disruption. If available, multiple models are used for the same endpoint and the results integrated into a single value, considering the reliability of each value. Similarly, multiple experimental values are integrated. This approach results in a property value together with its uncertainty. The VEGAHUB currently includes 112 individual models related to human toxicity (50), ecotoxicity (26), environmental fate and distribution, physical-chemical properties and toxicokinetics. They allow predictions for a total of 48 endpoints, including acute and chronic toxicity, developmental toxicity, endocrine disruption, genotoxicity, and mutagenicity. These models are developed by different groups, organizations and projects and combined into a single software, making it a practical option for implementation into an EWS.

## 4.3 Bioinformatic tools

Bioinformatics provide tools and techniques for analyzing and interpreting large amounts of data related to exposure and hazards that could be useful in an EWS. Algorithms will provide the edges and nodes with other scientific disciplines (e.g., statistics, mathematics, biology,

D8.1: Report on the conceptual design of the SSbD toolbox	Dissemination level: PU
WP8: Concepts and toolboxes	Version: 1.0
Main Authors: Denis Sarigiannis	Page: 25

physics, and chemistry) in a machine-readable format. Bioinformatics can contribute to an EWS via tools that scrutinize genetic information, as well as metabolic data which can provide indications that link a compound with an adverse outcome. In addition, the algorithms can also deliver the connections between a decision network that will evaluate a chemical and its terms of use. These techniques can also enable the development of new biomarkers for use in EWS. Effects can be explored across multiple scales of complexity, from molecular and cellular to tissue and organismal levels, to time and dose. Network sciences allow the integration of heterogeneous data sources and the quantification of their interactions. In this concept, the perception that human diseases are associated with a few dominant factors (reductionist) is replaced by the view of diseases as the outcome of many weak contributors (holistic). Therefore, they are sensitive approaches that can be of interest to consider. For example, entire pathways rather than single genes associated with diverse toxicity endpoints or adverse outcomes can be considered.

In network sciences, features such as node, i.e., the entity on which the network is building up (chemical, genes, pathways, clinical outcomes), edges (that can represent physical interactions, functional interactions or simply connection between multi-scale data), degree of distribution around a node and the forming hubs, degree of clustering or betweenness centrality are usually considered to assess the properties of the biological network. The integration of omics data plays a key role in creating these connections. Omics technologies, such as genomics, proteomics, and metabolomics, provide data for a comprehensive and holistic assessment of the biological effects of chemicals, and thus indications of potential health risks at an early stage. For example, over the last two decades, a number of toxicogenomics studies have been performed taking advantage of the mature microarray technology and, more recently, the RNA sequencing technology. These approaches have enabled the identification of differentially expressed genes (DEGs) in cells or tissues exposed to more than 20 000 chemical substances, which can be accessed through public repositories ([lincsproject.org](http://lincsproject.org)). The DEG profiles can be used to suggest biological mechanisms related to toxicity endpoints and/or to biomarkers of toxicity. ToxicDb is one of the major sources of harmonized toxicogenomics data, which allow researchers to evaluate the effects of toxicants by gene expression ([doi.org/10.1093/nar/gkaa390](https://doi.org/10.1093/nar/gkaa390)).

One of the advantages of using omics data in EWS is the ability to identify specific molecular mechanisms and pathways that are affected by a chemical ([doi.org/10.1289/ehp.1409157](https://doi.org/10.1289/ehp.1409157)). This information can be used to predict potential health risks, such as cancer or reproductive toxicity. Despite these advantages, there are also some limitations to using omics data in EWS. One limitation is that the data generated by omics technologies can be complex and difficult to interpret, requiring specialized expertise and advanced analytical tools (under development in PARC). Omics technologies can be directly connected to an EWS or retrieved from linked databases for meta-analysis or other data interpretation. Furthermore, bioinformatics also

D8.1: Report on the conceptual design of the SSbD toolbox	Dissemination level: PU
WP8: Concepts and toolboxes	Version: 1.0
Main Authors: Denis Sarigiannis	Page: 26

includes biostatistics, which can be an important component of an EWS. Statistics provide methodological tools to identify patterns and temporal trends in the data and to make predictions about the potential health risks associated with the exposure to a chemical. Biostatistics can also be used to determine the level of uncertainty associated with the predictions that will be provided by AI models, which is essential for making informed risk management decisions. Tools such as score matching, regression analysis, and sensitivity analysis can be used to analyze these uncertainties.

#### 4.4 Other computational tools

Other computational approaches are available in the literature, however, some of these methods are highly complex, chemical-specific and require much experimental information and manual processing, which may limit their applications in an automated EWS. Molecular dynamics simulations are one of these computational approaches which can be applied to analyze toxicodynamic interactions of emerging compounds with potential target receptors. However, it requires in-depth knowledge on specific receptors, 3D structures of compounds and is computationally expensive (Lu et al., 2018; Sakkiah et al., 2019).

Another source of information that could be analyzed through bioinformatic approaches is the image-based high content screening technology. Cell painting assays use a mixture of 6 fluorescent dyes to stain different compartments of a cell which are then represented by morphological features (i.e., shape, intensity, texture, among others). These assays allow the investigation of the morphological perturbation profiles induced by large sets of chemicals (more than 30 000) to study phenotypic changes associated with different modes of action. Combinations of these data with omics information could be promising, for example to assess the potential link between transcriptomic changes, the corresponding alterations in cell morphology and toxicology endpoints.

Systems biology (SB) models are mathematical models that aim to simulate the complex interactions between different biological processes and systems (Figure 3). They are the missing link between genotype and phenotype incorporating multiscale biological networks. These models involve quantitative predictions and can be a relevant tool in EWS, as they can provide insights into potential health risks associated with chemical exposure. Chemicals can affect human health by alterations in cellular uptake and efflux controlled by membrane transporter proteins, interactions of chemicals and metabolites with tissue macromolecules, immune response activation, metabolic transformations, and much more. To understand these interactions at different levels of biological organization, molecular and cellular targets, different SB models are required.

SB models use mathematical equations and computational methods to simulate the interactions between different biological systems and processes. They can be used to predict the effects of chemicals at the tissue, cellular, and molecular level. These mathematical

D8.1: Report on the conceptual design of the SSbD toolbox	Dissemination level: PU
WP8: Concepts and toolboxes	Version: 1.0
Main Authors: Denis Sarigiannis	Page: 27

models incorporate metabolic control analysis (MCA), flux balance analysis (FBA), and elementary model analysis (EMA) to understand network associated toxicity pathways and much more. SB models often contain multiple pathways to elucidate the crosstalk occurring at the cellular level after exposure. These models can be simplified and converted to AOPs. They follow the mechanism of action of a disruptor when a human is exposed to a chemical. As a result of this exposure, hazards can be identified, and the compound can be evaluated for its terms of use. In the same direction, systems biology models also provide the identification of genetic and epigenetic changes. These changes include anomalies in the genome, transcriptome, proteome, or metabolome of a human and the way these processes are regulated.

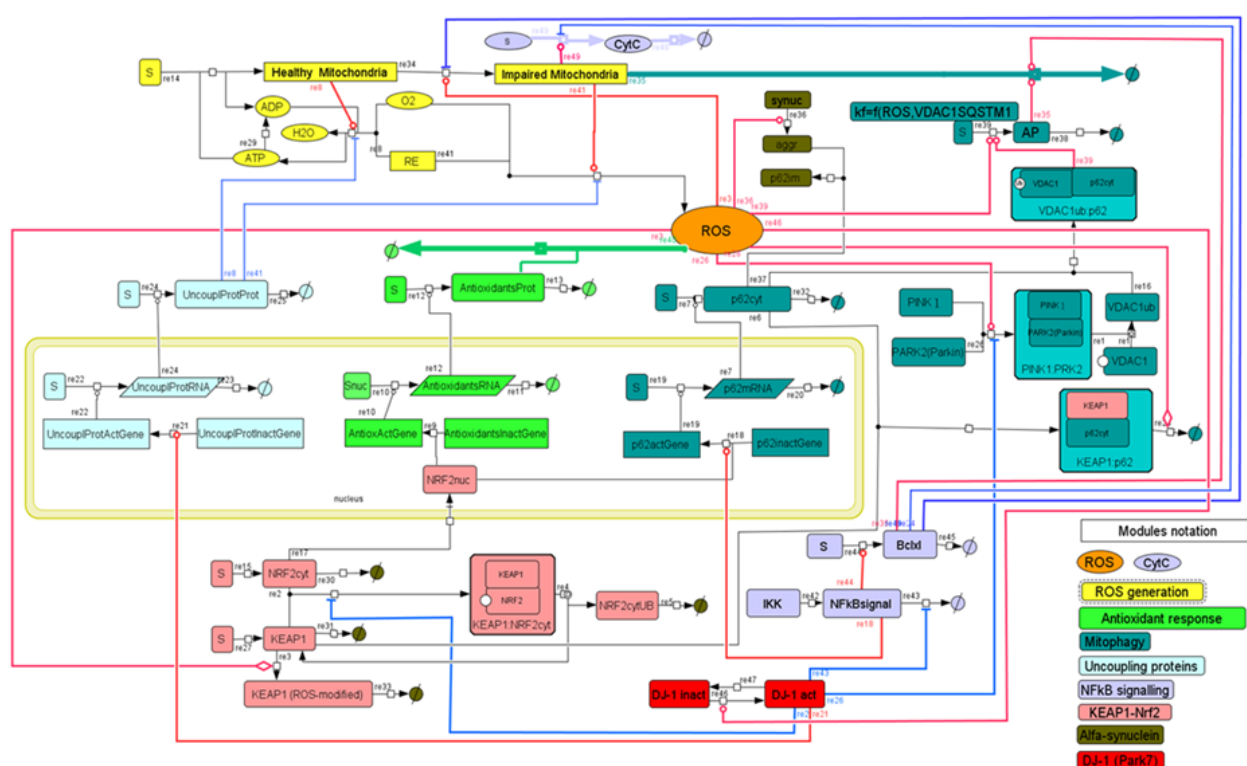


Figure 3. Example of Systems Biology model of Reactive oxygen species (ROS) management (Kolodkin et al., 2020)

Biological models can be used as a tool to identify interactions between chemicals at different states and levels of biological processes. It includes cases of all-or-none type of effect, i.e., additive responses (all effect) or antagonistic actions (none effect). This procedure involves the integration of SB and AI methodologies with pharmacokinetics to study such phenomena (Figure 4). Computational systems biology models have been reported to integrate chemical-disease information, epidemiology studies and diverse sources of mechanisms of action in a human environmental disease network that can help in the decision making for chemical risk assessment.

D8.1: Report on the conceptual design of the SSbD toolbox	Dissemination level: PU
WP8: Concepts and toolboxes	Version: 1.0
Main Authors: Denis Sarigiannis	Page: 28

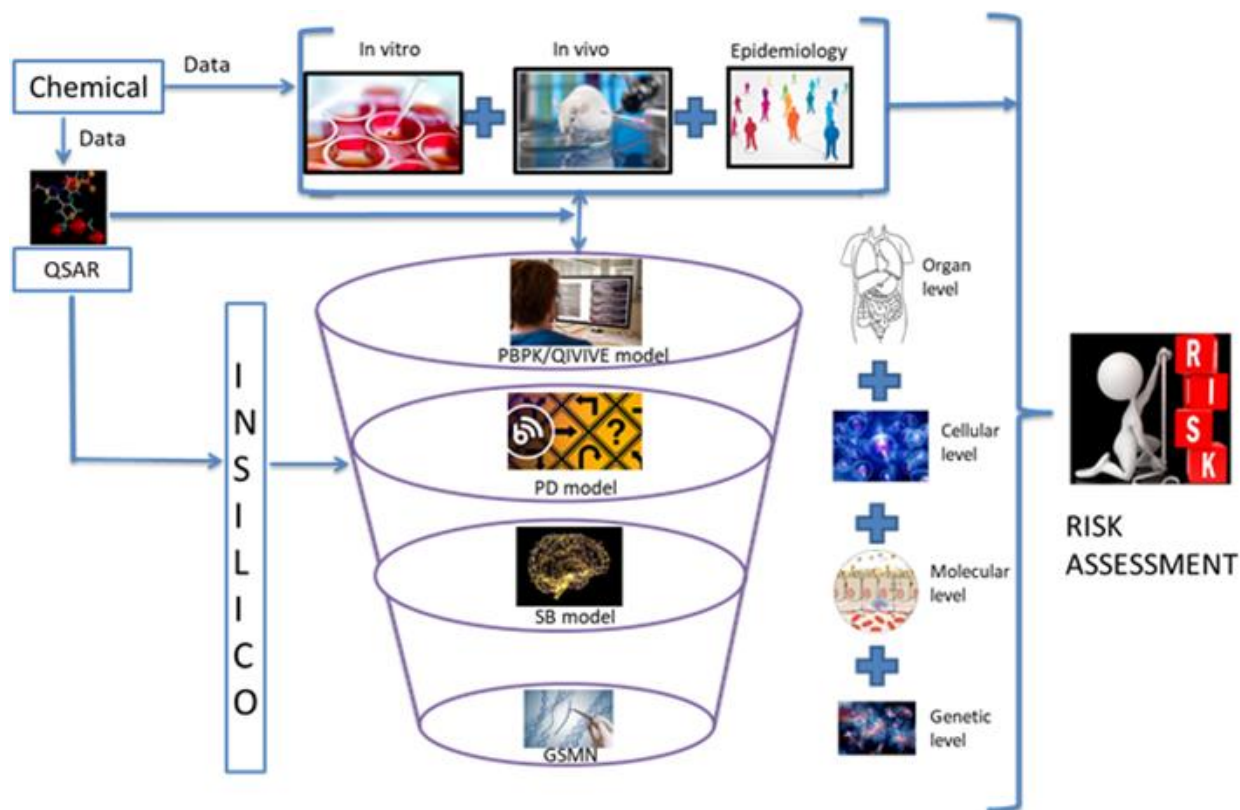


Figure 4. Integrated Translational approach for risk assessment. System biology models can act as a tool for translation and prediction of risk to human species (Deepika et al., 2020).

## 5. Integration of data for identification of potential NERCs

The final step of the EWS is the identification of NERCs based on data predicted or collected in previous steps. A critical aspect in this phase is the quality of data. Data curation as described earlier aimed at reaching chemical structures of high quality to be used in computational tools. However, experimental data identified in the first step and estimated data will have varying quality and may as well include erroneous data. It will thus be important to analyze quality of data and identify potential outliers in the weighing and scoring phase of the EWS. Overall, integrating this large variety of data sources and endpoints is complex and challenging and will require suitable cut-off values and simulation scenarios. Expert judgements will be needed to evaluate the different types of data, including potential use of weighting factors. At the same time, the process should be automated as much as possible, to enable a fast and impartial identification of NERCs from big data. Possible approaches include decision trees, scoring systems and grouping or clustering of chemicals or endpoints. In Figure 5, a decision tree illustrates how new signals could be treated, resulting in possible NERCs, uncertain NERCs and chemicals of possible low concern.

D8.1: Report on the conceptual design of the SSbD toolbox	Dissemination level: PU
WP8: Concepts and toolboxes	Version: 1.0
Main Authors: Denis Sarigiannis	Page: 29

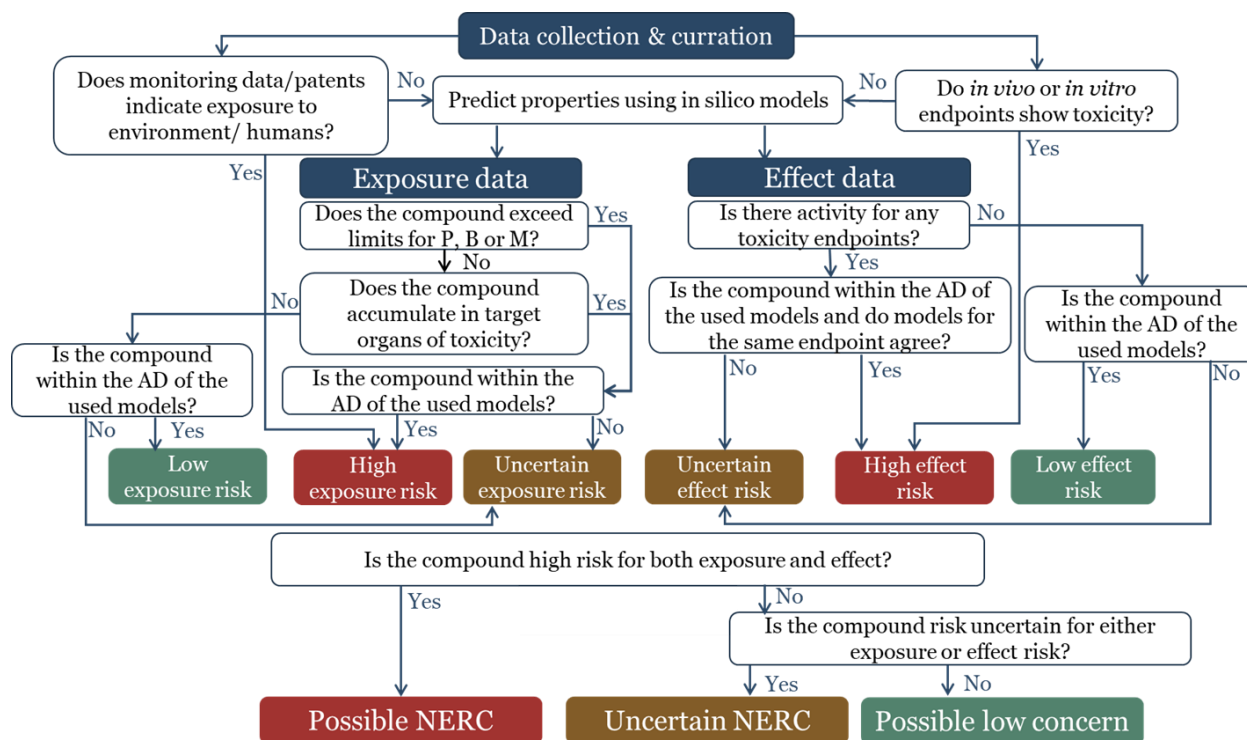


Figure 5. Example decision tree for integrating data of the EWS and classifying compounds into three broad categories

The NORMAN network has presented an early warning system (NormaNEWS) which is based on compounds that have been identified as potentially hazardous. Digitally archived high-resolution mass spectrometry data ([dsfp.norman-data.eu](https://dsfp.norman-data.eu)) are searched for these suspected compounds, providing semi-quantitative data on environmental occurrence. This information can be combined with the NORMAN prioritization framework. In the prioritization approach of the Swedish Chemicals Agency cut-off values were used for various predicted hazard properties including the reliability of the prediction in terms of model AD (Bruks et al., 2022). Another approach applied by the Swedish Chemicals Agency included effect predictions for various endocrine receptors combined with patent information to indicate likely human exposure (Swedish Chemicals Agency, 2020). Additional activities in search of NERCs include the annual IT mass screening by ECHA of registration dossiers covering both hazard profiles and exposure estimates ([echa.eu](https://echa.eu)). The Danish EPA is using combinations of QSAR models for hazard classification and self classification. ([danish epa](https://danish.epa)) More complex scoring and grouping approaches can be applied, including quantitative scoring values and cut-offs instead of binary classification schemes or grouping into specific classes of concern, such as PBTs, EDCs, and neurotoxicants. An example of such an approach was presented by Hartmann et al., (2023), where a scoring system integrating multiple models for PMT endpoints was developed. These authors used different weights for different endpoints and structural alerts resulting in a final score from 0 to 1.

D8.1: Report on the conceptual design of the SSbD toolbox	Dissemination level: PU
WP8: Concepts and toolboxes	Version: 1.0
Main Authors: Denis Sarigiannis	Page: 30

A suitable integration and scoring system for EWS should incorporate the following key aspects: (i) weighing of exposure and effect data of different kinds to yield normalized hazard scores, (ii) weighing of experimental versus estimated data, (iii) reliability assessment of scores including consideration of AD, (iv) quantitative cut-off values of scores to indicate potential risks, and (v) suitability of the scoring system for automation.

## 5.1 Scoring of chemicals

A simple approach to combine and score data from different sources is to apply heat mapping as was illustrated in Figure 2. This approach provides an easy visualization and ranking of data. However, it requires thresholds of each parameter which can be linked to color codes. Ideally, these thresholds would follow existing legislation, but will likely not be available for NERCs. Scores can be user-defined, based on expert judgement and criteria chosen for the EWS. A more elaborate approach in combining evidence would be to use multicriteria decision analysis (MCDA). MCDA covers various methodologies that can simultaneously handle data of different kinds including qualitative, semi-quantitative and quantitative data to guide decision making. MCDA can be divided into full aggregation methods, using all data that are converted to a comparable scale, or outranking methods which apply pairwise comparisons. Recently, MCDA was used to compare substitutes of brominated flame retardants based on several hazard estimates (Zheng et al., 2019) and was later extended to include their transformation products (Zheng et al., 2021). It still requires to set thresholds, as for heat mapping, with the difference that several data types are evaluated simultaneously in MCDA. This is an advantage in an EWS application that aims to combine different types of data. The scoring in itself can be binary but also condense raw data in continuous but normalized scales.

Scoring of chemicals in a prioritization context is also addressed in WP4 of PARC, where criteria, indicators and thresholds are currently being reviewed and assessed for their potential use in a prioritization scheme for environmental monitoring projects. Since the EWS includes similar elements, a coordination of these activities is useful and has been ensured through regular information exchange and collaboration.

## 5.2 Grouping and clustering

The collected data of the EWS can be used to group or cluster chemicals allowing for read-across and identification of potentially hazardous chemicals with patterns similar to known pollutants. One option would be to apply unsupervised machine learning approaches to analyze data. Since some chemicals are data rich and their risks are well documented, these could form references, i.e., positive controls in order to identify NERCs with similar descriptor patterns. Methods include principal component analysis (PCA) and clustering techniques (Chelcea et al., 2020; Rännar & Andersson, 2010). Multivariate statistical analysis can be used to explore a complex chemical space and for reducing dimensionality before applying grouping approaches (Chelcea et al., 2020; Oprea & Gottfries, 2001; Rännar & Andersson, 2010;

D8.1: Report on the conceptual design of the SSbD toolbox	Dissemination level: PU
WP8: Concepts and toolboxes	Version: 1.0
Main Authors: Denis Sarigiannis	Page: 31

Stenberg et al., 2009; Virshup et al., 2013). Hierarchical clustering is a computational method which allows for grouping of chemicals based on their descriptors and this has been widely used in the field of chemistry for varying purposes (Backman et al., 2011; Chelcea et al., 2020; Haranczyk & Holliday, 2008; Miller, 2002; Rännar & Andersson, 2010). Clustering is the basis of the VEGA tool (Viganò et al., 2022) of the VEGAHUB for read-across based on chemical structures, and other approaches including k-nearest neighbors have recently been used for grouping and read-across (Golden et al., 2023; Kutsarova et al., 2021).

Clustering can also be used to characterize hazard profiles that potential NERCs may possess, thus giving an indication of what testing would be required to verify these suspected properties. Figure 6 offers a visualization of this approach with reference compounds. For example, if compound 1 is a known non-persistent EDC and compound 2 is a known persistent and bioaccumulating compound, while compounds 3 and 4 are potential NERCs, they are likely to share properties with compound 1 and 2, respectively.

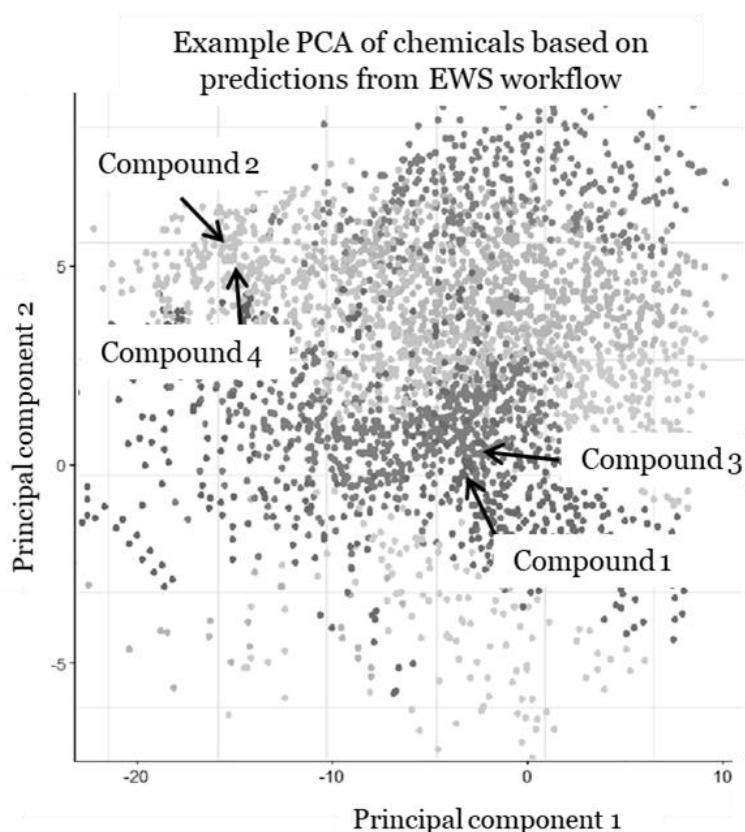


Figure 6. Example of a clustering exercise which can be used to group compounds into various categories of NERCs based on the outcome of predictive models within the EWS

### 5.3 Computational platforms for data integration

This report has described various computational tools which may be integrated into an EWS for identifying NERCs. Various programming languages, such as Python, R or Javascript could

D8.1: Report on the conceptual design of the SSbD toolbox	Dissemination level: PU
WP8: Concepts and toolboxes	Version: 1.0
Main Authors: Denis Sarigiannis	Page: 32

be employed to script the EWS. Since the individual tools have no common programming language, the workflow would have to be programmed in one script which synchronously invokes and runs scripts of the various tools. Translating each tool to a common language would be another option, however this would be more time-consuming and require expert maintenance, making it a less practical option.

Another possibility to build the EWS is to implement it in software, such as KNIME, that has been previously employed to build workflows for chemical data collection, processing and chemical prioritization (Bruks et al., 2021, 2022; Chelcea et al., 2020; Gadaleta et al., 2018). Additionally, some of the presented tools, such as the VEGAHUB and the Chemical Identifier Resolver are implemented as separate units in KNIME, called nodes, making them simple to integrate in the EWS. Another advantage is the fact that scripts can be easily integrated in KNIME as nodes, which offers a practical way to incorporate R, Python and Javascript models separately into the workflow. Finally, KNIME has a graphical user interface which makes it easy to visualize the whole workflow as well as the different components of the EWS. Such a software may therefore present a more visualizable and user-friendly alternative compared to a large script.

## 6. Conclusions and Future Aspects

Developing a sensitive, fast and effective early warning system is a great challenge for science and stakeholders. Nevertheless, pollution issues that have been able to build up over years and decades highlight the necessity for quick and efficient early warnings, upon which further studies can be initiated or actions be taken. Developments in analytical chemistry and data science have made early warnings more feasible. New screening techniques and bioanalytical tools can identify an unprecedented number of chemicals and effects, respectively. Models can predict exposure and effects, and natural language processing and bioinformatic tools can process large amounts of highly diverse data including grey literature and social media information. Despite these promising developments, combining those techniques in an efficient EWS that is needed for identifications of emerging issues is a challenging task. Experimental data are still the key to many computational approaches and essentially lacking for NERCs, and applicability domains of existing models might not accommodate NERCs. Transformation products and mixtures are often not considered, neither are certain compound groups, such as ionizable chemicals, polymers and UVCBs, very few species have been targeted, and uncertainties are difficult to quantify and include in an evaluation and interpretation of results. In addition, practical challenges exist in merging different types of models and approaches which might not be compatible a priori. The actual evaluation of EWS results, including the development of a scoring system, also poses a great challenge with regard to the definition of critical warning levels. Depending on the context, these might also vary, potentially requiring some flexibility, for example in weighting of evidence. Finally, the

D8.1: Report on the conceptual design of the SSbD toolbox	Dissemination level: PU
WP8: Concepts and toolboxes	Version: 1.0
Main Authors: Denis Sarigiannis	Page: 33

system has to be built on a computational platform that can be maintained and implemented, still being user friendly and open access. In summary, building an EWS with a strong computational component is timely and would provide a major step towards an early recognition and consequently, better assessment of chemical risks.

## 7. References

- Armitage, J. M., Wania, F., & Arnot, J. A. (2014). Application of Mass Balance Models and the Chemical Activity Concept To Facilitate the Use of in Vitro Toxicity Data for Risk Assessment. *Environmental Science & Technology*, *48*(16), 9770–9779. <https://doi.org/10.1021/es501955g>
- Armitage, J.M., Cousins, I.T., Hauck, M., Harbers, J.V., & Huijbregts, J.M.A. (2007). Empirical evaluation of spatial and non-spatial European-scale multimedia fate models: Results and implications for chemical risk assessment. *Journal of Environmental Monitoring*, *9*(6), 572–581. <https://doi.org/10.1039/B700680B>
- Arnot, J. A., Mackay, D., & Bonnell, M. (2008a). Estimating metabolic biotransformation rates in fish from laboratory data. *Environmental Toxicology and Chemistry*, *27*(2), 341–351. <https://doi.org/10.1897/07-310R.1>
- Arnot, J. A., Mackay, D., Parkerton, T. F., & Bonnell, M. (2008b). A database of fish biotransformation rates for organic chemicals. *Environmental Toxicology and Chemistry*, *27*(11), 2263–2270. <https://doi.org/10.1897/08-058.1>
- Arp, H. P. H., Brown, T. N., Berger, U., & Hale, S. E. (2017). Ranking REACH registered neutral, ionizable and ionic organic chemicals based on their aquatic persistency and mobility. *Environmental Science: Processes & Impacts*, *19*(7), 939–955. <https://doi.org/10.1039/C7EM00158D>
- Backman, T. W. H., Cao, Y., & Girke, T. (2011). ChemMine tools: An online service for analyzing and clustering small molecules. *Nucleic Acids Research*, *39*(suppl\_2), W486–W491. <https://doi.org/10.1093/nar/gkr320>
- Biryol, D., Nicolas, C. I., Wambaugh, J., Phillips, K., & Isaacs, K. (2017). High-throughput dietary exposure predictions for chemical migrants from food contact substances for use in chemical prioritization. *Environment International*, *108*, 185–194. <https://doi.org/10.1016/j.envint.2017.08.004>
- Blümmel, T., Rehn, J., Mereu, C., Graf, F., Kneuer, C., Wittkowski, P., Sonnenburg, A., Moore, A., Bech, K., van der Lugt, B., Bouwmeester, H., Kramer, N., & Dobrikov, T. (2023). Review of state-of-the-art AI tools and methods for screening, extracting and evaluating NAMs literature in the context of chemical risk assessment. *EFSA Supporting Publications*, *20*(1), 7815E. <https://doi.org/10.2903/sp.efsa.2022.EN-7815>

D8.1: Report on the conceptual design of the SSbD toolbox	Dissemination level: PU
WP8: Concepts and toolboxes	Version: 1.0
Main Authors: Denis Sarigiannis	Page: 34

- Breen, M., Wambaugh, J. F., Bernstein, A., Sfeir, M., & Ring, C. L. (2022). Simulating toxicokinetic variability to identify susceptible and highly exposed populations. *Journal of Exposure Science & Environmental Epidemiology*, 32(6), Article 6. <https://doi.org/10.1038/s41370-022-00491-0>
- Breivik, K., Eckhardt, S., S. McLachlan, M., & Wania, F. (2021). Introducing a nested multimedia fate and transport model for organic contaminants (NEM). *Environmental Science: Processes & Impacts*, 23(8), 1146–1157. <https://doi.org/10.1039/D1EM00084E>
- Brochot, C., & Quindroit, P. (2018). Modelling the Fate of Chemicals in Humans Using a Lifetime Physiologically Based Pharmacokinetic (PBPK) Model in MERLIN-Expo. In P. Ciffroy, A. Tediosi, & E. Capri (Eds.), *Modelling the Fate of Chemicals in the Environment and the Human Body* (pp. 215–257). Springer International Publishing. [https://doi.org/10.1007/978-3-319-59502-3\\_10](https://doi.org/10.1007/978-3-319-59502-3_10)
- Bruks, S., Andersson, P. L., Franke, V., & Wiberg, K. (2021). *Methods for early identification of chemicals that have the potential to harm human health or the environment*. The Toxicological Council. Swedish Chemicals Agency.
- Bruks, S., Zheng, Z., & Andersson, P. L. (2022). *Methods for early identification of chemicals that have the potential to harm human health or the environment*. The Toxicological Council. Swedish Chemicals Agency.
- Chang, C., Duignan, D. B., Johnson, K. D., Lee, P. H., Cowan, G. S., Gifford, E. M., Stankovic, C. J., Lepsy, C. S., & Stoner, C. L. (2009). The Development and Validation of a Computational Model to Predict Rat Liver Microsomal Clearance. *Journal of Pharmaceutical Sciences*, 98(8), 2857–2867. <https://doi.org/10.1002/jps.21651>
- Chelcea, I. C., Ahrens, L., Örn, S., Mucs, D., & Andersson, P. L. (2020). Investigating the OECD database of per- and polyfluoroalkyl substances – chemical variation and applicability of current fate models. *Environmental Chemistry*. <https://doi.org/10.1071/EN19296>
- Ciffroy, P., Alfonso, B., Altenpohl, A., Banjac, Z., Bierkens, J., Brochot, C., Critto, A., De Wilde, T., Fait, G., Fierens, T., Garratt, J., Giubilato, E., Grange, E., Johansson, E., Radomyski, A., Reschwann, K., Suci, N., Tanaka, T., Tediosi, A., ... Verdonck, F. (2016). Modelling the exposure to chemicals for risk assessment: A comprehensive library of multimedia and PBPK models for integration, prediction, uncertainty and sensitivity analysis – the MERLIN-Expo tool. *Science of The Total Environment*, 568, 770–784. <https://doi.org/10.1016/j.scitotenv.2016.03.191>
- Cronin, M. T. D., Spînu, N., & Worth, A. P. (2022). Will qAOPs modernise toxicology? *Computational Toxicology*, 21, 100199. <https://doi.org/10.1016/j.comtox.2021.100199>

D8.1: Report on the conceptual design of the SSbD toolbox	Dissemination level: PU
WP8: Concepts and toolboxes	Version: 1.0
Main Authors: Denis Sarigiannis	Page: 35

Csiszar, S. A., Ernstoff, A. S., Fantke, P., Meyer, D. E., & Jolliet, O. (2016). High-throughput exposure modeling to support prioritization of chemicals in personal care products. *Chemosphere*, *163*, 490–498. <https://doi.org/10.1016/j.chemosphere.2016.07.065>

Daniels, W., Lee, S., & Miller, A. (2003). EPA's Exposure Assessment Tools and Models. *Applied Occupational and Environmental Hygiene*, *18*(2), 82–86. <https://doi.org/10.1080/10473220301430>

Deepika D, Sharma RP, Schuhmacher M, Kumar V. (2021). Risk Assessment of Perfluorooctane Sulfonate (PFOS) using Dynamic Age Dependent Physiologically based Pharmacokinetic Model (PBPK) across Human Lifetime. *Environmental Research*, 199:111287. doi: 10.1016/j.envres.2021.111287. Epub 2021 May 14. PMID: 34000270.

Deepika, D., Sharma, R. P., Schuhmacher, M., Kumar, V. (2020). Integrated Translational Approach for Neurotoxicity- A Systematic Review, *Critical Reviews in Toxicology* 50 (5), 424-438. DOI: 10.1080/10408444.2020.1763253

Deepika, D., & Kumar, V. (2023). The Role of “Physiologically Based Pharmacokinetic Model (PBPK)” New Approach Methodology (NAM) in Pharmaceuticals and Environmental Chemical Risk Assessment. *International Journal of Environmental Research and Public Health*, *20*(4), Article 4. <https://doi.org/10.3390/ijerph20043473>

European Commission. (2021). *Pilot of an EU early warning system for emerging chemical risks to the environment*. doi:10.2779/3213

Falakdin, P., Terzaghi, E., & Di Guardo, A. (2022). Spatially resolved environmental fate models: A review. *Chemosphere*, *290*, 133394. <https://doi.org/10.1016/j.chemosphere.2021.133394>

Fantke, P., Ernstoff, A. S., Huang, L., Csiszar, S. A., & Jolliet, O. (2016). Coupled near-field and far-field exposure assessment framework for chemicals in consumer products. *Environment International*, *94*, 508–518. <https://doi.org/10.1016/j.envint.2016.06.010>

Foran, C., Rycroft, T., Keisler, J., Perkins, E., Linkov, I., & Garcia-Reyero, N. (2019). A modular approach for assembly of quantitative adverse outcome pathways. *ALTEX - Alternatives to Animal Experimentation*. <https://doi.org/10.14573/altex.1810181>

Fourches, D., Muratov, E., & Tropsha, A. (2010). Trust, but verify: On the importance of chemical structure curation in cheminformatics and QSAR modeling research. *Journal of Chemical Information and Modeling*, *50*(7), 1189–1204. <https://doi.org/10.1021/ci100176x>

Fragki, S., Piersma, A. H., Westerhout, J., Kienhuis, A., Kramer, N. I., & Zeilmaker, M. J.

D8.1: Report on the conceptual design of the SSbD toolbox	Dissemination level: PU
WP8: Concepts and toolboxes	Version: 1.0
Main Authors: Denis Sarigiannis	Page: 36

(2022). Applicability of generic PBK modelling in chemical hazard assessment: A case study with IndusChemFate. *Regulatory Toxicology and Pharmacology*, 136, 105267. <https://doi.org/10.1016/j.yrtph.2022.105267>

Gadaleta, D., Lombardo, A., Toma, C., & Benfenati, E. (2018). A new semi-automated workflow for chemical data retrieval and quality checking for modeling applications. *Journal of Cheminformatics*, 10(1), 60. <https://doi.org/10.1186/s13321-018-0315-6>

Geng, Z., Zhao, S., Tao, G., & Han, Y. (2017). Early warning modeling and analysis based on analytic hierarchy process integrated extreme learning machine (AHP-ELM): Application to food safety. *Food Control*, 78, 33–42. <https://doi.org/10.1016/j.foodcont.2017.02.045>

Georgopoulos, P. G., & Lioy, P. J. (2006). From a Theoretical Framework of Human Exposure and Dose Assessment to Computational System Implementation: The Modeling ENvironment for TOtal Risk Studies (MENTOR). *Journal of Toxicology and Environmental Health, Part B*, 9(6), 457–483. <https://doi.org/10.1080/10937400600755929>

Gobas, F. A., Burkhard, L. P., Doucette, W. J., Sappington, K. G., Verbruggen, E. M., Hope, B. K., Bonnell, M. A., Arnot, J. A., & Tarazona, J. V. (2016). Review of existing terrestrial bioaccumulation models and terrestrial bioaccumulation modeling needs for organic chemicals. *Integrated Environmental Assessment and Management*, 12(1), 123–134. <https://doi.org/10.1002/ieam.1690>

Gobas, F. A. P. C., Kelly, B. C., & Arnot, J. A. (2003). Quantitative Structure Activity Relationships for Predicting the Bioaccumulation of POPs in Terrestrial Food-Webs. *QSAR & Combinatorial Science*, 22(3), 329–336. <https://doi.org/10.1002/qsar.200390022>

Golden, E., Ukaegbu, D. C., Ranslow, P., Brown, R. H., Hartung, T., & Maertens, A. (2023). The Good, The Bad, and The Perplexing: Structural Alerts and Read-Across for Predicting Skin Sensitization Using Human Data. *Chemical Research in Toxicology*, 36(5), 734–746. <https://doi.org/10.1021/acs.chemrestox.2c00383>

Grech, A., Brochot, C., Dorne, J.-L., Quignot, N., Bois, F. Y., & Beaudouin, R. (2017). Toxicokinetic models and related tools in environmental risk assessment of chemicals. *Science of The Total Environment*, 578, 1–15. <https://doi.org/10.1016/j.scitotenv.2016.10.146>

Haranczyk, M., & Holliday, J. (2008). Comparison of Similarity Coefficients for Clustering and Compound Selection. *Journal of Chemical Information and Modeling*, 48(3), 498–508. <https://doi.org/10.1021/ci700413a>

Hartmann, J., Rorije, E., Wassenaar, P. N. H., & Verbruggen, E. (2023). Screening and

D8.1: Report on the conceptual design of the SSbD toolbox	Dissemination level: PU
WP8: Concepts and toolboxes	Version: 1.0
Main Authors: Denis Sarigiannis	Page: 37

prioritising persistent, mobile and toxic chemicals: Development and application of a robust scoring system. *Environmental Sciences Europe*, 35(1), 40.  
<https://doi.org/10.1186/s12302-023-00749-w>

Hendriks, A. J., van der Linde, A., Cornelissen, G., & Sijm, D. T. H. M. (2001). The power of size. 1. Rate constants and equilibrium ratios for accumulation of organic substances related to octanol-water partition ratio and species weight. *Environmental Toxicology and Chemistry*, 20(7), 1399–1420. <https://doi.org/10.1002/etc.5620200703>

Hollander, A., Schoorl, M., & van de Meent, D. (2016). SimpleBox 4.0: Improving the model while keeping it simple. *Chemosphere*, 148, 99–107.  
<https://doi.org/10.1016/j.chemosphere.2016.01.006>

Holmberg, R., Wedebye, E. B., Nikolov, N. G., & Tyle, H. (2021). How many potential vPvM/PMT substances have been registered under REACH? - VPvM/PMT-screening by using the Danish (Q)SAR database. In *How many potential vPvM/PMT substances have been registered under REACH? - vPvM/PMT-screening by using the Danish (Q)SAR database* [Report]. Danmarks Tekniske Universitet (DTU).

Honda, G. S., Pearce, R. G., Pham, L. L., Setzer, R. W., Wetmore, B. A., Sipes, N. S., Gilbert, J., Franz, B., Thomas, R. S., & Wambaugh, J. F. (2019). Using the concordance of in vitro and in vivo data to evaluate extrapolation assumptions. *PLOS ONE*, 14(5), e0217564.  
<https://doi.org/10.1371/journal.pone.0217564>

Isaacs, K. K., Glen, W. G., Egeghy, P., Goldsmith, M.-R., Smith, L., Vallero, D., Brooks, R., Grulke, C. M., & Özkaynak, H. (2014). SHEDS-HT: An Integrated Probabilistic Exposure Model for Prioritizing Exposures to Chemicals with Near-Field and Dietary Sources. *Environmental Science & Technology*, 48(21), 12750–12759.  
<https://doi.org/10.1021/es502513w>

Jaylet, T., Coustillet, T., Jornod, F., Margaritte-Jeannin, P., Audouze, K. (2023). AOP-helpFinder 2.0: Integration of an event-event searches module. *Environment International*, 177, 109017, <https://doi.org/10.1016/j.envint.2023.108017>

Jongeneelen, F. J., & Berge, W. F. T. (2011). A Generic, Cross-Chemical Predictive PBTK Model with Multiple Entry Routes Running as Application in MS Excel; Design of the Model and Comparison of Predictions with Experimental Results. *The Annals of Occupational Hygiene*, 55(8), 841–864. <https://doi.org/10.1093/annhyg/mer075>

Kelly, B. C., & Gobas, F. A. (2001). Bioaccumulation of persistent organic pollutants in lichen-caribou-wolf food chains of Canada's Central and Western Arctic. *Environmental Science & Technology*, 35(2), 325–334. <https://doi.org/10.1021/es0011966>

Kolodkin, A., Sharma, R.P., Colangelo, A., Ignatenko, A., Martorana, F., Jennen, D., Briede, J.J., Brady, N., Barberis, M., Mondeel, Thierry D.G.A., Papa, M., Kumar, V., Peters, B.,

D8.1: Report on the conceptual design of the SSbD toolbox	Dissemination level: PU
WP8: Concepts and toolboxes	Version: 1.0
Main Authors: Denis Sarigiannis	Page: 38

Skupin, A., Alberghina, L., Balling, R., Westerhoff, H.V. (2020). ROS networks: designs, aging, Parkinson's disease and precision therapies, *NPJ Systems Biology and Applications*, 6, 34 (2020). DOI: 10.1038/s41540-020-00150-w

Khalil, H., Ameen, D., & Zarnegar, A. (2022). Tools to support the automation of systematic reviews: A scoping review. *Journal of Clinical Epidemiology*, 144, 22–42. <https://doi.org/10.1016/j.jclinepi.2021.12.005>

Kim, K.-E., Jung, J. E., Lee, Y., & Lee, D. S. (2018). Ranking surface soil pollution potential of chemicals from accidental release by using two indicators calculated with a multimedia model (SoilPCA). *Ecological Indicators*, 85, 664–673. <https://doi.org/10.1016/j.ecolind.2017.11.010>

Kiritchenko, S., de Bruijn, B., Carini, S., Martin, J., & Sim, I. (2010). ExaCT: Automatic extraction of clinical trial characteristics from journal publications. *BMC Medical Informatics and Decision Making*, 10(1), 56. <https://doi.org/10.1186/1472-6947-10-56>

Kutsarova, S., Mehmed, A., Cherkezova, D., Stoeva, S., Georgiev, M., Petkov, T., Chapkanov, A., Schultz, T. W., & Mekenyan, O. G. (2021). Automated read-across workflow for predicting acute oral toxicity: I. The decision scheme in the QSAR toolbox. *Regulatory Toxicology and Pharmacology*, 125, 105015. <https://doi.org/10.1016/j.yrtph.2021.105015>

Lautz, L. S., Dorne, J. L. C. M., Oldenkamp, R., Hendriks, A. J., & Ragas, A. M. J. (2020). Generic physiologically based kinetic modelling for farm animals: Part I. Data collection of physiological parameters in swine, cattle and sheep. *Toxicology Letters*, 319, 95–101. <https://doi.org/10.1016/j.toxlet.2019.10.021>

Li, L., Westgate, J. N., Hughes, L., Zhang, X., Givehchi, B., Toose, L., Armitage, J. M., Wania, F., Egeghy, P., & Arnot, J. A. (2018). A Model for Risk-Based Screening and Prioritization of Human Exposure to Chemicals from Near-Field Sources. *Environmental Science & Technology*, 52(24), 14235–14244. <https://doi.org/10.1021/acs.est.8b04059>

Liu, N., Bouzembrak, Y., van den Bulk, L. M., Gavai, A., van den Heuvel, L. J., & Marvin, H. J. P. (2022). Automated food safety early warning system in the dairy supply chain using machine learning. *Food Control*, 136, 108872. <https://doi.org/10.1016/j.foodcont.2022.108872>

Lu, L., Zhan, T., Ma, M., Xu, C., Wang, J., Zhang, C., Liu, W., & Zhuang, S. (2018). Thyroid Disruption by Bisphenol S Analogues via Thyroid Hormone Receptor  $\beta$ : In Vitro, in Vivo, and Molecular Dynamics Simulation Study. *Environmental Science & Technology*, 52(11), 6617–6625. <https://doi.org/10.1021/acs.est.8b00776>

MacLeod, M., Woodfine, D. G., Mackay, D., McKone, T., Bennett, D., & Maddalena, R. (2001).

D8.1: Report on the conceptual design of the SSbD toolbox	Dissemination level: PU
WP8: Concepts and toolboxes	Version: 1.0
Main Authors: Denis Sarigiannis	Page: 39

BETR North America: A regionally segmented multimedia contaminant fate model for North America. *Environmental Science and Pollution Research*, 8(3), 156–163.  
<https://doi.org/10.1007/BF02987379>

Miller, M. A. (2002). Chemical database techniques in drug discovery. *Nature Reviews Drug Discovery*, 1(3), Article 3. <https://doi.org/10.1038/nrd745>

Neumann, M., & Schliebner, I. (2019). *Protecting the sources of our drinking water: The criteria for identifying persistent, mobile and toxic (PMT) substances and very persistent and very mobile (vPvM) substances under EU Regulation REACH (EC) No 1907/2006*. Umwelt Bundesamt (UBA).

NORMAN. (2016). *NORMAN network of reference laboratories, research centres and related organisations for monitoring of emerging environmental substances*.  
<http://www.norman-network.net/>

OECD. (2021). *Guidance document on the characterisation, validation and reporting of Physiologically Based Kinetic (PBK) models for regulatory purposes* (Guidance No. 331; OECD Series on Testing and Assessment).  
<https://www.oecd.org/chemicalsafety/risk-assessment/guidance-document-on-the-characterisation-validation-and-reporting-of-physiologically-based-kinetic-models-for-regulatory-purposes.pdf>

Oprea, T. I., & Gottfries, J. (2001). Chemography: The Art of Navigating in Chemical Space. *Journal of Combinatorial Chemistry*, 3(2), 157–166.  
<https://doi.org/10.1021/cc0000388>

Paini, A., Leonard, J. A., Joossens, E., Bessems, J. G. M., Desalegn, A., Dorne, J. L., Gosling, J. P., Heringa, M. B., Klaric, M., Kliment, T., Kramer, N. I., Loizou, G., Louise, J., Lumen, A., Madden, J. C., Patterson, E. A., Proença, S., Punt, A., Setzer, R. W., ... Tan, Y. M. (2019). Next generation physiologically based kinetic (NG-PBK) models in support of regulatory decision making. *Computational Toxicology*, 9, 61–72.  
<https://doi.org/10.1016/j.comtox.2018.11.002>

Paixão, P., Gouveia, L. F., & Morais, J. A. G. (2010). Prediction of the in vitro intrinsic clearance determined in suspensions of human hepatocytes by using artificial neural networks. *European Journal of Pharmaceutical Sciences*, 39(5), 310–321.  
<https://doi.org/10.1016/j.ejps.2009.12.007>

Palmen, N. G. M. (2016). *Early warning systems to detect new and emerging risks in Europe* (No. 2016–0022). Dutch National Institute for Public Health and the Environment (RIVM).

Park, E. S., & Thomas, J. (2018). *Evidence synthesis software*. <https://doi.org/10.1136/bmjebm-2018-110962>

D8.1: Report on the conceptual design of the SSbD toolbox	Dissemination level: PU
WP8: Concepts and toolboxes	Version: 1.0
Main Authors: Denis Sarigiannis	Page: 40

- Pawar, G., Madden, J. C., Ebbrell, D., Firman, J. W., & Cronin, M. T. D. (2019). In Silico Toxicology Data Resources to Support Read-Across and (Q)SAR. *Frontiers in Pharmacology*, *10*. <https://www.frontiersin.org/articles/10.3389/fphar.2019.00561>
- Perkins, E. J., Gayen, K., Shoemaker, J. E., Antczak, P., Burgoon, L., Falciani, F., Gutsell, S., Hodges, G., Kienzler, A., Knapen, D., McBride, M., Willett, C., Iii, F. J. D., & Garcia-Reyero, N. (2019). Chemical hazard prediction and hypothesis testing using quantitative adverse outcome pathways. *ALTEX - Alternatives to Animal Experimentation*, *36*(1), 91–102. <https://doi.org/10.14573/altex.1808241>
- Punt, A., Louise, J., Pinckaers, N., Fabian, E., & van Ravenzwaay, B. (2022). Predictive Performance of Next Generation Physiologically Based Kinetic (PBK) Model Predictions in Rats Based on In Vitro and In Silico Input Data. *Toxicological Sciences*, *186*(1), 18–28. <https://doi.org/10.1093/toxsci/kfab150>
- Qin, X., Liu, J., Wang, Y., Liu, Y., Deng, K., Ma, Y., Zou, K., Li, L., & Sun, X. (2021). Natural language processing was effective in assisting rapid title and abstract screening when updating systematic reviews. *Journal of Clinical Epidemiology*, *133*, 121–129. <https://doi.org/10.1016/j.jclinepi.2021.01.010>
- Ram, R. N., Gadaleta, D., & Allen, T. E. H. (2022). The role of ‘big data’ and ‘in silico’ New Approach Methodologies (NAMs) in ending animal use – A commentary on progress. *Computational Toxicology*, *23*, 100232. <https://doi.org/10.1016/j.comtox.2022.100232>
- Rännar, S., & Andersson, P. L. (2010). A Novel Approach Using Hierarchical Clustering To Select Industrial Chemicals for Environmental Impact Assessment. *Journal of Chemical Information and Modeling*, *50*(1), 30–36. <https://doi.org/10.1021/ci9003255>
- Ring, C. L., Pearce, R. G., Setzer, R. W., Wetmore, B. A., & Wambaugh, J. F. (2017). Identifying populations sensitive to environmental chemicals by simulating toxicokinetic variability. *Environment International*, *106*, 105–118. <https://doi.org/10.1016/j.envint.2017.06.004>
- Rosenbaum, R. K., Bachmann, T. M., Gold, L. S., Huijbregts, M. A. J., Jolliet, O., Juraske, R., Koehler, A., Larsen, H. F., MacLeod, M., Margni, M., McKone, T. E., Payet, J., Schuhmacher, M., van de Meent, D., & Hauschild, M. Z. (2008). USEtox—the UNEP-SETAC toxicity model: Recommended characterisation factors for human toxicity and freshwater ecotoxicity in life cycle impact assessment. *The International Journal of Life Cycle Assessment*, *13*(7), 532–546. <https://doi.org/10.1007/s11367-008-0038-4>
- Sakkiah, S., Kusko, R., Tong, W., & Hong, H. (2019). Applications of Molecular Dynamics Simulations in Computational Toxicology. In H. Hong (Ed.), *Advances in Computational Toxicology: Methodologies and Applications in Regulatory Science* (pp.

D8.1: Report on the conceptual design of the SSbD toolbox	Dissemination level: PU
WP8: Concepts and toolboxes	Version: 1.0
Main Authors: Denis Sarigiannis	Page: 41

181–212). Springer International Publishing. [https://doi.org/10.1007/978-3-030-16443-0\\_10](https://doi.org/10.1007/978-3-030-16443-0_10)

Sarigiannis, D. A., Karakitsios, S., Dominguez-Romero, E., Papadaki, K., Brochot, C., Kumar, V., Schuhmacher, M., Sy, M., Mielke, H., Greiner, M., Mengelers, M., & Scheringer, M. (2019). Physiology-based toxicokinetic modelling in the frame of the European Human Biomonitoring Initiative. *Environmental Research*, *172*, 216–230. <https://doi.org/10.1016/j.envres.2019.01.045>

Sarigiannis, D., Karakitsios, S., Dominiguez-Romero, E., Papadaki, K., Brochot, C., Kumar, V., Schumacher, M., Sy, M., Mielke, H., Greiner, M., Mengelers, M., & Scheringer, M. (2018). *Review paper on PBTK/D models for the 1st set of priority compounds (D12.1)*. HBM4EU.

Sarigiannis, D., Karakitsios, S., Gotti, A., Loizou, G., Cherrie, J., Smolders, R., De Brouwere, K., Galea, K., Jones, K., Handakas, E., Papadaki, K., & Sleuwenhoek, A. (2014). Integra: 7th International Congress on Environmental Modelling and Software. *Proceedings - 7th International Congress on Environmental Modelling and Software: Bold Visions for Environmental Modeling, IEMSs 2014*, *2*, 1001–1008.

Silva, M., & Kwok, R. K.-H. (2023). Use of computational toxicology models to predict toxicological points of departure: A case study with triazine herbicides. *Birth Defects Research*, *115*(5), 525–544. <https://doi.org/10.1002/bdr2.2144>

Spinu, N., Cronin, M. T. D., Enoch, S. J., Madden, J. C., & Worth, A. P. (2020). Quantitative adverse outcome pathway (qAOP) models for toxicity prediction. *Archives of Toxicology*, *94*(5), 1497–1510. <https://doi.org/10.1007/s00204-020-02774-7>

Stadnicka, J., Schirmer, K., & Ashauer, R. (2012). Predicting Concentrations of Organic Chemicals in Fish by Using Toxicokinetic Models. *Environmental Science & Technology*, *46*(6), 3273–3280. <https://doi.org/10.1021/es2043728>

Stenberg, M., Linusson, A., Tysklind, M., & Andersson, P. L. (2009). A multivariate chemical map of industrial chemicals – Assessment of various protocols for identification of chemicals of potential concern. *Chemosphere*, *76*(7), 878–884. <https://doi.org/10.1016/j.chemosphere.2009.05.011>

Swedish Chemicals Agency. (2022). *PM1/22: Dataanalys av patentinformation med hjälp av artificiell intelligens*. <https://www.kemi.se/publikationer/pm/2022/pm1-22-dataanalys-av-patentinformation-med-hjalp-av-artificiell-intelligens#:~:text=Patentdata%20som%20r%C3%B6r%20en%20grupp,analysera%20deras%20text%2D%20och%20bildinneh%C3%A5ll>.

Swedish Chemicals Agency. (2020). *Rapport 5/17: Bisfenoler – en kartläggning och analys* [Text]. <https://www.kemi.se/en/publications/reports/2017/rapport-5-17-bisfenoler-->

D8.1: Report on the conceptual design of the SSbD toolbox	Dissemination level: PU
WP8: Concepts and toolboxes	Version: 1.0
Main Authors: Denis Sarigiannis	Page: 42

-en-kartläggning-och-analys

- Tebby, C., van der Voet, H., de Sousa, G., Rorije, E., Kumar, V., de Boer, W., Kruisselbrink, J. W., Bois, F. Y., Faniband, M., Moretto, A., & Brochot, C. (2020). A generic PBTK model implemented in the MCRA platform: Predictive performance and uses in risk assessment of chemicals. *Food and Chemical Toxicology*, *142*, 111440. <https://doi.org/10.1016/j.fct.2020.111440>
- Testai, E., Bechaux, C., Buratti, F. M., Darney, K., Di Consiglio, E., Kasteel, E. E. J., Kramer, N. I., Lautz, L. S., Santori, N., Skaperda, Z.-V., Kouretas, D., Turco, L., & Vichi, S. (2021). Modelling human variability in toxicokinetic and toxicodynamic processes using Bayesian meta-analysis, physiologically-based modelling and in vitro systems. *EFSA Supporting Publications*, *18*(4), 6504E. <https://doi.org/10.2903/sp.efsa.2021.EN-6504>
- Thompson, V. C., Firman, W. J., Goldsmith, R. M., Grulke, M. C., Tan, Y.-M., Paini, A., Penson, E. P., Sayre, R. R., Webb, S., & Madden, C. J. (2021). *A Systematic Review of Published Physiologically-based Kinetic Models and an Assessment of their Chemical Space Coverage*.
- U.S. Occupational Safety and Health Administration. (n.d.). *Chemical Exposure Health Data | Occupational Safety and Health Administration*. Retrieved June 9, 2023, from <https://www.osha.gov/opengov/health-samples>
- Viganò, E. L., Colombo, E., Raitano, G., Manganaro, A., Sommovigo, A., Dorne, J. L. C., & Benfenati, E. (2022). Virtual Extensive Read-Across: A New Open-Access Software for Chemical Read-Across and Its Application to the Carcinogenicity Assessment of Botanicals. *Molecules*, *27*(19), Article 19. <https://doi.org/10.3390/molecules27196605>
- Virshup, A. M., Contreras-García, J., Wipf, P., Yang, W., & Beratan, D. N. (2013). Stochastic Voyages into Uncharted Chemical Space Produce a Representative Library of All Possible Drug-Like Compounds. *Journal of the American Chemical Society*, *135*(19), 7296–7303. <https://doi.org/10.1021/ja401184g>
- Wambaugh, J. F., Bare, J. C., Carignan, C. C., Dionisio, K. L., Dodson, R. E., Jolliet, O., Liu, X., Meyer, D. E., Newton, S. R., Phillips, K. A., Price, P. S., Ring, C. L., Shin, H.-M., Sobus, J. R., Tal, T., Ulrich, E. M., Vallero, D. A., Wetmore, B. A., & Isaacs, K. K. (2019a). New approach methodologies for exposure science. *Current Opinion in Toxicology*, *15*, 76–92. <https://doi.org/10.1016/j.cotox.2019.07.001>
- Wambaugh, J. F., Wetmore, B. A., Ring, C. L., Nicolas, C. I., Pearce, R., Honda, G., Dinallo, R., Angus, D., Gilbert, J., Sierra, T., Badrinarayanan, A., Snodgrass, B., Brockman, A., Strock, C., Setzer, W., & Thomas, R. S. (2019b). Assessing Toxicokinetic Uncertainty and Variability in Risk Prioritization. *Toxicological Sciences : An Official Journal of the Society of Toxicology*, *172*(2), 235–251. <https://doi.org/10.1093/toxsci/kfz205>

D8.1: Report on the conceptual design of the SSbD toolbox	Dissemination level: PU
WP8: Concepts and toolboxes	Version: 1.0
Main Authors: Denis Sarigiannis	Page: 43

- Wang, Z., Walker, G. W., Muir, D. C. G., & Nagatani-Yoshida, K. (2020). Toward a Global Understanding of Chemical Pollution: A First Comprehensive Analysis of National and Regional Chemical Inventories. *Environmental Science & Technology*, 54(5), 2575–2584. <https://doi.org/10.1021/acs.est.9b06379>
- Wania, F., Breivik, K., Persson, N. J., & McLachlan, M. S. (2006). CoZMo-POP 2 – A fugacity-based dynamic multi-compartmental mass balance model of the fate of persistent organic pollutants. *Environmental Modelling & Software*, 21(6), 868–884. <https://doi.org/10.1016/j.envsoft.2005.04.003>
- Weiss, J., Fischer, S., & Eriksson, C.-H. (2022). *Identification of new and emerging risk chemicals and validation of the exposure index with a focus on humans* [Institutionen för Miljövetenskap (ACES) Stockholms universitet]. <https://www.imm.ki.se/Datavard/Rapporter/215-20-010%20Identification%20of%20new%20and%20emerging%20risk%20chemicals%20and%20validation%20of%20the%20exposure%20index%20with%20a%20focus%20on%20humans.pdf>
- Wiecek, W., Dorne, J.-L., Quignot, N., Bechaux, C., & Amzal, B. (2019). A generic Bayesian hierarchical model for the meta-analysis of human population variability in kinetics and its applications in chemical risk assessment. *Computational Toxicology*, 12, 100106. <https://doi.org/10.1016/j.comtox.2019.100106>
- Zheng, Z., Arp, H. P. H., Peters, G., & Andersson, P. L. (2021). Combining In Silico Tools with Multicriteria Analysis for Alternatives Assessment of Hazardous Chemicals: Accounting for the Transformation Products of decaBDE and Its Alternatives. *Environmental Science & Technology*, 55(2), 1088–1098. <https://doi.org/10.1021/acs.est.0c02593>
- Zheng, Z., Peters, G.M., Arp, HP., Andersson, P.L. Combining in Silico Tools with Multicriteria Analysis for Alternatives Assessment of Hazardous Chemicals: A Case Study of Decabromodiphenyl Ether Alternatives. *Environmental Science & Technology*. 53, 6341–6351, 2019. DOI: 10.1021/acs.est.8b07163