INTEGRA

Integrated External and Internal Exposure Modelling Platform

(INTEGRA)

B11 - Realistic estimation of exposure to substances from multiple sources
(CEFIC Long-range Research Initiative funded project)

D1.2 FUNCTIONAL SPECIFICATIONS OF THE INTEGRA PLATFORM

Due date of delivery: *30/06/2013*

Actual submission date: *30/11/2013*

Start date of the project: *01/01/2013*

Duration: *27 Months*

Lead contractor organisation name for this deliverable: *CERTH*

Dissemination Level		
PU	Public	
PP	Restricted to other programme participants	X
RE	Restricted to a group specified by the consortium	
CO	Confidential, only for members of the consortium	

Document Information

Document Name	D1.2 FUNCTIONAL SPECIFICATIONS OF THE INTEGRA PLATFORM
ID	-
Revision	Version 1
Revision Date	28/11//2013
Author	D.A. Sarigiannis / CERTH

Approvals

	Name	Organization	Date
Author	D.A. Sarigiannis	CERTH	30/06/2013
WP Leader	D.A. Sarigiannis	CERTH	30/06/2013
Coordinator	D.A. Sarigiannis	CERTH	30/11/2013

Documents history

Revision	Date	Modification	Author
Version 0	31/05//2013	1st draft	D.A. Sarigiannis
Version 1	30/11/2013	Final editing	D.A. Sarigiannis

TABLE OF CONTENT

A. Introduction/Scope

The scope of this document is to provide the basic concept for the development of the INTEGRA computational platform for "full-chain" exposure assessment.

The INTEGRA platform will be a web-based computational environment that will bring together the different tools available from the INTEGRA consortium addressing all the aspects of the full chain for aggregated exposure assessment in environmental and occupational settings.

The document describes the functional requirements and technical specifications of the different platform modules. For this reason a large number of stakeholders including academics, chemical and consumer goods industry, and regulators from both the EU and the USA were contacted in a a series of events organized by the INTEGRA team in the frame of international conferences that have taken place in the period August-October 2013. These comprised namely the International Society for Exposure Science conference organized in Basel (19-23 August 2013), the annual conference of the European Society of Toxicology organized in Interlaken (1-4 September 2013), the International Symposium on Human Biomonitoring organized in Manchester, UK (9-11 September 2013) and the 17th Symposium of the Mediterranean Scientific Association for Environmental Protection organized in Istanbul (28 September – 1 October 2013). The stakeholder consultation highlight was a dedicated workshop that took place in the frame of the ISBM in Manchester. The stakeholders were approached to provide an overview of the user requirements and the applicability of integrated exposure models such as the ones under development in INTEGRA.

The consensus of all stakeholders who participated in the INTEGRA workshop was that when it comes to exposure models it is advisable to start simple and gradually add complexity to the model structure to improve its performance. In this context, PBPK models are seen as the next best thing to extensive measurements in different biological fluids and tissues in humans. Sensitivity analysis would allow us to determine the most critical parameters for model performance. Model usability is also associated to the possibility to maintain live links with parameter databases that would allow model adaptation to scientific progress.

Mechanistic models allow us to approach the causal association between exposure and effect. Models need to capture as much biology or disease mechanisms as possible to be successful. There is a clear need for a metabonomic signature database that is widely accessible, since biomarker data for chronic disease are already available, but not centrally collected and managed.

Reconstruction of exposure will be the significant improvement that will come from toxicity data. If we know when concentrations cause toxicity then we can work back to reconstruct exposure. There is, however, a need to differentiate between "near" and "far field" exposure data. Near field may need more refined exposure data, where far field may be more in need of better toxicity data. The conclusion was, nonetheless, that in general we need both better and more toxicity data and better data appropriate for correct exposure characterisation.

A key problem is the misuse / misinterpretation / misrepresentation of model output by non-expert users. Thus, it is deemed that a user-friendly and transparent user interface is required for any such integrated exposure model to be widely and correctly used. Uncertainty needs to be represented in an understandable way - in this context, graphical representations of uncertainty may provide a viable solution.

The exposure models can be the same for both occupational and environmental or consumer exposure scenarios, however different parameterisation schemes are warranted in the different exposure settings. Normal populations are more sensitive than workers due to their enhanced diversity. It has to be noted that occupational exposure regulation often works with clear limit values - this is not the case with environmental exposure.

Overall, the stakeholders agreed that models do provide useful insights. Non-modellers erroneously take the modelled data with being aware of the model limitations, even though the modellers themselves understand that a model is only a representation of reality. For this reason, a "stop light" approach was proposed. If a chemical is in the "green" zone of a very simple tier 0 model, then there is no need for refinement. However, if it is in the "orange" or "red" zones, the we need to proceed to more complicated tier 1 and tier 2 models. In a nutshell, the stakeholder recommendation is to start simple and add complexity when needed, keeping in mind that the model output needs to be easy to interpret.

The scope of this document is not to provide the complete guidebook towards the development of the INTEGRA computational platform; rather, it seeks to foster discussion and exchange of ideas towards making final decisions among the project teams and the involved stakeholders. To this aim this concept document has been circulated to the INTEGRA partners to come to a final and stable version of it which will represent the official roadmap of the INTEGRA platform and finally to start software development.

B. Description of the main components of the platform

B.1   General concept

The INTEGRA platform will offer a number of generally applicable models ranging from multimedia environmental model to indoor air quality model and from exposure models for the different exposure routes (inhalation, oral and dermal) to a suite of three Physiology Based ToxicoKinetic (PBTK) models to evaluate internal doses and a database containing links to several types of data ranging from human physiological parameters to human biomonitoring data, from emission data from consumer products to indoor (occupational and residential) concentration levels and from outdoor pollution levels to building characteristics. Data will be stored as far as possible along with their geographical information in order to allow users to build realistic exposure scenarios to represent typical exposure conditions for specific countries and/or cities in Europe. Together, the database and models provide the tools to assess exposure for a wide range of scenarios covering a wide chemical space, whereby only limited additional information will be needed. Furthermore, the software also will accept stochastic distributions as input to a wide range of exposure parameters assessed via Markov Chain-Monte Carlo methods.

An exposure assessment in INTEGRA will follow two distinct tiers.

Tier 1 – Occupational exposure: Tier 1 will focus on the development of an external exposure model linked to a PBTK model to be applied in occupational settings. The main aim is for a number of scenarios to utilize the models, taking account of variability and uncertainty in model parameters, to estimate the levels that would be measured by biological monitoring and to then compare these data and attempt to reconcile any differences using sensitivity and other statistical analysis. Tier 1 will make use of the Advanced REACH Tool (ART) for inhalation, the Dermal Advanced REACH Tool (DART) for dermal exposure assessment or the GuLF DREAM model (van Wendel de Joode

et al, 2005) if DART will not be fully validated and available within the life of the project for dermal exposure assessment. Depending on their complexity and after a critical assessment of the operational feasibility the above models will be directly implemented within the platform or linked to the INTEGRA computational platform through hyperlink which will redirect the user to the web server where the models are actually installed.

Preliminary assessment suggests the ART model will be linked to the INTEGRA platform as external model while the GuLF DREAM model will be directly implemented into the INTEGRA platform.

Tier 2 – Environmental and consumer exposure: Tier 2 will focus on environmental exposure and will entail a multi-media and multi-route exposure models coupled with a suite of PBTK models ranging from a simple PK model suitable for data poor chemicals to a life-stage changing mother-fetus generic PBTK model.

In Tier 2 the modeling environment comprises five main modules, as follows:

1. **Multimedia environmental modeling** module to estimate the concentration of chemicals in different environmental matrixes (i.e. air, water, soil and food) taking into consideration the exchange between the different environmental media and different geographical scales.

2. **Indoor emissions-concentrations module**, linking sources to indoor concentrations, taking into account the physicochemical processes in indoor settings: dispersion, ventilation, gas-particle-dust partitioning, etc.

3. **Exposure module** including several models for the dermal, inhalation and oral routes, taking into account time-microenvironment-activity patterns and inhalation rates based on activity, gender and body weight.

4. **Internal dosimetry module**, which computes aggregate exposure linking temporal patterns to internal dose through a suite of PBTK models. It estimates the internal doses of contaminants and their metabolites at the target tissues. This module will contain automatic link to external Quantitative structure–activity relationship (QSAR) models to cover as much as possible the chemical space.

5. Quantitative linkage between exposure and internal dose allowing for assimilation of biomarker data will be implemented in the **reverse dosimetry module** for exposure reconstructions based on human biomonitoring data. To this aim the PBTK models will be applied using a computational inference engine for inverse modeling optimization and Bayesian hierarchical population modeling for extrapolation of single individual samples to population-relevant exposure estimates.

The wide variety of exposure determinants is associated with very diverse population groups both from the gender/age class, lifestyle habits, and from geographical point of view. This variety poses a problem for exposure assessors, who often do not have measured exposure data related to these determinants for all conceivable exposure scenarios and all population groups. To assist in the risk assessment associated to chemical exposure, INTEGRA can be used to estimate exposure for different scenarios for which there is paucity of measured data. This is reflected in the overall philosophy of the platform, which allows the user to choose which data is necessary to configure a simulation: automatic data retrieving from the INTEGRA database or manually inputted data when the user has newer or more relevant information. Furthermore, the approach followed permits the

user to use the platform as a tool to create hypothetical scenarios and investigate how the modified determinants affect the resulting exposure.

In the development of the computational platform special effort will be made to improve transparency, flexibility and ease of use of the software. According to data availability, the INTEGRA computational platform flexibly will allow the user to begin from different starting points along the source to dose continuum. For example, to estimate indoor air quality in residential settings the user can start the simulation either from the consumer products emission data, by utilizing a multizone Indoor Air Quality model implemented into the platform or directly from the indoor concentration data if the emission data are not available.

Moreover to improve the usability the platform and to facilitate the user in developing an exposure scenario the platform will propose default values for several parameters needed to configure the different models implemented in the platform. This however will not prejudice the user to edit and modify them if better and newer information will become available to the user.

Every user will access the platform with her/his own credentials and all her/his simulations results will be stored together with the associated input data in a dedicated folder of the CERTH server accessible to that user only. The user can access all his/her previous simulations to see them and modify the input parameters as needed to run a new simulation.

C. Platform structure

C.1 IT architecture

The INTEGRA platform will be based on the existing platform developed in the frame of the CEFIC-LRI INTERA (Asikainen et al. 2011) and TAGS (Sarigiannis et al. 2012) projects. It will be a multi-modular software especially designed to meet the end-user real needs and requirements for executing aggregate exposure assessment. More analytically it will enable the integration and assimilation of model and data needed to execute a "full-chain" aggregate exposure assessment focusing on both environmental and occupational exposure.

The IT architecture of INTEGRA computational platform system will follows the open architecture paradigm in order to allow future models to be readily used within the same framework on the basis of a standardized input / output protocol. In this way, the tool will easily adapt to scientific / technical progress.

The INTEGRA computational platform will be a web based 3-tier application working on the most common used web browsers The components that constitute the platform are the Graphical User Interface (GUI), the computational module and the Database (Db). Furthermore, closely related to the platform are the external databases (e.g. Popgen and Symcyp) from where to retrieve data not stored in the INTEGRA Database but necessary to configure a simulation and the ART, DART or GuLF Dream external models to be used in Tier 1 simulations. The proposed IT web-based architecture is graphically depicted in Figure C1 and in Figure C2.
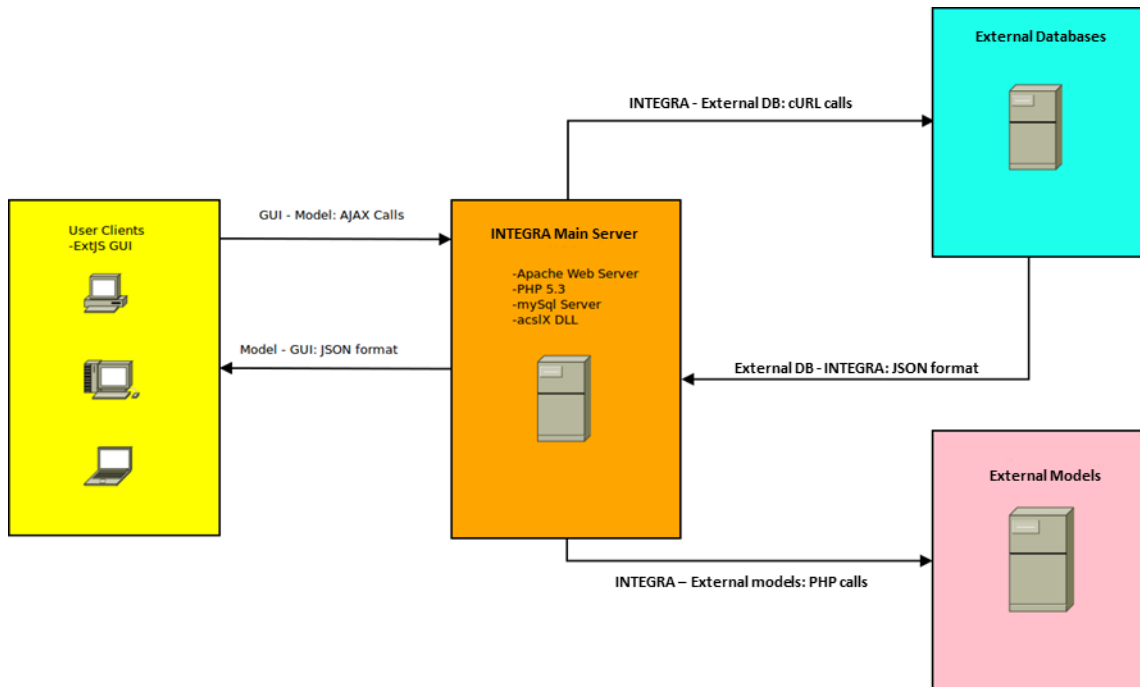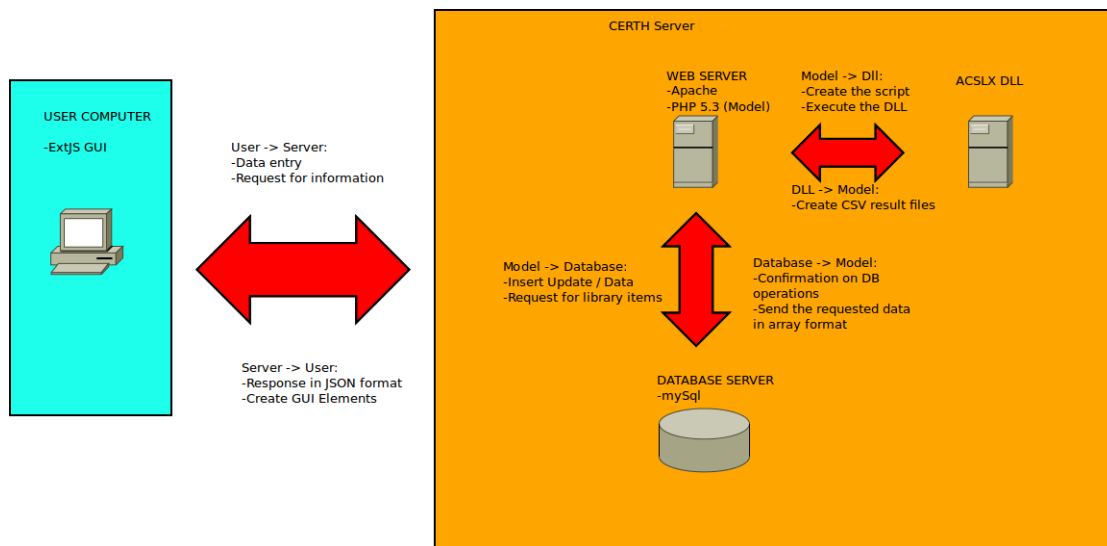
Figure C1: overall INTEGRA Platform architecture

Figure C2: INTEGRA Server-side architecture

a) Graphical User Interface (GUI)

The GUI will be a client side application, executed inside the web browser's window. The GUI is the means through which end users interact with the model allowing the user to configure and run a simulation easily and in an intuitive way. Using appropriate GUI elements (such as text boxes, drop down lists and grid controls) the users will configure the simulation, view the results in the form of charts, download the results in an Excel format for further use and create and download simulation reports in PDF format.

The GUI will be coded in Javascript (using the ExtJS library) and the interaction with the Model is implemented through AJAX calls (Asynchronous Javascript And XML). Every server-side script related with the creation of the Javascript GUI is coded in PHP-Hypertext Preprocessor. Furthermore, the GUI will utilize the Highcharts Javascript library for the creation of the charts.

b) Computational module

This will be the core module of the application. It will be hosted on CERTH's server. The model consists of two "parts"; PHP and acslX© (http://www.acslx.com/).

Whilst the acslX part will be the "heart" of the platform, the PHP part is the "glue" that joins the user's requests, the database queries, the remote queries (e.g. to external Databases) and the process of the scripts creation into one "action" transparent to the end users. Not only that, the PHP part of the model is responsible for the post-processing of the simulation results.

INTEGRA computational platform will be a web application, so a web server is required. The INTEGRA server side scripts will be executed within the Apache Web Server environment. The PHP part will be written in PHP following the Object Oriented paradigm. It is designed in a flexible and extensible way, so new functionality can be added with the minimum effort.

The PHP part can conceptually be separated in layers, each one of them plays a different role in the application:

- The *Authentication Layer* will handle the User Registration and the User login procedure.

- The *Presentation Layer* will be tightly coupled with the GUI. It will receive the user data and it will create the appropriate Javascript elements for presenting the data back to the users.

- The *Data Access Layer* will handle all the database related functions. It will utilize the AdoDB PHP library, thus being "database engine" agnostic. Furthermore, the Data Access Layer will contain Classes for reading Excel files (utilizing the PHPExcel Library), XML files (utilizing the SimpleXML library) and executing remote queries using the cURL library and parsing the results given in JSON (JavaScript Object Notation) format (www.json.org).

- The *Reporting Layer* will utilize the tcPdf library for the creation of pdf reports.

- The *SimulationModel Layer* will be a set of interrelated PHP classes representing the model's structure and data.

- The *Execution Layer* will use the Simulation Model in order to create and execute the acslX scripts.

c) Database

The database server will be the latest version (version 5.5.10) of mySql Server. Its role will be to provide an effective persistence mechanism for all the application's storage needs. The model will communicate with the Database using the AdoDB library, helping these two tiers to be independent.

The Database tier will be involved in three major aspects of the application. It will store the user's accounts and checks for authenticated users. Moreover it will store and manage all the types of the underlying information (environmental data, modelling results, human biomonitoring data, consumer products data, use, patterns, toxicological data, population data etc.), necessary for the configuration of the simulation. The Database will also include links to external data sources that may not necessarily have been incorporated into the INTEGRA Database (e.g. from other relevant CEFIC-LRI projects, or the JRC EXPOFACTS database). Finally, the Database will provide a storage place for the input parameters and the output results of a model execution.

The end-user computer has only to connect to the INTEGRA Server in order to run the platform, other specific software requirements are not requested since the end-user computer will not perform any calculations.

The "hearth" of the platform software will be built in software environment based on acslX. The core platform coding will be implemented in acslX version 3.0 developed by Aegis Technologies. AcslX is a dynamic modeling and simulation software environment that includes an optimization package that contains a robust parameter estimation, min/max optimization, global sensitivity analysis, and Monte Carlo analysis for performing uncertainty assessment. The software runs under MS-Windows Operating System and it includes Application Programming Interface (API) to interface acslX executable with other applications. The development language is native but Fortran and C can also easily supported.

D. Computational implementation of the full chain assessment

The scope of this section is to provide the basic concept for the development of the INTEGRA computational platform for aggregate exposure assessment according to the "full-chain" approach.

In our vision the INTEGRA platform has to bring together the different tools available inside the INTEGRA consortium as well as outside it, both addressing all the aspects of the full chain for aggregate exposure assessment rather than develop a single computational model.

A large list of different models has been identified by the INTEGRA consortium each one addressing a specific step along the full-chain: they range from indoor air quality models to multimedia model and from exposure models to PBTK models. The objective is to integrate all of them in a suitable way into the INTEGRA platform.

From the point of view of software development the platform will be built so as to guide the generic user through intelligent user friendly graphical interfaces to select and execute the appropriate model among the ones implemented in the INTEGRA computational platform.

A flexible approach consisting of three different options will be used with regard to the data needed to run each specific model. For those databases which are available to the project partners or that can be freely downloadable from the web they will be directly implemented in the INTEGRA Database. A second approach will regard those databases which are not freely downloadable but they are freely accessible through the web: in this case ad-hoc query scripts will be developed and implemented with the objective of interrogating these databases and extracting the data of interest in a format usable by the platform. A final option will entail the manual insertion of specific data from literature and/or ad-hoc studies. Tailored user-friendly GUIs will be implemented to allow the user to insert and store the desired data into the INTEGRA Database.

D.1. Modules

The INTEGRA platform will be implemented focusing on single chemical exposures through multiple exposure routes: inhalation, dermal and ingestion covering a chemical space as wide as possible. The INTEGRA overall methodological approach is illustrated in Figure D1
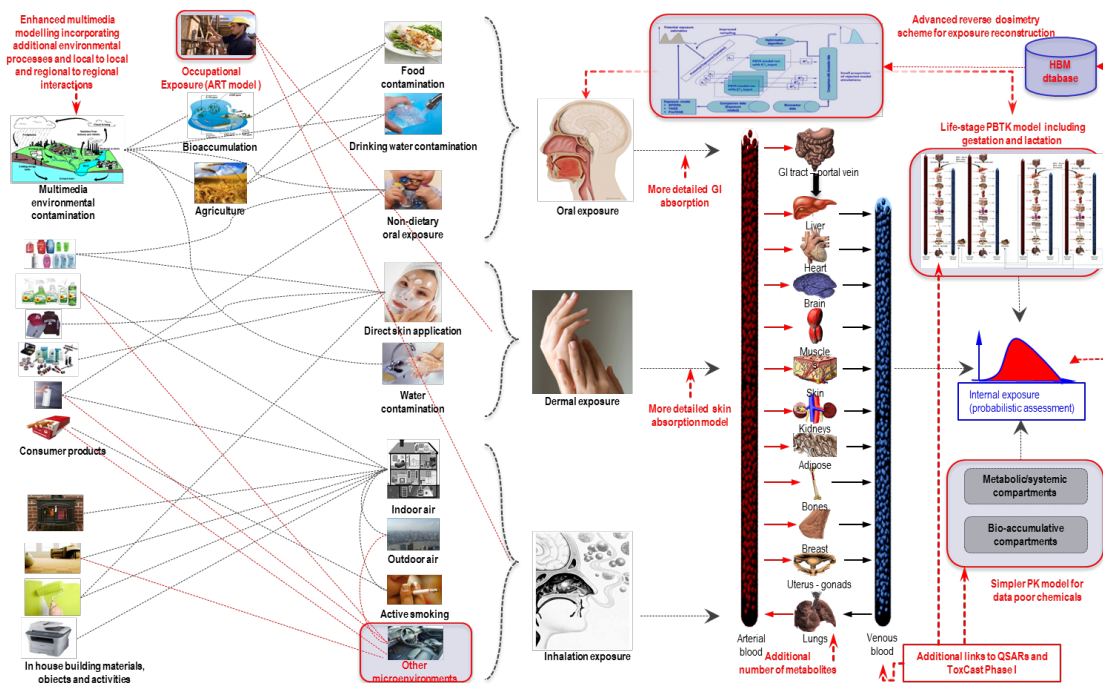

Figure D1: INTEGRA methodological approach

The platform architecture will encompass three horizontal modules and six vertical modules (Figure D2):

HORIZONTAL MODULES

1. INTEGRA Knowledge database
2. Uncertainty module
3. Report and visualization module

VERTICAL MODULES

4. Model run settings
5. Multimedia model module including a dietary contamination sub-module
6. Microenvironment module
7. Exposure scenario definition module
8. Internal dosimetry module
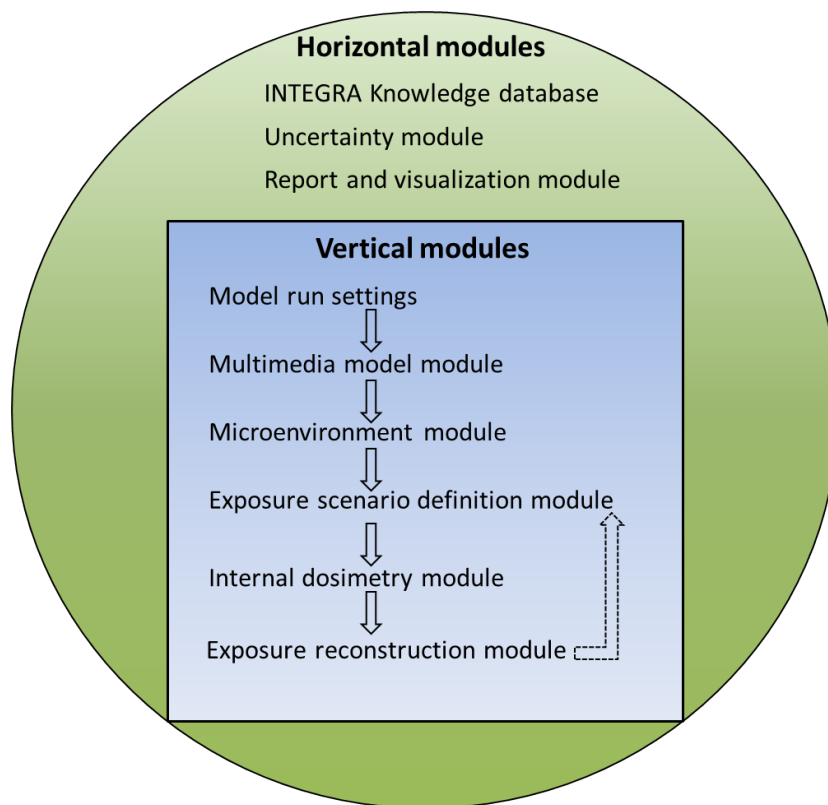9. Exposure reconstruction module

Figure D2: Graphical representation of the INTEGRA platform modules

Each vertical module will be designed to address a specific part of the full-chain approach and it will "talk" to the others through the Database that represents the core of the system.

The first horizontal module, the INTEGRA knowledge database, will be specifically designed to manage the Database which stores and manages all the types of the underlying information.

The second module will address the quantification and propagation of uncertainty along each step of the full-chain. For this the AcslX environment includes a sophisticated optimization package that contains besides robust parameter estimation and global sensitivity analysis, a Markov Chain - Monte Carlo tool specifically designed to address uncertainty analysis.

The main functionality of the third horizontal module will be the presentation of the results obtained through the INTEGRA platform. These will be displayed and exported both as graphs and reports.

D.1.1.   *Horizontal modules*

D.1.1.1. INTEGRA Knowledge database

The INTEGRA Database will enable the effective and user-friendly management, storing and updating of the data required for executing a full-chain integrated exposure assessment.

This module will act as the *"main input gate"* of the INTEGRA platform of all the available datasets to the end-users. A further important role of this module is that it will be the main information server for all the INTEGRA application modules. Having a predefined database schema the application modules can directly access these datasets without having to perform any format conversion either manually or automatically.

The INTEGRA Database will have a publicly available and a user-specific section. Although the generic user can see and access both the sections he/she can add, modify, delete data only in her/his user section. To this aim each entry will be flagged with the name of the user who added this entry and it will be visible only to him/her. The database Administrator will be responsible for the quality assurance of the data entered and accordingly to decide if to flag it as a public data. In this case all the users can see and use the entry.

This module offers the following functionalities:

a) *Management*: enables the Database Administrator to query, import, update, delete data from the public section of the INTEGRA Database. The same functionalities will be granted to the generic user with regard to the user section of the INTEGRA Database. The stored data can be classified according to:

- Emission data (i.e. releases from consumer products)
- Consumer products list (classified according chemical composition)
- Use patterns of consumer products
- Exposure data (Intake fraction, uptake factors)
- Physical-chemical properties of the substances
- Toxicological properties of the substances
- PBTK Models parameters (physiological, biochemical and metabolic)
- Human biomonitoring data
- Environmental contamination data both for occupational and residential settings
- Modelling results
- Supporting data (such as population and population data and time activity patterns).

Data on physiological parameters as well as on environmental contamination and supporting data will be stored along with their geographical information in order to allow users to build realistic exposure scenarios which can represent typical exposure conditions for specific countries and/or cities in Europe.

b) *Compact*: Enables to compact and reindex the INTEGRA database to minimize the required disk storage and ease of access to the complex structure of the database. This process is taking place automatically after importing a new dataset into the database.

c.) *Remote querying:* It was mentioned that some already existing databases (e.g. Popgen) coupled to specific modelling tools will not be integrally duplicated into the INTEGRA internal Database. As a consequence of that, particular attention will be paid to the development of suitable query scripts having the objective of interrogate these databases and extract from them the data of interest in the proper format usable by the platform and store them into the INTEGRA database for further use. A set of standard data format (such as ASCII, CSV, XLS) will be identified which will render connecting these databases with the INTEGRA platform easy.

D.1.1.2. Uncertainty module

Within the INTEGRA computational platform an advanced stochastic tool will be implemented based on the use of Monte Carlo - Markov Chain technique (MCMC). This technique allows the

user to execute a simulation using Probability Distribution Functions for the main input parameters rather than using fixed single value for them and consequently to obtain distribution functions also for the results. Samples of parameter values are randomly drawn from the specified distribution functions. For these sets, exposure is calculated and a probability distribution of the exposure is constructed.

According to the option selected in the "model run settings" (see paragraph D.1.2.1) probabilistic calculation will be taken into account or not. AsclX development environment already includes an optimization package that contains a robust parameter estimation, min/max optimization, global sensitivity analysis, and Markov Chain Monte Carlo (MCMC) analysis for performing uncertainty assessment. Monte Carlo (MC) methods are the most widely used means for uncertainty analysis, with a wide variety of applications. These methods involve random sampling from the distribution of inputs and successive model runs until a statistically significant distribution of outputs is obtained which represents a refined approximation of the estimated exposure for the population (Figure D3).

The user will be prompted to select a distribution by clicking on a pull-down menu and make a selection from the list that is displayed. Then he/she will enter the values for each parameter characterizing the distribution (min. max, mean and standard deviation). To support the user, if one or more values are incorrect or incompatible with the Probability Distribution Function selected, the background of the container box will turn grey and the user will not be permitted to go to the next screen.

After having selected a Probability Distribution Function the system will sample a value from the corresponding distribution for that input, set the inputs to the sampled values, run the simulation and finally collect the desired outputs. The process will be repeated, collecting appropriate statistical information on the outputs until the desired accuracy has been reached.

In INTEGRA the user can determine the number of MC samples that is drawn. It should be noted that a higher number of samples in general leads to a more representative distribution of the exposure, but such at the cost of more calculation time which, due to the high number of differential equations to be simultaneously solved by numerical integration, can be very high for a high number of iterations resulting in substantial computational demands.

For that reason a global sensitivity analysis will be part of the uncertainty module allowing the user to identify which input parameters will most affect the final outcomes and so facilitating the selection of the parameters to be used as input to the MCMC engine.
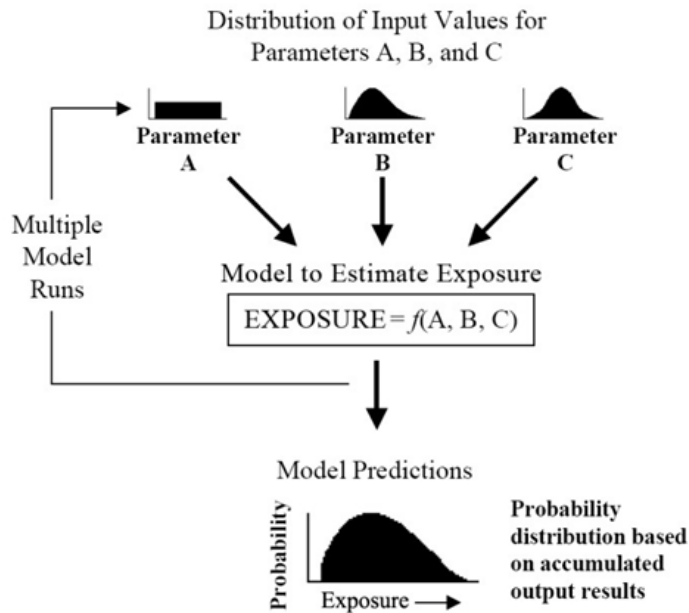
Distribution of Input Values for
Parameters A, B, and C

Parameter A Parameter B Parameter C

Multiple Model Runs

Model to Estimate Exposure

$$EXPOSURE = f(A, B, C)$$

Model Predictions

Probability

Exposure →

Probability distribution based on accumulated output results

Figure D3: Diagrammatic representation of the application of Monte Carlo analysis to a model.

D.1.1.3. Reporting and visualization module

The results of an INTEGRA platform simulation will be presented in different ways: as a chart, as a distributed result from a MC simulation, and in a textual report.

Outputs

The outputs obtained through the execution of the INTEGRA platform will be displayed in the visualization module after their storing in to the INTEGRA Database.

A first list of outcomes variables will encompass the following:

- **External inhalation exposure** is calculated as the air concentration during exposure. Results will be shown as time dependent air concentration through a determined period of time.

- **Internal inhalation doses** expressed as daily average of the amount taken up per kg bodyweight.

- **External dermal exposure** is calculated as the amount that can potentially be taken up per kg bodyweight).

- **Internal dermal doses** expressed as daily average of the amount taken up per kg bodyweight.

- **External oral exposure** is calculated as the amount that can potentially be taken up per kg bodyweight.

- **Internal oral doses** expressed as daily average of the amount taken up per kg bodyweight.

- **Total aggregated external exposure** is calculated as the summations of the corresponding external exposure per route.

- **Total aggregated internal doses** is calculated as the summations of the corresponding doses per route

- **Biologically effective doses (BED)** in the target tissues resulting from the application of PBTK models. These results could be either estimated as total BED (aggregated over the different exposure routes) or disaggregated according to the exposure routes. Different measures are: time course through a determined period of time and Area Under the Curve (AUC), to have the integration of internal dose through time.

- **Concentrations of metabolite** through excretion (e.g. urine, breath) and in venous blood, allowing the interpretation to biomonitoring data. Different measures are: time course through a determined period of time and AUC, to have the integration over time.

It has to be mentioned that the platform can generate a huge number of output variables but the user may not be interested in exporting all of them into the Database. Furthermore, as the INTEGRA computational platform will have the unique characteristic of tracking dynamically in the time the exposure using a small integration step to account for very short exposure events, the output files may reach huge sizes. For these reasons and to speed up the process of storing the results in the Database the INTEGRA Visualization module will present automatically a list of all the output variables generated after a simulation so that the user can select the output variables he/she would like to store in the Database and to further display as graphs.

Graphs

All the above can be examined via graphs of the time course of the exposure. The time axis will automatically be set to the exposure duration specified by the user in the "model run settings" (see paragraph D1.2.1) module, but can be adjusted manually by using the zoom function.

Similarly, the y-axis will automatically be fitted to include the whole exposure range encountered during the event, but may be adjusted manually. The data points of the resulting graph can be exported as standard Microsoft Excel file format. The graph itself can be saved as a print screen shot in .png or .jpg format.

Distribution

When a probabilistic simulation has been selected in the "model run settings" module (see paragraph D1.2.1), the INTEGRA platform will execute a probabilistic calculation. The program will draw a set of random numbers from the specified distributions for distributed parameters and calculates the endpoint of choice with this set. This procedure is repeated for a user-specified number of times (the number of MC samples). The result of this procedure will be a distributed set of outcomes. Some characteristics of the distribution such as median, 5th and 95th percentile of the calculated distribution will be reported. The calculated distributions will be plotted in a graph, and can be exported to a Microsoft Excel file.

Report

The INTEGRA report will give a complete (textual) overview of the simulation calculation. It will display all the model run settings, all the parameter values as well as the outputs obtained. The report can be downloaded and saved as PDF file directly from the INTEGRA platform.

D.1.2.   Vertical modules

D.1.2.1. Model run settings

This module will serve to define some general settings for each platform run. These should be defined before to start any run as they will define the overall configuration of a specific platform simulation.

In this module the user will be able to:

- **Create a new simulation** entering a new name

- **Define the tier**: choosing between Tier 1 (occupational) and Tier 2 (environmental)

- **Edit/Copy a selected simulation to a new simulation** to restore the contents of a previous run into a new one and allow it to be edited.

- **Delete** an existing run.

- **Select the chemical** for which a new simulation will be executed.

- **Specify the simulation length** allowing the user to specify the duration (in hours) of the simulation which will be executed.

- **Define the population targeted** by the simulation. Through this function the user will be able to define the genders and ages of interest in the simulation (all ages, or some subset of those ages will be available for selection). The INTEGRA platform will use age groups for several purposes. Indeed the model will take from the INTEGRA Database age-specific physiological parameters (e.g. bodyweight, inhalation rates, body surface etc.) as well as time activity pattern according to the age classes selected. It is worth mentioning that the body weight (Opasnet 2011) will represent an important parameter within the INTEGRA computational platform as many other parameters necessary to run the PBTK model will be automatically scaled on its value such as the organ volumes, the cardiac output and the blood flows. Furthermore, on the basis of the combination of gender and age selected the platform will retrieve age-dependent model inputs such as dust and soil ingestion daily amount.

- **Specify the geographic location** of the population selected. Through this option the platform will query the INTEGRA Database to retrieve the environmental contamination data and time activity pattern of specific climatic regions. The platform will display the choices among different European Countries as well as among some pre-defined climatic regions according to the prevalent meteorological conditions.

- **Probabilistic run or not**: this option will entail the possibility to run a simulation defining probability distribution functions (PDF Probability Distribution Functions) for the main input parameters used by each model along the full-chain instead of using fixed values for them. The "no" option will represent the default option. Some basic distributions will be provided through the platform such as: uniform, normal, log-normal.

It is worth mentioning that according to the Tier selected the user will be redirected to different user interfaces. Specifically, if the user will select to run a Tier 1 simulation (i.e. for occupational settings) he/she will be redirected to the web server where ART and eventually DART models are installed. A common shared authentication procedure will be implemented so as to allow the user to access these external models in a transparent way. Furthermore, a common standard

Input/Output protocol will be developed to allow a seamless integration of Input/Output with the INTEGRA platform.

The user will configure and run these models through their own user interface obtaining thus the main output exposure estimates in occupational settings (additionally the GuLF DREAM model developed by IOM will be integrated in the INTEGRA platform as an alternative solution to specific occupational exposure settings). The latter will be automatically imported into the INTEGRA computational platform and will represent the input of the Internal Dosimetry module (see D.1.2.5). Therefore the next three modules illustrated hereinafter, which will provide exposure estimates in environmental settings, are intended to be used for Tier 2 simulations only.

D.1.2.2. Multimedia model module

This module will include a multimedia model for the estimation of environmental media concentrations in different environmental matrixes (air, soil, water) relevant to the compound of interest. All the stages of the compound life cycle, including production, industrial, consumer use and disposal will be considered. The model will be based on the Simple-box model (Klepper and Den Hollander 1999) already implemented in the TAGS platform (DA. Sarigiannis et al. 2012), which will be enhanced incorporating interactions between similar smaller spatial scales (e.g. local to local, regional to regional), new environmental pathways (e.g. groundwater), and improved accuracy in the estimation of parameters affecting environmental fate.

In addition, a dietary contamination sub-module will be part of this module. It will consist of two major elements, namely food chain transfer which will be coupled to the multimedia model and migration from food contact materials. The latter will entail the development of a food contact material migration model that takes into account the physicochemical properties of the compound of interest as well as the conditions of use (e.g. temperature, time in contact to the material). Work done at the JRC Food Contact Materials EU Reference Laboratory (Simoneau et al. 2012) will be the basis for this model.

D.1.2.3. Microenvironment module

This module will estimate chemical environmental concentration for indoor locations. The model will start from data on releases from consumer products and building materials stored in the INTEGRA Database and will implement a multi-zone indoor air quality model to estimate the indoor chemical concentration.

The IAQ model will incorporate interactions within major media components (including gaseous, particles and settled dust phase), including dynamic equilibrium of emissions through consumer products (volatilization, direct abrasion) and a detailed description of settled dust phase mass balance.

There will be two main options to generate concentration data for scenario-relevant media.

Transfer functions/Dispersion Model: with this option the user chose to apply a multizone Indoor Air Quality model which will be directly implemented in the platform core (figure D4).
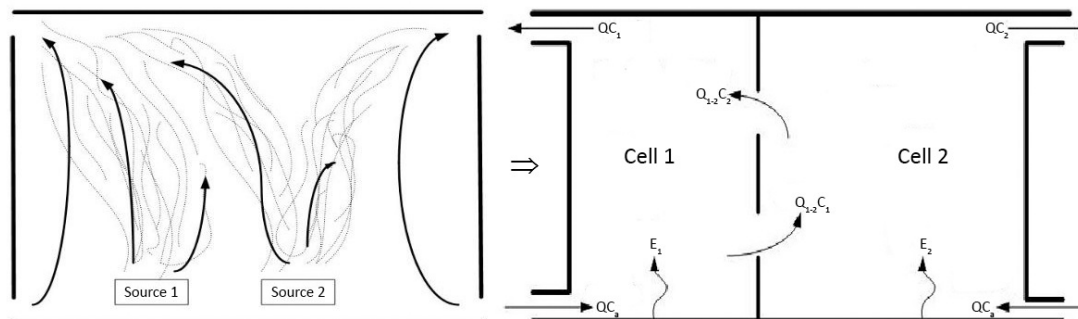
Figure D4: A partially mixed enclosure example and its numerical transformation

User-specified concentration values: The user will supply directly the indoor air concentration levels. This functionality is thought to enhance the flexibility of the platform as it is recognized that the user may not have enough data on emission from consumer products to properly create an emission scenario. In this case the user can decide to start directly from the concentration data which are in general easier to find.

Independently from the option selected the output of this module will be the indoor air concentration levels for residential settings for the selected chemical and will represent one of the inputs to the subsequent modules for exposure assessment.

D.1.2.4. Exposure scenario definition module

The main aim of this module is to build an exposure scenario for the population group(s) selected in the "module run settings" module aggregating all the possible routes of exposure for a selected chemical.

There are various ways in which one can be exposed to a compound. For example, a volatile compound may be emitted into air which subsequently is inhaled. A product may be applied to the hands resulting in dermal exposure to a compound in the product. By hand to mouth contact some of the compound can be transferred to the mouth resulting in oral exposure. Consequently, a compound may be taken up in the body via these different routes (inhalation, dermal and oral), resulting in an internal systemic dose.

For sake of clarity, it is convenient and easier to consider separately the different exposure routes.

Exposure by inhalation will be quantitatively estimated starting from the output of the microenvironment module and of the multimedia model with regard to outdoor air. The exposure resulting from inhalation is the sum of the fraction of time spent in each environment (outdoor, residential indoor, other indoor locations) multiplied by the concentration of the selected chemical in each of the above environment.

Others input data which will be gathered from the INTEGRA database accordingly with the model run settings (i.e. gender, age, geographical position) will include:

- Inhalation rate (according with the type of physical activity carried out)
- Time activity pattern

Based on the INTERA and TAGS database, the INTEGRA platform will list a series of different activity types (Hanninen et al. 2002; Jantunen 1999; Koistinen et al. 2001) to determine the appropriate inhalation and consequently intake rate (ICPR 2002).

The second and the third routes of exposure will be directly calculated inside this module.

For dermal exposure the loading mechanisms considered will encompasses:

- Instant application, which simulates the direct application onto the skin of a consumer product during a short period. Examples include body creams, baby lotions, hair gel, skin cleaners etc.

- Migration from other material in direct contact with the skin which describes the exposure due to migration of a chemical to the skin when dermal contact with the material happen. Examples include clothes, gloves, toys etc.

- Rubbing-off, which describes a secondary exposure situation. Indeed it describes a situation in which a surface (e.g. table top, floor) is treated with a product or when a chemical deposits onto it (in particle phase) and dermal exposure arises from contact with the that surface.

For each of the above types different data will be needed (i.e. product amount, weight fraction compound, contact time, transfer coefficient, exposed area, etc.) which will be gathered from the INTEGRA database or manually inserted by the user if he/she has more detailed data.

For oral exposure the loading mechanisms considered will encompass at least the followings:

- Soil ingestion

- Dust ingestion

- Food ingestion

- Migration from objects brought to the mouth

- Inadvertent Ingestion of Personal Care Products

Data needed to properly configure the above loading mechanisms, such as the amount of chemical ingested together with the weight fraction of the chemical, the ingestion rate, the contact area and the migration rate will be queried from the INTEGRA database or manually inserted by the user if he/she has more detailed data.

Applying to each of the above exposure routes an uptake factor (for inhalation) or uptake fraction (for oral and dermal) an exposure scenario for a selected chemical will be built taking into consideration all the possible routes of exposure.

It is important to remind that, according with the choice of probabilistic simulation defined in the "model run settings" module all the above parameters can be defined together with their probability distribution function. To this aim the GUI will display a button "MCM" next to each parameter: clicking this button the user will be able to select and configure the Probability Distribution Function desired.

The main output will be the actual exposure level of the population (or selected sub-groups) for the contaminant considered together with its uncertainty provided in terms of median, 5th and 95th percentile.

D.1.2.5. Internal dosimetry module

The internal dosimetry module will be a key component of the INTEGRA platform aiming at estimating the internal doses of a chemical and its metabolites in the different tissues composing the human body on the basis of the exposure scenario obtained by the "exposure scenario definition" module. It will be implemented in the in acslX version 3.0 developed by Aegis Technologies Group software. AcslX provides a flexible PBTK modelling tool that predicts drug dispersion and residual drug levels for internal dose analysis. AcslX supports both classical and compartmental models so that there is no limit on model simplicity or complexity.

Despite the complexity imposed by the internal dose module, the estimation of internal dose instead of external exposure facilitates the need for refined exposure assessment as for many chemicals, toxicity arises not from the parent compound but from the metabolites justifying also the assumption that internal dose is the correct exposure metric to be considered assessing the health risk for the population.

The whole concept will be based on a suite of three different PBTK models according with their level of detail and complexity as follows:

- **A simpler PK** model for data poor chemicals. This model will include a limited number of compartments allowing simple PK mass balance considerations and the assessment of bioaccumulation potential, capturing both the effect of lipophilicity and metabolism on chemical bioaccumulation.
- **A generic PBTK** model which incorporates multi-route exposure, detailed compartmental description, plasma protein/red cell binding and estimation of concentration of parent compound and metabolites in several biological fluids including breast milk and the contribution of lactation to breast feeding exposure. This model is already implemented in the TAGS/INTERA platforms.
- **A life-stage changing mother-fetus generic PBTK model** (i.e. an enhanced version of the generic PBTK model described above) incorporating in utero exposure, describing internal exposure from conception onwards, including transfer through gestation and lactation.

On the basis of the data availability and the aim of the assessment the graphical user interface will enable the user to choose the PBTK model to be used in his/her simulation.

Every PBTK model requires several parameters that are critical determinants of chemical uptake and disposition. These determinants can be classified into three main categories, namely, anatomical/physiological, physicochemical, and biochemical.

Validation of the PBTK model with data sets other than the working set (or training set) is a fundamental and critical phase of the model development: extensive data sets are used on the PBTK model, ensuring robustness when investigating its predictive capability. Once a PBTK model is validated for one chemical the vales of the physicochemical and biochemical parameters are determined and the PBTK model is ready to be used for accurate estimations of the time course of a chemical and its metabolites in the human tissues.

For chemicals with a PBTK model already validated the value of physicochemical and biochemical parameters will be stored in the INTEGRA Database unequivocally linked to the relevant chemical. When the user will select that chemical in the "model run settings" module the platform will query

the INTEGRA Database and will display in an *ad-hoc* user interface the values of all the relevant physicochemical and biochemical parameters.

Although it would not recommended to modify these values as they are derived through the model validation process, the user can change them simply entering the new value in the corresponding edit box. A column with the default values (i.e. the values obtained from the validation process) will be displayed close to the editable one to assist the user during this phase.

To enhance the flexibility of the platform and taking advantage of the generic nature of the PBTK models implemented, for new chemicals added by the user or chemical not having a pre-validated PBTK model all the PBTK parameters will be set to a default dummy value: the user can modify them entering the desired values so that the new values will be stored in the INTEGRA Database and used in the simulation.

This module will include also specific links to QSAR models in order to derive the necessary parameterization of PBTK models for new chemicals or chemicals with limited information. Furthermore links with PopGen (http://xnet.hsl.gov.uk/popgen/service.aspx) and Simcyp (http://www.simcyp.com/ProductServices/FreeADMETools/) will allow the user to retrieve information on realistic anatomical and physiological parameters in human populations.

D.1.2.6. Exposure reconstruction module

The key objective of this module is to reconstruct exposure scenarios consistent with human biomonitoring data allowing for a better exploitation of biomonitoring data, in terms of estimating overall exposure, route and pathway attribution, or reconstruction of a complete exposure scenario.

Assimilation of human biomonitoring data and their translation into intake distribution amounts to a computational inversion problem, where the objective is to identify the specific input distributions that best explain the observed outputs while minimizing an appropriate metric of error. In our case the inputs involve spatial and temporal information on environmental media concentrations of chemicals and corresponding information on human activities that result in intakes; the outputs are the observed biomarkers.

In the INTEGRA computational platform a tiered approach will be followed based on the data availability (periodicity and size of sampling, specimen type) and the requirements of the exposure reconstruction analysis (temporal analysis of exposure, contribution from different routes): different methods will be explored starting from simple exposure conversion factors (ECF), up to Maximum Likelihood Estimates–PBTK modelling with synthetic biomarker data and MCMC analysis (Georgopoulos et al. 2009)..

Accordingly the user interface will display different options among which the user can select the most appropriate one. The reverse dosimetry module will be implemented in the core of the INTEGRA platform (D. Sarigiannis et al. 2012) through ad-hoc m-language scripts implemented in ACSLx which will use the human biomonitoring data together with the *a-priori* knowledge of the exposure determinants, physiological, metabolic and biochemical parameters for individuals as input data and will take advantage from the PBTK suite of models and the MCMC engine implemented in the platform to infer plausible exposure scenarios consistent with the human biomonitoring data which represents the final output of this module

E. Conclusion

The overall purpose of this document is to provide the basic concepts for the development of the INTEGRA computational platform for source to dose aggregated exposure assessment in occupational and environmental settings covering a wide chemical space..

The document describes the functional specifications of a proof-of-concept computational toolbox, the core of which will be a synthetic dynamic modelling environment able to track and describe in mathematical terms all the steps of the source to dose approach, implementing both mechanistic (e.g. multimedia fate modelling, Physiology Based ToxicoKinetic Models) and probabilistic methodologies (Markov Chain Monte Carlo) for human biomonitoring data assimilation.

The INTEGRA platform will bring together a large list of different models and databases each one addressing a specific step along the source to dose approach integrating them into the INTEGRA platform either implementing them directly into the computational platform or accessing them through suitable hyperlinks in a transparent way for the user.

All the above characteristics will be implemented in a multi-modular publicly available web based computational platform providing clearly a series of advantages compared to the currently existing exposure assessment modelling software. These advantages mainly lying on:

- the flexibility - adaptation of the INTEGRA platform that allows the user to tackle contaminants with completely different exposure pathways and exposure route by different parameterization and individual model selection;
- the fusion of mechanistic and probabilistic approaches in order to minimize the total boundaries of uncertainty;
- the development of an exposure biology based approach, taking into account inter-individual susceptibility and the different biological response to the same exposure levels;
- occupational, environmental, consumer exposure and metabolic processes are tackled simultaneously and continuously, describing in a realistic way the interaction among human body and the continuously changing surrounding environment.
- the accessibility and the user-friendliness of the platform which facilitate and support non-expert users in the exposure scenario development process.

Its implementation in a user-friendly web-based computational environment allows any user to easily build exposure scenarios and run them to estimate both external and internal exposure in both occupational and environmental settings for the whole population as well as for some sensitive groups and to assist risk assessors and public authorities in the risk assessment associated to chemical exposure.

In addition the approach followed permits the user to use the INTEGRA platform as a tool to create hypothetical scenarios and investigate how the critical determinants affect the resulting exposure. This could be useful to simulate impact of risk management measures, thus rendering INTEGRA an easily adaptable tool for use in the context of several regulatory frameworks, including but not limited to REACH.

F. References

Asikainen A, DeBrouwere K, Doust E, Galea K, Gotti A, Happonen E, Sanchez Jimenez A, Karabelas A, Karakitsios S, Kontoroupis P, Kuipers E, Mplatsis E, Nikolaki S, Sarigiannis D, Semple S, Standaert A, Torfs R, Jantunen M. 2011. INTERA Final report Kuopio, Finland:National Institute for Health and Welfare (THL).

Fransman W, Van TM, Cherrie JW, et al. (2011) Advanced reach tool (ART): development of the mechanistic model. Ann Occup Hyg; 55: 957–79.

Fransman W, Van Tongeren M, Cherrie JW, Tischer M, Schneider T, Schinkel J, Kromhout H, Warren N, Goede H, Tielemans E. 2011. Advanced reach tool (ART): Development of the mechanistic model. Annals of Occupational Hygiene 55:957-979.

Georgopoulos PG, Sasso AF, Isukapalli SS, Lioy PJ, Vallero DA, Okino M, Reiter L. 2009. Reconstructing population exposures to environmental chemicals from biomarkers: Challenges and opportunities. Journal of Exposure Science and Environmental Epidemiology 19:149-171.

Hanninen OO, Alm S, Kaarakainen E, Jantunen M. 2002. The EXPOLIS Databases, B13/2002. Publications of the National Public Health Institute, Kuopio, KTL.

ICPR. 2002. Basic anatomical and physiological data for use in radiological protection: reference values. (The International Commission on Radiological Protection).

Jantunen M, Katsouyanni, K, Knoppel, H., Kuenzli, N., Lebret, E., Maroni, M., Saarela, K., Sram, R., Zmirou, D. 1999. Final report: Air pollution exposure in European cities: the EXPOLIS study. Kuopio:Kuopion Yliopiston Painatuskeskus.

Klepper O, Den Hollander HA. 1999. A comparison of spatially explicit and box models for the fate of chemicals in water, air and soil in Europe. Ecological Modelling 116:183-202.

Koistinen KJ, Hänninen O, Rotko T, Edwards RD, Moschandreas D, Jantunen MJ. 2001. Behavioral and environmental determinants of personal exposures to PM2.5 in EXPOLIS - Helsinki, Finland. Atmospheric Environment 35:2473-2481.

Opasnet. 2011. Available at: http://en.opasnet.org/w/Category:Intera.

Sarigiannis D, Gotti A, Karakitsios S, Kontoroupis P, Nikolaki S. 2012. Intera Platform: A Tool for Mechanistic Risk Assessment of Indoor Air Pollutants. Epidemiology 23.

Sarigiannis D, Karakitsios S, Van Tongeren M. 2012. TAGS Final report. Thessaloniki, Greece:CERTH, 25.

Schinkel J, Ritchie P, Goede H, Fransman W, Van Tongeren M, Cherrie JW, Tielemans E, Kromhout H, Warren N. 2013. The advanced REACH tool (ART): Incorporation of an exposure measurement database. Annals of Occupational Hygiene 57:717-727.

Simoneau C, van den Eede L, Valzacchi S. 2012. Identification and quantification of the migration of chemicals from plastic baby bottles used as substitutes for polycarbonate. Food Additives and Contaminants - Part A Chemistry, Analysis, Control, Exposure and Risk Assessment 29:469-480.

van Wendel de Joode B, Brouwer DH, Vermeulen R, van Hemmen JJ, Heederik D and Kromhout H. (2003). DREAM: A Method for Semi-quantitative Dermal Exposure Assessment. Ann Occup Hyg; 47(1):71-87.

van Wendel de Joode B, Vermeulen R, van Hemmen JJ, Fransman W, & Kromhout H. (2005). Accuracy of a semiquantitative method for Dermal Exposure Assessment (DREAM). Occupational and Environmental Medicine; 62: 623-632.