This project is funded by
the Horizon 2020 Framework
Programme of the European Union

Deliverable

EuroMix

Grant Agreement n°633172

Collaborative project H2020-SFS-2014-2

Deliverable 1.2 Data management plan

Data management plan

WP 1– Management

Due month of deliverable: M6

Actual submission month: M18

Deliverable leader: RIVM

Document status: version number Creation date:

Project co-funded by the European Commission within the Seventh Framework Programme		
Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission)	
RE	Restricted to a group specified by the consortium (including the Commission)	
CO	Confidential, only for members of the consortium (including the Commission)	

This project is funded by
the Horizon 2020 Framework
Programme of the European Union

D1.2

Data management plan

15 November 2016

This project is funded by
the Horizon 2020 Framework
Programme of the European Union

D1.2

Data management plan

Authors:

Jacob van Klaveren, RIVM, Partner nr 1
Ad Peijnenburg, DLO, Partner nr 2
Trine Husøy, NIPH, Partner nr 15
Hubert Dirven, NIPH, Partner nr 15
Jan Jongenelen, RIVM, Partner nr 1
Corinne Sprong, RIVM, Partnert nr 1

Summary

1. Introduction

The EuroMix project is participating in the Horizon 2020 pilot action on open access to research data. The Data Management Plan (DMP) of the EuroMix project describes the data management life cycle including specific standards of the databases in terms of formats, metadata, sharing, archiving and preservation. The DMP will also clarify which data will be open access, and thereby available to others, and the procedure to gain access to these.

The EuroMix data management will continue to develop during the project and the DMP will be periodically updated. This document represents only an initial draft of the data management life cycle for all datasets to be collected, processed or generated by EuroMix.

In EuroMix, data will be generated and collected within five work-packages. Four data types can be distinguished. Data collection of toxicity data in (WP2) to set first priority for testing, test data generated from *in- vitro* tests and *in-vivo* tests (WP3 and 4), data collections of detailed food consumption and residue data of involved countries needed for future risk assessment of mixtures (WP5), and epidemiological data questionnaires and biomarker data (WP7).

The EC in 2016 provided the *Guidelines on Data Management in Horizon 2020*[1]. According to these guidelines, we need to describe:

1. Dataset reference and name;
2. Dataset description;
3. Standards and metadata;
4. Data sharing;
5. Archiving and preservation.

The next paragraph describes these five requirements for data collected and generated in EuroMix.

[1] Version 2.1. Brussels: European Commission. Available online: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

2. Description of data collected and generated in EuroMix, status November 2016

2.1. Workpackage 2: priority list of chemicals for testing

2.1.1 Dataset reference, name and description

The aim of WP 2 is as follow:

- to apply suitable computational methods enabling prioritisation for testing;
- to prioritise compounds for inclusion in cumulative assessment groups;
- to identify mixtures for further testing in relation to organ or system toxicity;
- to contribute to the development of a tiered testing strategy for the risk assessment of mixtures.

This will be done by searching literature for critical endpoints regarding the toxicity of chemicals to be grouped in cumulative assessment groups. It also includes literature search for intake statistics. When no such data can be found, substitution is considered, e.g. read- across or TTC values for lacking toxicity data.
This type of data is referred to as 'Chemical Inventory list for testing' and the toxicity data of the chemicals on the list are stored in the 'EuroMix Toxicity Database'. Deliverable 2.1 provides a full description of the Chemical Inventory and the EuroMix Toxicity Database.

At this stage of the EuroMix project, the Chemical Inventory list and the EuroMix Toxicity Database serve as input for WP3. Therefore, it is not regarded as useful data to be stored, shared, archived or preserved. Standards and meta data are not applicable.

2.2. Work package 3 and 4: test results

2.2.1 Dataset reference, name and description

The aim of WP3 and WP 4 is:

- to identify the optimal *in-vitro* models for detecting (putative) modes of action for liver toxicity, developmental toxicity, endocrine effects, and immune toxicity, and to define a **bioassay toolbox** for such in vitro testing;
- to investigate the potential of these *in-vitro* test systems for refining cumulative assessment groups (CAGs):
- to test the combined effects of chemicals having similar and dissimilar MoA in a quantitative manner (dose-response measurements) *in vitro*;
- to identify optimal *in-vitro* testing strategies for grouping chemicals and for predicting the effect of **mixtures** based on **molecular signatures;**
- To anchor the key events identified *in vitro* ('omics' and other methods) to animals *in vivo* by using kinetic modelling approaches;
- To develop kinetic models (PB-PK) for IVIVE of mixture exposure;

- To assess the effect of combined exposures to compounds with similar vs dissimilar mode of action;
- To contribute to test strategies with less animal testing (in collaboration with WP2 and WP3);
- To contribute to CAG refinement (in collaboration with WP 2 and WP 3).

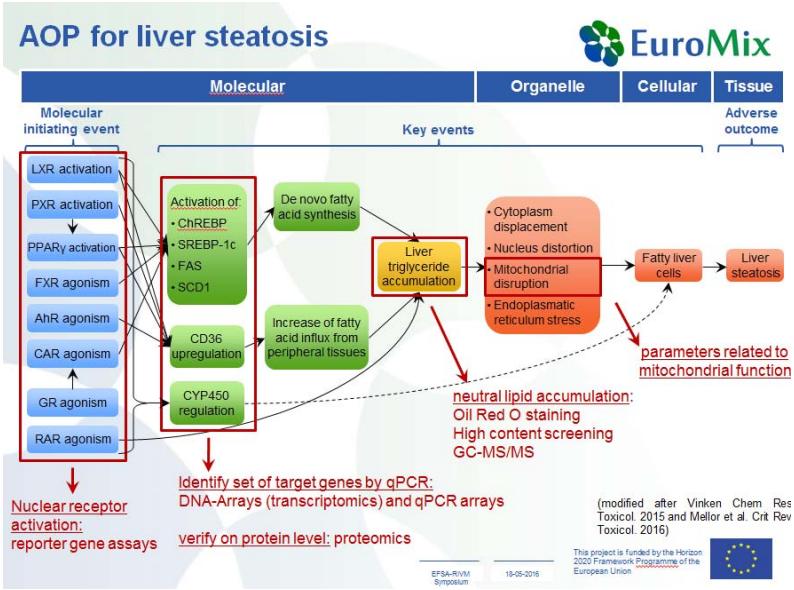
In WP3 15-20 tests have been identified that can be used for mixture testing. At least three chemicals will be tested in all the assays available. The data set is referred to as 'WP3 test results'.

In WP4 3 animal experiments will be performed according to OECD guidance.

2.2.2 Standards and meta data

The selected tests cover a number of key-events in an Adverse Outcome Pathway (AOP). EuroMix has agreed to work on 3 conversing AOPs (liver steatosis, skeletal malformation and endocrine misbalance). An example is given below in Figure 1, tests available in the AOP liver steatosis.

Figure 1: an example of tests available in the Adverse Outcome Pathway liver steatosis



The test results obtained in WP3 will be stored in a web-based platform database/model Proast.RIVM.nl, except for data from -omics studies which will be stored at server of the European Bioinformatics Institute (EBI) within Biostudies (http://www.ebi.ac.uk/biostudies/)

The meta data associated with the test results are, among other things: AOP-number, Key-event number, Test description, Test dose and Response. Regarding data obtained from omics studies, meta data are data related to the set-up and conditions of the -Omics experiments (e.g. compound concentrations/ doses used and time of treatment).

2.2.3 Data sharing

The Proast.rivm.nl web-based version is under construction and still has to be tested. The data is stored in an AOP-wise manner, to better understand the underlying quantitative relationships between key-events of the AOP. This will serve as a 'prove of principle'. If the concept works, the data and structure will be open for sharing with others, providing that others contribute to the maintenance of the EuroMix open data and model platform.

Data from -Omics studies
The data from transcriptomics studies will be stored at a server of the European Bioinformatics Institute (EBI) within Biostudies (http://www.ebi.ac.uk/biostudies/). This concerns:

1. metadata: data that are related to the set-up and condtions of the -Omics experiments (e.g. compound concentrations/ doses used and time of treatment);
2. CEL files that store the results of the intensity calculations on the pixel values. This includes an intensity value, standard deviation of the intensity, the number of pixels used to calculate the intensity value, a flag to indicate an outlier as calculated by the algorithm and a user defined flag indicating the feature should be excluded from future analysis;
3. processed data that can be used for bioinformatics analyses;
4. results of all data analyses.

Only microarrays that pass the quality control requirements (according to recommendations of Affymetrix) will be included in the database.

Microarray data have to be submitted to the database in the ISA-Tab format (Gonzalez-Beltran et al., 2014) securing that all relevant information concerning experimental settings and results are completely and uniformly represented.

Data processing will be performed in an uniform way as well and will include combination of probe IDs to unique Gene Symbols (Gonzalez-Beltran, et al., 2014), RMA normalization (http://www.bioconductor.org/packages/release/bioc/html/affy.html) and flooring. RMA is a common algorithm for converting raw Affymetrix data to gene expression values. In addition, the MBNI Custom CDF, which contains updated probe set definitions for Entrez Gene IDs, will be applied to obtain one expression value per gene (http://brainarray.mbni.med.umich.edu/Brainarray/Database/CustomCDF/11.0.1/entrezg.asp).

Finally, 2 log-ratios of treatments vs. the average of the respective control s (vehicle controls) will be calculated, which enables comparisons of the effects of all treatments within one file. This latter file enables further processing for bioinformatics analyses to detect genes that are up- or down-regulated and to identify pathways or cellular processes that are affected.

A small number of users can add and modify their own data in the EuroMix project account at the EBI server. Other members of EuroMix can access the data but cannot alter the data. Access is provided with use of a password. An example of a Biostudy with some data as tryout (can be assessed without

password) is provided at http://www.ebi.ac.uk/biostudies/preview/studies/S-BSST3. In addition, as a backup, all data will also be stored at a repository at RIKILT (DLO-R).

2.2.4 Archiving and preservation

The aim is to archive data for a number of years beyond the project lifetime, but archiving depends on usefulness of these data for follow-up actions. If the data is useful to better understand system toxicology, it will be useful to store them and to share the data with other data sources in the same infrastructure. RIVM has sufficient infrastructure and capacity to do so. Currently RIVM and EFSA have a partnership agreement on making data and mixture risk assessment models available to all European stakeholders responsible for implementing dietary pesticide risk assessment. However, the follow-up activities are depending on budget and agreements needs to be made as part of a follow-up strategy.
If the follow-up strategy does not work, it does not make sense to store data for a longer period.
The EuroMix project partners are open to discuss data sharing with the Joint Research Infrastructure (IPChem). However, international organization should also discuss cooperation with member states or initiatives that generates the data. Until now, cooperation on these issues are a point of further discussions.

2.3. Workpackage 5: detailed food consumption data and residue data needed for future risk assessment of mixtures

2.3.1. Dataset reference, name and description

Residue and consumption data are needed to perform future mixture risk assessment. National residue and consumption data will be provided by several EuroMix partners. Some of them are data owner, others have agreed with third parties to use the data. The data are organised in WP5. Data will be stored in SQL databases and will be available for use via the EuroMix open data and model platform. All data can be used for future mixture risk assessment by all stakeholders, the raw data underlying the calculations will, depending ownership and rules in the General Food Law, not be available.

The aim of WP5 is:

- To organize relevant data for exposure assessment and modelling;
- To prioritize chemicals to be included in mixture assessment based on exposure profiles;
- To overview relevant aggregated exposure models with relation to food;
- To perform aggregated and cumulative exposure assessment;
- To test integration of PB-PK/D models in aggregated exposure assessment.

This type of data is referred to as ‘detailed food consumption data and residue data’.

Data set description for each involved partner
From 2017 onwards, Horizon2020 requires research data to be open by default, although some options to disclose backgound data are possible (see Chapter 3). In general, this mean that for EuroMix partners being data owner, that their public data are open data. For data owned by third parties, other rules apply. Furhtermore, the residue and consumption data are considered as background as described in the

EuroMix Consortium Agreeement. The tables 1 to 3 describes the data and data owners, they have to decide about openess of background data. For more details on the agreement of data use. Please refer to the Consortium Agreement.

This project is funded by
the Horizon 2020 Framework
Programme of the European Union

Table 1. Data description Food Consumption data. Data owned by third parties are indicated by *italics*

Country	Consumption survey	Age range	Nr individuals	Nr of reporting days per person	Data Owner	Available in MCRA
France	INCA2	3-79	4079	7	ANSES	yes
Netherlands	VCP-kids	2-6	1279	2	*Dutch Ministry Public Health*	yes
Netherlands	VCP –basic	7-69	3818	2	*Dutch Ministry Public Health*	yes
Belgium	Diet_National_2004	14-105	3245	2	UGENT	yes
Czech Republic	SISP04	4-64	2353	2	SZU	yes
Iceland	IS National FC 2010-2011	18-80	1312	2	*Directorate of Health, the Icelandic Food and Veterinary Authority and the Unit for Nutrition Research*	yes
United Kingdom	NDNS	19-64	1724	7	*EFSA, FSA*	yes
Slovenia	PG_SI_child_follow up	11-14	1206	1	NIJZ	yes
Greece	Regional Crete	4-6	874	3	*University of Crete, Faculty of Medicine, Department of Preventive Medicine and Nutrition*	Yes
Cyprus	Childhealth	11-15	303	3	*Research and Education Institute of Child Health*	yes
Spain	Encuesta ENIDE	18-71	3386	3	*Scientific Committee of the Spanish Agency for Consumer Affairs, Food Safety and Nutrition (AECOSAN)*	yes
Denmark	Danish Dietary Survey	4-75	4120	7	DTU	yes
Denmark	DANSDA 2005-08	4-75	2700	7	DTU	yes
Norway	to be transferred				*University of Oslo*	Not yet

This project is funded by
the Horizon 2020 Framework
Programme of the European Union

Table 2. Data description Concentration data. *Data owned by third parties are indicated by italics*

Country	Concentration data	Format	Years	Data owner	Available in MCRA
France	Monitoring data	SSD1	2011-2014	*French General Directorate for Competition Policy, Consumer Affairs and Fraud Control, from the French Ministry for the Economy and Finance, the French Ministry of Health and the French General Directorate for Food from the French Ministry for Agriculture*	yes
France	TDS2	TDS	2007-2009	ANSES	yes
Netherlands	Monitoring data	SSD1	2008-2014	*Dutch Food and Consumer Product Safety Authority, being part of the Dutch Ministry of Economic Affairs*	yes
Belgium	Monitoring data	SSD1	2011-2013	*FAVV (Belgian Food Safety Authority)*	yes
Czech Republic	Monitoring data	SSD1	2011	*Ministry of Agriculture and Ministry of Health*	yes
Czech Republic	TDS	TDS	2004	CZU	?
Iceland	-	-	-	-	-
UK	-	-	-	-	-
Slovenia	Monitoring data	SSD1	2012-2015	*The Ministry of Agriculture, Forestry and Food*	yes
Greece	Monitoring data	SSD1	2011-2015	*Hellenic Food Authority*	yes
Cyprus	Monitoring data	SSD1	2011-2014	SGL	yes
Spain	TDS	TDS	2006-2008	*Catalan Food Safety Agency (ACSA)*	yes
Spain	Monitoring data	SSD1			To be transferred
Denmark	Monitoring data	SSD1	2011-2015	*Danish Veterinary and Food Administration (DVFA).*	yes
Norway	-	-	-	-	-

This project is funded by
the Horizon 2020 Framework
Programme of the European Union

Table 3. Other data sources in WP5. *Data owned by third parties are indicated by italics*

Country	Description data	Data owner	Permission obtained	Available in MCRA
UK	Pesticide usage data	*DEFRA*	yes	Not yet
France	ENNS survey on food consumption and biomonitoring.	*French institute for public health surveillance (INVS)*	Anses is waiting for access rights from Usen, InVS-Paris 13, to use ENNS data in the Euromix project.	Not yet
France	Pesticide concentration in air	*the AASQA (Certified Associations of Air Quality Monitoring)*	Anses will ask the AASQA for access rights to use this data in the Euromix project	Not yet
France	metal concentrations in dust, soil and water	*EHESP School of public health and CSTB*	Anses will ask the EHESP and CSTB for access rights to use this data in the Euromix.	Not yet
France	on pyrethroid and Bisphenol A concentrations in dust and indoor air	*EHESP*	Anses will ask the EHESP for access rights to use this data in the Euromix	Not yet

2.3.2 Standard and metadata

The level of detail of the monitoring and consumption data brought to EuroMix platform is comparable with the level of detail of the consumption and residue data send to EFSA in SSD formats and FoodEx coding.

Standards

Consumption data are coded with the EFSA FoodEx1 code, which is the standard in Europe. Residue data are coded according to the EFSA standards (Standard Sample Description version 1), which are described in the EFSA Data Collection Framework. This framework is agreed between EFSA and the member states.

Metadata

The meta data included background about sampling location, time of sampling, analytical method used, quality aspects of the analytical method, limit of detection of the analytical method etc.

2.3.3 Data sharing

The data will be shared by the partners openly and in all detail. However, the partners cannot capitalize or distribute data owned by others as laid down in the EuroMix Consortium Agreement.

Use conditions are described by the partners in the EuroMix Consortium Agreement and also under paragraph 2.3.1 of this Deliverable. Basic principles are:

- Data can be used by users of the EuroMix data and model platform after registration as a user.
- Data can be used to generate exposure assessment results.
- The data described in table 1 and 2 will be collated in one database to avoid conflicts with the General Food Law.
- The raw data cannot be downloaded from the EuroMix data and model platform.

2.3.4 Archiving and preservation

The data and models will be archieved for 3 years after the finalisation of the project. It is not foreseen to store data longer, because countries will frequently generate newer consumption and residue data.

For this newly generated data, countries will send them to EFSA and those data will be stored in the EFSA dataware house. In the future, member states can retreive their (new) data from the EFSA dataware house and link it to the EuroMix platform. Then they have to comply with the EFSA dataware house policy rules.

If needed additional rules or deviations form the above stated rules can be described in the EuroMix exploitation plan.

2.4. Epidemiological data (WP7)

The NIPHs contribution is subjected to approval from the regional committee of medical research ethics in Norway and/or the Norwegian Data Inspectorate and should comply with standard operating procedures within the Norwegian Institute of Public Health.

All biological materials (regardless of the form or medium in which they are disclosed or stored) collected and/or used in the Project as described below; and all recorded information or data extracted from such material is to be considered as NIPH's background.

All data are anonymised and will be referred to by a study specific reference number for each subject.

2.4.1 Dataset reference and name

The dataset from the human study will be called **Dataset_human_WP7.**

2.4.2 Dataset description

The Dataset_human_WP7 will include several sub-sheets with individual data from 140 subjects compromising of:

- Data extracted from Food frequency questionnaire
- Data extracted from the 24h food diary (for 1 or 2 days)
- Data extracted from the 24h cosmetic use diary (for 1 or 2 days)
- Data on the concentrations of 3 metabolites of interest as biomarker for exposure
- Data on at least 2 analytes of interest as determined as biomarker for effect

2.4.3 Standards and metadata

The Food frequency questionnaire and the diet diary have previously been developed, standardised and used in MoBA. The cosmetics diary has been developed for the purpose of this particular study. Metadata will include background information on time of sampling, days between sampling (if relevant), analytical methods used (with reference to described analytical methods). Where possible relevant standard EFSA data formats will be used.

2.4.4 Data sharing

The background data (country specific data) will, however, become foreground as defined in chapter 8 and can be used by partners in the EuroMix project. The data will be shared with the partners openly according to the Consortium Agreement.

2.4.5 Archiving and preservation

-All paper-based data will be archived for at least 10 years after the finalisation of the project. Biological materials will be stored for at least 2 years after the finalisation of the project.

3. International rules on open data

Horizon2020 mandates open access to all scientific publications. In addition, the data underlying the scientific publication must be open, in order that peers can reassess the conclusions of the research project. Furthermore, all Horizon signed projects were encouraged to provide open access to research data as much as possible.

From 2017 onwards, Horizon2020 requires research data to be open by default. Not only data underpinning a publication, but also all the data gathered in the project should be open by default. However, there are possibilities to opt out. The motto is "as open as possible, as closed as necessary".

The opt out possibilities under Horizon2020 are in case:

- Privacy of data-subjects is harmed;
- Intellectual property rights lie elsewhere;
- Publication of the data will jeopardize the objective of the project itself.

EuroMix will comply with open data by default rule, but if this is not possible because of the reasons mentioned above, the reasons for not doing so will be explained.

Because EuroMix will combine different kind of data from different institutions in Europe it will be very well possible that certain datasets will be open and others will not. The providers of the data have the rights of the data. They will be the ones to decide whether the data will be open or not. Therefore, there is a necessity for clear data delivery protocols and agreements amongst the participants. Such protocols and agreements state under which conditions data are provided, to prevent discussion in a later stage about open access and ownership.

Under Dutch regulation, the breach of privacy is also one of the most important legitimate reasons to refrain from publishing data. However, there are also other reasons to opt out, e.g. if the safety of the nation or the crown is jeopardized. Forensic data and data gathered for inspection and auditing goals are excluded. All these opt out clauses are presented in a decision tree that helps data-owners to assess systematically if data can or should be open. More (Dutch) information for opt outs under the Dutch regulation can be found at the following website : http://www.rivm.nl/Documenten_en_publicaties/Algemeen_Actueel/Nieuwsberichten/2016/Beslisboom_voor_publiceren_open_data_ontwikkeld

Some researcher might be reluctant to hand over their data when they fear that others might reap the benefits before they themselves have done so. Under Dutch regulation, this will not be an acceptable excuse to refrain from publishing. This sentiment remains to be a factor to be aware of when striving for being as open as possible.