Data Management Plan (DMP)

Deliverable Report

D 10.1

WP 10 - Data Management and Analysis

Deadline: June 2017

Upload by Coordinator: [31.07.2017]

With separate attachment: HBM4EU Data Policy

Entity	Name of person responsible	Short name of institution	Received [Date]
Coordinator	Marike Kolossa-Gehring	UBA	20/07/2017
Grant Signatory	Greet Schoeters	VITO	14/07/2017
Pillar Leader	Argelia Castaño	ISCIII	20/07/2017
Work Package Leader	Greet Schoeters	VITO	14/07/2017
Task leader	Sylvie Remy	VITO	14/07/2017

Responsible author	Sylvie Remy	E-mail	sylvie.remy@vito.be
Short name of institution	VITO	Phone	+32 14 33 51 07
Co-authors (alphabetical order)	André Conrad (UBA); Silvia Dalla Costa (JRC); Catherine Ganzleben (EEA); Eva Govarts (VITO); Greet Schoeters (VITO); Nina Vogel (UBA); Xenia Trier (EEA)		

Table of contents

Authors and acknowledgements

Lead authors

Sylvie Remy (VITO);

Silvia Dalla Costa (JRC);

Eva Govarts (VITO);

Greet Schoeters (VITO)

Contributors

Nina Vogel (UBA);

André Conrad (UBA);

Catherine Ganzleben (EEA);

Xenia Trier (EEA)

Acknowledgements

We would like to thank the above mentioned partners for their valuable contributions to this deliverable.

Introduction

The HBM4EU project aims at the coordination and harmonisation of existing HBM (human biomonitoring) initiatives in 26 countries, including 22 European Member States and three associated countries, as well as Switzerland. These national HBM initiatives are managed by leading HBM experts in Europe. It is encouraged to make existing data available for research within HBM4EU, and to the broader research community, stakeholders, and policy makers. In addition, new HBM data, i.e. data generated during the project with HBM4EU co-fund, will be used for research during the course of the project. It will be mandatory to allow use of data generated with HBM4EU co-fund as single measurement data (i.e. individual records obtained from data subjects) to meet the objectives of the HBM4EU project. These data will be shared to highest possible resolution to the broader research community, stakeholders, and policy makers, while respecting the ethico-legal framework.

This document is the HBM4EU data management plan (DMP). The DMP describes the data management life cycle for all datasets to be collected, processed and/or generated by the research project. The H2020 DMP describes, among others:

- the handling of research data during and after the project
- the type of data that will be collected, processed, or gathered
- what methodlogy and standards will be applied
- whether and how the data will be made (openly) accessible
- how the data is stored

As a separate attachment to the DMP, the HBM4EU data policy has been designed. **The procedures described in the HBM4EU data policy shall be followed by all members of the consortium** and ensure that data on human subjects are transferred and used in a secure setting; that use of the data is compliant with ethico-legal requirements (including signed informed consent, ethics approval, and the applicable data protection laws, furthermore the EU data protection regulation, which is applicable from May 2018); and that the use of both existing as well as new data occurs in agreement with the Data Owner/Data Provider. Management of datasets that include personal information and health information of study participants will be compliant with the General Data Protection Regulation (GDPR, Regulation (EU) 2016/679). The GDPR is a regulation by which the European Parliament, the European Council and the European Commission intend to strengthen and unify data protection for individuals within the European Union (EU).

To share data on human subjects between HBM4EU partners, a secure platform has been established: the HBM4EU repository. Beyond this, it is a principle aim of the project to increase the availability of human biomonitoring data to policy makers, stakeholders and the broader research community, in order to multiply the benefits that can be generated through its use. As such, metadata of all datasets that are subject to this data policy will be directly integrated into IPCheM[1] – the Information Platform for Chemical Monitoring - as a minimal requirement. This will allow identification of existing datasets and enable to contact the Data Owner/Data Provider to request access to use the data. Integration of aggregated and single measurement data will be stimulated, while respecting the ehtico–legal framework.

---

[1] https://ipchem.jrc.ec.europa.eu/

1   Data summary

Points to be addressed:

Provide a summary of the data addressing the following issues:

- State the purpose of the data collection/generation
- Explain the relation to the objectives of the project
- Specify the types and formats of data generated/collected
- Specify if existing data is being re-used (if any)
- Specify the origin of the data
- State the expected size of the data (if known)
- Outline the data utility: to whom will it be useful

1.1   What is the purpose of the data collection/generation and its relation to the objectives of the project?

The purpose is to collect and harmonize human biomonitoring data across Europe. The data collection will comply with all national and EU ethics and legal requirements. Access to use these data is needed to address the different HBM4EU objectives as specified in the grant agreement and summarized below.

- **Objective 1**: Laying the foundations for a pan-European HBM platform that builds on national hubs and existing expertise;
- **Objective 2**: Developing a common methodology for the interpretation and use of HBM data in policy-making;
- **Objective 3**: Harmonising and optimising the practices of national HBM programmes, including sample collection, quality assurance and data management;
- **Objective 4**: Identifying gaps where further data are needed to inform current policy questions and design new, targeted studies to address these knowledge gaps;
- **Objective 5**: Including new HBM data and, where possible, existing HBM data in the European Commission´s Information Platform for Chemical Monitoring (IPCheM, https://IPCHEM.jrc.ec.europa.eu/);
- **Objective 6**: Linking external to internal exposure in order to improve exposure models for risk assessment;
- **Objective 7**: Developing, validating, and applying exposure and effect biomarkers to improve our understanding of the health risks associated with aggregate exposures;

- **Objective 8**: Identifying chemicals of concern through novel methods for the holistic analysis of HBM samples and improving the use of HBM data in assessing exposure to and the risks of chemical mixtures;

- **Objective 9**: Enhancing our understanding of the causal association between chemical exposure and adverse health outcomes by combining mechanistic studies with existing cohort data;

- **Objective 10**: Promoting capacity building at national level through training and exchange programmes;

- **Objective 11**: Engaging with stakeholders, including the general public, throughout the programme to ensure the credibility, accountability and legitimacy of activities and results.

1.2 What types and formats of data will the project generate/collect?

Types of the data:

- Data from self-administered paper-based/online questionnaires, CAPI, CATI: lifestyle, health, exposure-relevant behaviours

- Data from clinical assessment: physiological, cognitive and anthropometric measures

- Measurements in biological matrices/tissues (such as hair, urine, nails, blood, placenta, adipose tissue)

- Molecular data: data on (part of/whole) genome, transcriptome, proteome, metabonome

- Exposure data: internal biomarkers of exposure

- Modelling data: estimated exposure and/or effect parameters

Formats of the data:

- Data and metadata will be requested, stored and transferred (across partners and in IPCheM) in a comma-separated values (CSV) format.

- To facilitate the data exchange, MS Excel compatible files including comma separated and .xls(x) format will be also accepted.

- For statistical purposes, other formats include .sas7bdat (SAS), .RData (R), .SAV (SPSS), .mat (matlab).

- Where applicable data formats may be migrated when new technologies become available and are proved robust enough to ensure digital continuity and continued availability of data.

1.3 Will you re-use any existing data and how?

Yes, it is encouraged to make existing data available for research within HBM4EU. WP10 will provide data templates and lists of pre-defined values, in order to be able to harmonize the different datasets that are provided.

For single measurement data, a template will be created and filled out by the providers of the data.

For aggregated data, an R-script (and - if necessary for data weighting - an SPSS-Syntax) will be provided to generate comparable summary statistics.

1.4 What is the origin of the data?

Human biomonitoring initiatives and projects in which human biomonitoring and/or exposure data have been collected across Europe.

1.5 What is the expected size of the data?

To be evaluated during the course of the project. The expected size depends on the extend and the nature of the data that are made available.

1.6 To whom might it be useful ('data utility')?

- HBM4EU consortium;
- European Commission services and European Agencies;
- EU National Bodies;
- The general public including the broader scientific community

2 FAIR data

Points to be addressed:

- In general terms, your research data should be 'FAIR' that is findable, accessible, interoperable and re-usable. These principles precede implementation choices and do not necessarily suggest any specific technology, standard or implementation-solution.

2.1 Making data findable, including provisions for metadata:

Points to be addressed:

- Outline the discoverability of data (metadata provision)
- Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?
- Outline naming conventions used
- Outline the approach towards search keyword
- Outline the approach for clear versioning
- Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how

2.1.1 Metadata provision

A specific Human biomonitoring metadata template, as an extension of the IPCheM metadata schema, has been defined, in order to describe, discover and trace existing data collected by the HBM4EU project and the data that will be generated by it over the next years. The templates will be sent and filled out by the data owners/ data providers and saved in the HBM4EU repository. Afterward a metadata fiche, for each HBM4EU data collection will be made publicly accessible through IPCheM.

In addition a list of studies and available variables will be compiled, which will help HBM4EU data users to identify potential datasets for analysis.

2.1.2 Standards for metadata creation

The Human biomonitoring metadata proposal is a (thematic) extension of the IPCheM metadata schema, used to describe chemical monitoring data collections, e.g.:

- chemical occurrence data collected as result of legal obligations on ad-hoc or regular basis for reporting /monitoring at European or national levels;
- data generated as result of targeted research on the presence of known or unknown chemical substances in specific media in a European country/region.

The IPCheM metadata is compliant with two European metadata standards, namely:

1. INSPIRE metadata elements for spatial data sets and services (see these elements in the INSPIRE Metadata Regulation: http://data.europa.eu/eli/reg/2008/1205/oj#d1e600-14-1
2. The "DCAT application profile for European data portals" (DCAT-AP), developed in the framework of the EU ISA Programme. The European Data Portal is implementing the DCAT-AP as the common vocabulary for harmonising descriptions of datasets harvested from several data portals of 34 countries. The DCAT-AP specification is available at: https://joinup.ec.europa.eu/asset/dcat_application_profile/

2.1.3 Approach towards search keyword

The dataset information reported into the metadata fiche will be published in IPCheM, where specific filters, based on the metadata elements, will allow to refine the search across datasets (e.g. search dataset by chemical or chemical group, by temporal or spatial coverage of the data, by key words, etc.)

2.1.4 Use of persistent and unique identifiers

The assignment and management of persistent identifiers (PIDs) to the data will be assessed in the course of first year of the project, starting with a proof of concept carried out by the IPCheM team and an evaluation of available PID services (either managed by the European Commission or third parties organisations)

2.1.5 Naming conventions

For metadata, dataset and template names we will define naming convention consisting in 3 mandatory parts:

- A prefix, indicating if it is a dataset, a metadata or a template
- A root composed by:
- the short and meaningful name of the dataset/template
- the acronym/short name of the data provider organisation(s) (HBM4EU by default for templates)
- A suffix indicating the date of the last upload into the Repository in YYYYMMDD format.

Each of these elements are separated by an underscore: _

2.1.6 Clear versioning

The versioning management of the data, metadata template and in general the files stored into the Repository will be applied at two levels:

1. Via the naming convention and the use of the date as suffix, indicating the last version of the file uploaded into the Repository;
2. As capabilities of the Nextcloud Repository sets up for the project, since the solution supports the simple version control system for the uploaded files (more info are available here: https://docs.nextcloud.com/server/9/user_manual/files/version_control.html

2.2 Making data openly accessible:

Points to be addressed:

- Specify which data will be made openly available? If some data is kept closed provide rationale for doing so
- Specify how the data will be made available
- Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?

- Specify where the data and associated metadata, documentation and code are deposited
- Specify how access will be provided in case there are any restrictions

2.2.1 Directly accessible via HBM4EU repository to HBM4EU consortium partners

To share data with HBM4EU consortium partners, a repository has been set up (OSS: NextCloud 11: https://nextcloud.com/. It provides access to data through a web interface and also a platform to view, sync and share the files across devices easily — all under user's control. It uses an open architecture, extensible via API for applications and plugins and it works with any storage.

The **HBM4EU repository** has been set up for this project as it:

- Is a platform that facilitates sharing of data, intermediate results, and results
- Is hosted at JRC – (server location: JRC, Ispra, Italy) - as one of the components of the IPCheM architecture and allows integrating part of data in the repository into IPCheM, when agreed with the Data Owners/Data Providers.
- Is needed to enable the analysis of human biomonitoring data, but also of accessory external exposure data and health data to meet the goals of HBM4EU.
- Enables data users to work with selected quality controlled data sets and versions approved by the Data Owners/Data Providers.
- Enables that Data Owners/Data Providers can chose to which research they will contribute with their data; that use of the data can be detailed, diversified, and flexible according to purpose and to interests of the Data Owners/Data Providers
- Aims to reach the highest level of GDPR compliancy, amongst others by:
  - Relying on the EU authentication platform and security protocols for data sharing.
  - Applying a strict policy in granting and revoking access to the data.
  - Logging of user identity during data access, download, and upload, including version control. This enables to restore the availability and access to the data in a timely manner in the event of a physical or technical incident.

For **data generated with HBM4EU co-fund** during the course of HBM4EU, the Data Owner/Data Provider shall agree that these data are transferred at high level of granularity to the HBM4EU repository (anonymised or pseudonymised single measurement data); and that the accompanying variables of the study that are needed to solve the envisaged research purpose(s) are also provided as single measurement data. This is a necessity to meet the objectives of HBM4EU. Prior to generation of the data, the Data Owner/Data Provider shall confirm ethico-legal compliance of the study in which new data are generated; and fill out and sign the data transfer agreement. Data generated with HBM4EU co-fund and accompanying variables are, *by default*, directly accessible for use within HBM4EU following the procedures outlined in section 4.

For **existing data, not generated with HBM4EU co-fund**, the Data Owner/Data Provider specifies the level of granularity that data will be transferred: anonymised single measurement data; pseudonymised single measurement data; or aggregated data. The Data Owner/Data Provider indicates for each level of granularity whether the data are directly accessible for use within HBM4EU. In case the Data Owner/Data Provider indicates that the data are not directly accessible for use within HBM4EU, the Data Owner/Data Provider will be asked approval when consortium members request access to the data to meet the goals of a particular objective.

2.2.2 Openly accessible via IPCheM to HBM4EU consortium; European Commission services and European Agencies; EU National Bodies; The general public

From the HBM4EU repository, meta-data are *by default* directly integrated into IPCheM. The meta-data in IPCheM are openly available to European Commission services and European Agencies; EU National Bodies; HBM4EU consortium; and the general public. Aggregated data and individual data are only directly integrated into IPCheM if the conditions set by the Data Owner in the "IPCheM Participation for HBM4EU project" allow. For aggregated and single measurement data, the Data Owner can indicate to grant/revoke access and use permission to *a priori* defined IPCheM user groups: European Commission services and European Agencies; EU National Bodies; HBM4EU consortium; and the general public.

2.3 Making data interoperable:

Points to be addressed:

- Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.
- Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?

At present, no specific data and metadata vocabularies are available for the field of Human Biomonitoring.

A common vocabulary and code lists of pre-defined values for harmonising the descriptions of Human Biomonitoring metadata and data are under definition and will be defined in the course of the project, using when possible the code lists and their values, as defined in the INSPIRE implementing rules on metadata (Commission Regulation (EC) No 1205/2008): http://inspire.ec.europa.eu/metadata-codelist

2.4 Increase data re-use (through clarifying licenses):

Points to be addressed:

- Specify how the data will be licenced to permit the widest reuse possible
- Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed
- Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why
- Describe data quality assurance processes
- Specify the length of time for which the data will remain re-usable

This section will be compiled throughout the course of the project, when we get more information on the datasets that are made available for HBM4EU.

3 Allocation of resources

Points to be addressed:

Explain the allocation of resources, addressing the following issues:

- Estimate the costs for making your data FAIR. Describe how you intend to cover these costs
- Clearly identify responsibilities for data management in your project
- Describe costs and potential value of long term preservation

3.1 Estimation of costs

One of the objectives of HBM4EU is integration of HBM data into the European Commission´s Information Platform for Chemical Monitoring (IPCheM), which is managed by JRC. In this context, a collaboration between HBM4EU and JRC has been established. The HBM4EU repository is hosted at JRC as one of the components of the IPCheM architecture. Costs for establishing and maintaining the HBM4EU data repository are covered by the financial budget of IPCheM.

While the repository in itself is not maintained after the end of the project, all files stored within the HBM4EU repository shall be stored after the project to meet the requirements of good scientific practice. A strategy for storage of the files after the project is being developed and will be included in the DMP later.

3.2 Responsibilities for data management

The JRC-IPCheM team (ipchem-support@jrc.ec.europa.eu), VITO (HBM4EU.DATAMANAGEMENT@vito.be), and UBA (HBM4EU.DATAMANAGEMENT@uba.de) are co-managers of the data repository, with these specific responsibilities.

The JRC is responsible for:

a) Initial set-up of the hardware and software components of the data repository

b) Maintenance of the hardware and software components of the data repository

c) Carrying out the initial Security assessment of the repository

d) Perform Security Assessment on a regular basis (e.g. one year) in order to guarantee the agreed security level

e) Reporting and blocking any possible security threat, taking appropriate measures accordingly

f) Creation and management of the internal User Group Account database (LDAP), as one of the component of the data repository

The JRC responsibilities include:

- setting up and upgrading, when needed, of the hardware and software components of the repository
- creation, maintenance and upgrading of the User Group Account database
- co-creation, under specific instructions provided by VITO/UBA, of the data repository's folders/sub-folders for each user group and document type (e.g. data, metadata, templates)
- capacity management of hardware and software components

The JRC responsibilities does not include the definition of:

- The user groups
- the list of members belonging to one or more user groups
- the access, upload and download rights for each user group

The JRC is not responsible for the interruptions of the data repository services that are due to force majeure.

The JRC is not responsible of the content (of data and documents) reported into the data repository, that must be compliant with this HBM4EU Data Policy and the IPCheM Data Policy (since the data repository is part of).

In the event of an ICT Incident, the JRC will follow its internal ICT Incident Response procedure, which may entail granting access to the platform to EC Security staff, including generation of a forensic copy of the server for further analysis; EC Security Service will act in compliance with GDPR.

The JRC obligations

The JRC shall:

- Not distribute the list of members for each user groups, except to VITO and UBA.
- Inform VITO and UBA about any scheduled interruptions due to data repository services upgrading or technical interventions at the JRC

And in general term comply with personal data protection rules (Regulation EC 45/2001)

VITO is responsible for a-e and UBA is responsible for a-b:

a. Collecting the users request for access to and download of data
b. Preparing, checking the list of User groups and members for each user group
c. Providing to the JRC the list of User groups and members for each user group that will be transferred into the internal User Group Account database (LDAP)
d. Definition, creation, updating of the data repository structure, i.e.: structure of folders and subfolders, names, contents and access, upload, download rights
e. Co-creation with the JRC of the data repository's folders/sub-folders for each user group and document type (e.g. data, metadata, templates)

VITO responsibilities include a-e and UBA's responsibilities include a+c:

a. Ensuring that the list of members is constantly up to date and consistent with the assigned rights
b. Providing instructions to the JRC about the data repository structure
c. Timely communication with the JRC of any change in term of groups (e.g. name, rights) and related membership
d. Communication to the user any scheduled interruptions due to data repository services upgrading or technical interventions at the JRC.
e. Communication to the users any unplanned interruption of services

The VITO and UBA responsibilities do not include:

▪ The creation and maintenance of the internal User Group Account database (LDAP)

The VITO obligations

▪ Verify that the data and documents uploaded into the repository are following and compliant with what is reported into this HBM4EU Data Policy

▪ Timely communicate any possible compliance issue to the JRC

▪ Timely communicate any possible incompliance with this HBM4EU Data Policy to the Data Providers

Assistance to data providers and data users is coordinated via the HBM4EU helpdesk on data management that is accessible via the internal pages of the HBM4EU website. The helpdesk is managed by VITO and UI for HBM4EU consortium partners. Policy makers can address their questions to HBM4EU.DATAMANAGEMENT@vito.be.

4 Data security

Points to be addressed:

▪ Address data recovery as well as secure storage and transfer of sensitive data

With respect to Privacy and Data Protection, the forthcoming EU-legislation - the General Data Protection Regulation - imposes several new obligations upon the consortium partners being data processors. Moreover, several new rights are granted to data subjects and significant fines are introduced in case of a data breach.

Apart from this legislation, the consortium partners regard privacy and data protection as a fundamental principle and hence apply a strict policy on this matter.

4.1 Data confidentiality and integrity

The data confidentiality and integrity are implemented at various levels:

▪ Data at rest - stored at the JRC Data Repository - is protected against unauthorised access by means of standard EU Login (former ECAS authentication). Appropriate access levels will be granted by the creation of groups

- Data in transit is secured by means of secure data transfer mechanisms, such as TLS 1.2[2] (Transport Layer Security)

- Data access is logged by a tamper-proof logging mechanism built into NextCloud software, the log files are stored within an encrypted file system, and configured in append-only

- Consortium partners will impose a strict policy on all employees, co-workers, subcontractors ... having access to the data. This policy will include, but is not limited to,
  - allowing copies on local devices only during processing of the data with guaranteed erasure after being processed
  - extending the access control policies to the local copies
  - contractual clauses
  - agreement to terms and conditions before access is granted
  - etc

- Data will be pseudonmised up to the level as to not interfere with the quality of the research

- Lastly, awareness on data privacy and security will be enhanced (a.o. by attending a webinar on this matter prior to be granted access to the repository; attending this webinar shall be mandatory at least yearly during the course of the project.)

4.2 Data availability

Business continuity and data availability are guaranteed by JRC, with a Recovery Time Objective (RTO) of 7 days and a Recovery Point Objective of 2 days;
Moreover, a DPIA (Data Privacy Impact Assessment) will be carried out, as well as other obligations as stated in the GDPR (e.g. a register of processing activities, data breach notification procedures, etc).

5 Ethical aspects

Points to be addressed:

- To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former

The transfer of data on human subjects to the HBM4EU repository is only considered when: informed consents, ethics approval and – when applicable - approval by local data protection authorities cover the purpose that the data are envisaged to be used within HBM4EU and allow transfer of individual or aggregated data to the HBM4EU repository.

---

[2] https://www.ietf.org/rfc/rfc5246.txt

All data that are transferred to the HBM4EU repository shall be either pseudonymised or completely anonymized. The Data Owner/Data Provider is responsible for the anonymization or prseudonymation process and for ensuring that identifiable variables are not transferred to the HBM4EU repository. Directly identifiable variables include - but are not limited to - national ID number, name, phone number, ZIP-code, e-mail address, address, geographical coordinates (at a resoluation that risks identification). One shall also be aware that a combination of just of few indirect identifying variables (such as birth data, gender, and zip-code) can be used to identify a large portion of individuals on any dataset. In this context, the Data Owner/Data Provider shall only provide such variables at the lowest possible resolution that is necessary to for analysis, e.g. district instead of zip-code; year of birth or age instead of birth date.

6 Other

Points to be addressed:

- Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

The procedures for data sharing, and requesting access to use the data are described in the HBM4EU data policy (separate attachment).