Microbiology In Nuclear waste Disposal

(GRANT AGREEMENT: 661880)

DELIVERABLE D4.4

Data Management Plan (DMP)

Date of issue of this report: **01/12/2015**

Start date of project: 01/06/15 Duration: 48 Months

This project has received funding from the Euratom research and training programme 2014-2018 under Grant Agreement no. 661880		
Dissemination Level		
PU	Public	**PU**
RE	Restricted to a group specified by the partners of the MIND project project	
CO	Confidential, only for partners of the MIND project	

Contents

1 Introduction

The European Commission wishes to encourage good Data and Knowledge Management practice and is therefore asking each of the H2020-projects to submit a Data Management Plan as a deliverable for each of the actions supported by the Horizon 2020 framwork program.

Within the MIND-project we have decided to create a common Data Management plan which is this document. The document will be updated continuously by the project members on Syncplicity whenever new datasets are made available. An updated version will be published at the end of each reporting period.

This plan is based on the Guidelines on Data Management in Horizon 2020 Version 1.0 11 December 2013 http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

2 General guidelines

This plan is structured as a series of questions that should be ideally clarified for all datasets produced in the project.

Based on the Horizon 2020 Guidelines on Data Management; Scientific research data should be easily:

1. **Discoverable**
   DMP question: are the data and associated software produced and/or used in the project discoverable (and readily located), identifiable by means of a standard identification mechanism (e.g. Digital Object Identifier)?

2. **Accessible**
   DMP question: are the data and associated software produced and/or used in the project accessible and in what modalities, scope, licenses (e.g. licencing framework for research and education, embargo periods, commercial exploitation, etc.)?

3. **Assessable and intelligible**
   DMP question: are the data and associated software produced and/or used in the project assessable for and intelligible to third parties in contexts such as scientific scrutiny and peer review (e.g. are the minimal datasets handled together with scientific papers for the purpose of peer review, are data is provided in a way that judgments can be made about their reliability and the competence of those who created them)?

4. **Useable beyond the original purpose for which it was collected**
   DMP question: are the data and associated software produced and/or used in the project useable by third parties even long time after the collection of the data (e.g. is the data safely stored in certified repositories for long term preservation and curation; is it stored together with the minimum software, metadata and documentation to make it useful; is the data useful for the wider public needs and usable for the likely purposes of non-specialists)?

5. **Interoperable to specific quality standards**
   DMP question: are the data and associated software produced and/or used in the project interoperable allowing data exchange between researchers, institutions, organisations, countries, etc. (e.g. adhering to standards for data annotation, data exchange, compliant with available software applications, and allowing recombination with different datasets from different origins)?

The MIND-project has received funding from the European Union's EUROATOM research and training program (Horizon2020) under grant agreement 661880.

3 MIND Data-sets

The MIND Data Management Plan (DMP) has three main purposes:

1. To support the project partners when publishing the datasets generated by the project.
2. Support the Coordinator in assuring the availability of the data generated by the project.
3. Assist external parties in analysing the work performed and accessing the data generated by the project.

To ensure that all three purposes are fullfilled we have standardized the information that needs to be given regarding each of the datasets generated by the project.

The project partner responsible for producing a dataset will fill out the two tables show below which will be added to this dokument as individula sub-chapters to chapter 3 in this document (3.1, 3.2, 3.3 etc.). Assitance in doing this will be available through both the work package managers and the coordinators.

Metadata	Description
Dataset reference and/or name	Unique name identifying the dataset. Identifier should start with EU-MIND2020-WPx where "x" is the relevant work package number followed by a three digit number. Example: "*EU-MIND2020-WP4-001*"
MIND Datatype	Choose one or more of the relevant data types: Experimental Data, Observational Data, Raw Data, Derived Data, Physical Data (samples), Models, Images and Protocols. Alternatives are further described in Appendix 1.
Source	Source of the data. Reference should include work package number, task number and the main project partner or laboratory which produced the data.
Dataset description	Provides a brief description of the data set along with the purpose of the data and whether it underpins a scientific publication. This should allow potential users to determine if the data set is useful for their needs.
Standards and metadata	Provides a brief description of the relevant standards used and list relevant metadata in accordance with the description in Appendix 1. The usage of the Directory Interchange Format is optional.
Science Keywords	List relevant scientific key words to ensure that the data can be efficiently indexed so others may locate the data.
Data sharing	Description of how data will be shared both during and after the MIND2020 project. Include access procedures, embargo periods (if any), outlines of technical mechanisms for dissemination and necessary software and other tools for enabling re-use, and definition of whether access will be widely open or restricted to specific groups. Information should include a reference to the repository where data will be stored. In case the dataset cannot be shared, the reasons for this should be mentioned (e.g. ethical, rules of personal data, intellectual property, commercial, privacy-related, security-related).
Archiving and preservation	Description of the procedures that will be put in place for long-term preservation of the data.

The MIND-project has received funding from the European Union's EUROATOM research and training program (Horizon2020) under grant agreement 661880.

3.1 “Dataset one”

Metadata	Description
Dataset reference and/or name	…
Source	…
Dataset description	…
Standards and metadata	…
Data sharing	…
Archiving and preservation	…

The MIND-project has received funding from the European Union’s EUROATOM research and training program (Horizon2020) under grant agreement 661880.

3.2 "Dataset two"

Metadata	Description
Dataset reference and/or name	…
Source	…
Dataset description	…
Standards and metadata	…
Data sharing	…
Archiving and preservation	…

The MIND-project has received funding from the European Union's EUROATOM research and training program (Horizon2020) under grant agreement 661880.

APPENDIX A: MIND2020 DATA TYPES

1 Experimental Data

1.1 Dataset description

The experimental data originate from measurements performed in a laboratory environment, be it *in situ* or *ex situ*. The data comprise point or continuous numerical measurements (e.g. pH, temperature), microbial counts, dose measurements)

The data will be collected either on a sample basis (sampling an experiment at a certain point in time) or on an experiment scale (without sampling the experimental set-up). Data can be derived from either destructive of preservative analyses. Experimental data collection can occur automatically or manually, and will be available in a digital or a hard copy format. In the case of the latter, experimental data will first be copied to e.g. a lab book and then digitized.

Experimental data are supposed to be unique, in the way that new experiments will be set-up, producing fresh data. In some cases, similar data will be available from previous/other experiments within the project, within the partners' institution or from overlapping projects, allowing comparison and integration of the newly obtained data.

Experimental data will be used in downstream statistical analyses (hypothesis testing, correlations, etc.), interpretations, quantifications and modelling approaches.

1.2 Standards and metadata

Experimental data are obtained using standardized laboratory techniques which are calibrated when applicable. Positive and negative controls are used and standards, internal or external, are introduced.

Metadata (summary information) can optionally be provided according to a Directory Interchange Format (DIF). A DIF allows users of data to understand the contents of a dataset and contains those fields which are necessary for users to decide whether a particular dataset would be useful for their needs.

2 Observational Data

2.1 Dataset description

Observational research (or field research) is a type of correlational (i.e., non-experimental) research in which a researcher observes ongoing behaviour.

2.2 Standards and metadata

The metadata for observational data should include any standards used and the necessary information so that an external researcher has the possibility to analyse how the data was gathered.

3 Raw Data

3.1 Dataset description

Raw data are primary data collected from a source, not subjected to processing or any other manipulation.

The MIND-project has received funding from the European Union's EUROATOM research and training program (Horizon2020) under grant agreement 661880.

Raw data are derived from a source, including analysis devices like a sequencer, spectrometer, chromatograph etc. In most cases, raw data are digitally available. In some cases (e.g. sequencing), the raw data will be very extensive datasets.

Raw data has the potential to become information after extraction, organization, analysis and/or formatting. It is therefore used as input for further processing.

3.2 Standards and metadata

Raw data are obtained using standardized laboratory techniques which are calibrated when applicable. Positive and negative controls are used and standards, internal or external, are introduced.

Metadata should at least include standards, techniques and devices used. Metadata can optionally be provided according to a DIF. A DIF allows users of data to understand the contents of a dataset and contains those fields which are necessary for users to decide whether a particular dataset would be useful for their needs.

4 Derived Data

4.1 Dataset description

Derived data are the output of the processing or manipulation of raw data.

Derived data originate from the extraction, organization, analysis and/or formatting or raw data, in order to derive information from the latter. In most cases, derived data are digitally available, as are the raw data. Derived data will allow for the interpretation of laboratory experiments, e.g. through statistical analysis or bioinformatics processing.

4.2 Standards and metadata

Manipulation of data will be performed using a 'scientific code of conduct', i.e. maintaining scientific integrity and therefore not falsifying the output or its representation.

Metadata should include any standard or method or best practice used in the analysis. Metadata can optionally be provided according to a Directory Interchange Format. A DIF allows users of data to understand the contents of a dataset and contains those fields which are necessary for users to decide whether a particular dataset would be useful for their needs.

5 Physical Data (samples)

5.1 Dataset description

Physical data are samples that have been produced by an experiment or taken from a given environment. Sampling of an environment or experiment is performed in order to obtain information through analyses. As such, experimental, raw and or derived data will be obtained from physical data. When the analyses are destructive, the samples cannot be stored for later use. When the analyses are preservative, samples can be stored for later use, but only for a limited time. Environmental samples will primarily be samples from the Underground Research Facilities and analogue sites.

The MIND-project has received funding from the European Union's EUROATOM research and training program (Horizon2020) under grant agreement 661880.

5.2 Standards and metadata

When sampling an environment or experiment, blank samples are taken as well, as a reference. In case of microbiological samples, a blank can be a non-inoculated experiment.

Metadata should include description of the origin of the sample, age, processing, storage conditions and expected viability of the sample (as some sets of samples can only be stored for a limited time, due to their nature).

6 Models

6.1 Dataset description

Representation or simplified version of a concept, phenomenon, relationship, structure or system used for facilitating understanding by eliminating unnecessary components.

6.2 Standards and metadata

References and metadata should include existing standards of the discipline used, tools used in the modelling and focus of the modelling.

7 Images

7.1 Dataset description

Imaging data are optical semblances of physical objects.

Objects of macro- and microscopic scale can be imaged in a variety of ways (e.g. photography, electron microscopy), enabling the optical appearance to be captured for later use or for sharing. When required, the optical appearance can be magnified (e.g. microscopy) and manipulated to enable the interpretation of the objects (mostly samples from an environment or experiment). Imaging data support the interpretation of other data, like experimental data. Some imaging data will be raw data (3.3), which need to be derived through image processing to enable interpretation.

7.2 Standards and metadata

Advanced imaging devices are calibrated to ensure prospering visualization.

Metadata which are provided are time of imaging, device settings and magnification/scale when appropriate. In addition, metadata will be provided about the object that is being imaged.

8 Protocols

8.1 Dataset description

A protocol is a predefined written procedural method in the design and implementation of experiments or sampling. In addition to detailed procedures and lists of required equipment and instruments, protocols often include information on safety precautions, the calculation of results and reporting standards, including statistical analysis and rules for predefining and documenting excluded data to avoid bias.

8.2 Standards and metadata

Protocols enable standardization of a laboratory method to ensure successful replication of results by others in the same laboratory or by partners' laboratories.

The MIND-project has received funding from the European Union's EUROATOM research and training program (Horizon2020) under grant agreement 661880.

Metadata for Protocols should include the purpose of the protocols, references to standards and literature.

The MIND-project has received funding from the European Union's EUROATOM research and training program (Horizon2020) under grant agreement 661880.