D8.1

BEACON
Bentonite Mechanical Evolution

Beacon Data Management Plan (DMP)

DELIVERABLE D8.1
Dissemination type: Other

Editor: Mary Westermark, SKB

Reporting period: 01/06/2017 – 30/11/2018
Date of issue: **09/05/2018**

Start date of project: **01/06/2017** Duration: 48 Months

This project receives funding from the Euratom research and training programme 2014-2018 under grant agreement No 745 942		
Dissemination Level		
PU	**Public**	**X**

REVIEW

Name	Internal /Project internal/ External	Comments
Patrik Sellin	Project internal	
Distributed to all partners before submission	Project internal	In accordance with Beacon Consortium Agreement, section 4.2

DISTRIBUTION LIST

Name	Number of copies	Comments
Athanasios Petridis (EC) Christophe Davies (EC) Beacon partners	Digitally	

Table of contents

1. Introduction and background

The main focus of the Beacon Project is modelling, and mainly existing data from ongoing or previous experiments will be (re)used. This data has been assembled and evaluated in work package 2 and will be used in work packages 3 and 5. The database with the assembled data is a part of Beacon deliverable D2.2.

It is envisaged that there will be need to complement existing data and therefore Beacon has a work package (WP4) for experimental work. The experiments performed in the project will be

- complementary where the existing data is not sufficient
- only laboratory scale

This Data Management Plan concerns both experimental data, models and input and output data.

Good research data management is a key conduit leading to knowledge discovery and innovation, and to subsequent data and knowledge integration and reuse. Instructions from the Commission regarding data management can be found here http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm and here http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

2. Purpose

The purpose with this plan for the management of Beacon Project data is to support traceability, availability and quality assured handling of the data produced within the frames of the Beacon project, and to the extent it is possible, realistic and relevant it should support in making the research data in the Beacon project "findable, accessible, interoperable and reusable (**FAIR**)" as it is formulated in the H2020 online manual.

3. Delimitations

This DMP does not concern the data assembled in WP2 from other projects for reuse in the project. This reused data is documented in an excel database, published together with the Beacon deliverable D2.2.

4. Summary of the guidelines

→ Data and models produced, generated and/or developed in Beacon have to be saved, stored and backed up in a safe way for at least 5 years after the Beacon project is finalised.

→ The Commission requires that data is kept "Findable, Accessible, Interoperable and Reusable (**FAIR**)", to the extent it is possible, during and after the project. (Explanations to these concepts available further down)

→ Partners should normally use their own procedures and management systems to manage data and models.

→ If such procedures are lacking use the below guidelines.

→ 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).

5. Guidelines

5.1 General

Data from experiments, models, the input data used and output data generated in the Beacon project need to be stored, findable, accessible, and as far as possible interoperable and reusable (FAIR) for at least 5 years after the Beacon project is finalised.

Data types in Appendix 3.

The **participants that have a management system for handling data and information** that lives up to the relevant applicable parts of the guidelines in section 4 should follow that and inform the rest of the consortium how the data of the experiments and modelling tasks are managed through filling out the essential parts appendix 1 (modified if appropriate) and uploading it to the relevant folder in the Beacon Projectplace. The Beacon Projectplace can also be used to store project data.

The **participants that do not** have a management system for handling data and information that lives up to the relevant parts of the guidelines in section 4 will follow the below described guidelines and can use attachment 1(modified if appropriate) to describe data. Attachment 2 can be used as a template when planning experiments and modelling tasks where data will be produced.
In an experiment or modelling plan with an appropriate scope the relevant and applicable parts of the following issues should be addressed. Appendix 1 and 2 can be used as templates.
The partner(s) performing the experiments or modelling and thereby producing the data are responsible for managing the data.

Generally, the procedures normally applied within the organisation can and should be used.

5.2 Experiment data summary

State/explain/specify:

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

5.3 Modelling, Input Data and Output data

It is the responsibility of the Party performing the modelling that the models and codes used are quality assured, validated and traceably described.
Modelling files will be

- (compressed if appropriate) saved and stored in a safe and findable way
- backed up,
- together with
    - a description of which version of the model/calculation code has been used for which modelling/calculation
    - Input Data
    - Output Data
    - (if relevant: which compression tool, inc version, has been used)

5.4 FAIR data according to H2020 guidelines

(also assembled in appendix 4)
FAIR= Findable, Accessible, Interoperable and Reusable

Making the data Findable (inc. provisions for metadata (information about the data) [FAIR]

The discoverability of data should be outlined (metadata provision)
The identifiability of data should be outlined and standard identification mechanism referred to. Do you make use of persistent and unique identifiers such as **Digital Object Identifiers**?
Naming conventions used should be outlined. The approach towards search keyword should be outlined. The approach for clear versioning should be outlined. Standards for metadata (information about the data) creation (if any) should be specified.

Guidance:
The Research Data Alliance provides a Metadata Standards Directory that can be searched for discipline-specific standards and associated tools.

Making data openly Accessible [FAIR]

It should be specified which data will be made openly available, and how. If some data is kept closed provide rationale for doing so.

Guidance: Follow the principle "**as open as possible, as closed as necessary**". The Commission recognises that there are good reasons to keep some or even all research data generated in a project closed. Where data need to be shared under restrictions, explain why, clearly separating legal and contractual reasons from voluntary restrictions. It is possible for specific beneficiaries to keep their data closed if relevant provisions are made.

The Registry of Research Data Repositories provides a useful listing of repositories that you can search to find a place of deposit. If you plan to deposit in a repository, it is useful to explore appropriate arrangements with the identified repository in advance. 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 (information about the data), documentation and code are deposited.
Specify how access will be provided in case there are any restrictions.
For example is there a need for a data access committee.

Making data Interoperable [FAIR]

Assess the interoperability of your data. Specify what data and metadata (information about the data) vocabularies, standards or methodologies you will follow to facilitate interoperability.

Guidance:
Interoperability means allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (i.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins.
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?

Increase data Re-use (through clarifying licenses) [FAIR]

Specify how the data will be licenced (**if applicable**) to permit the widest reuse possible. Specify whether the data produced and/or used is useable by third parties, in particular after the end of the project. If the re-use of some data is restricted, explain why. Specify the length of time for which the data will remain re-usable.

Guidance:
The EUDAT B2SHARE tool includes a built-in license wizard that facilitates the selection of an adequate license for research data. Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed. Reasons for embargoes may include time to publish or seek patents. If an embargo is sought, specify why and for how long, bearing in mind that research data should be made available as soon as possible.

6. Quality assurance of data

Data in the project should be quality assured. Describe or refer to the processes used in

7. Allocation of resources for open access of data

Costs related to open access to research data are eligible as part of the Horizon 2020 grant (if compliant with the Grant Agreement conditions). Costs are eligible for reimbursement during the duration of the project under the conditions defined in the H2020 Grant Agreement, in particular Article 6 and Article 6.2.D.3 but also other articles relevant for the cost category chosen.
If applicable describe costs and potential value of long term preservation

8. Data security

Address:

- data recovery
- if applicable: secure storage and transfer of sensitive data

Open data can be stored in Beacon Projectplace. Regarding sensitive data; bring the issue to the Executive board for discussion. If necessary it should be safely stored in certified repositories for long term preservation and curation.

9. Ethical aspects

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.

Guidance:
Consider whether there are any ethical or legal issues than can have an impact on data sharing. For example, is informed consent for data sharing and long term preservation included in questionnaires dealing with personal data?

10. National/funder/sectorial/departmental procedures

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

APPENDIX 1: Data management description template

Upload to the Beacon Projectplace

Metadata (information about the data)	Description
Dataset reference and/or name	*Unique name identifying the dataset.* *Identifier should start with EU-Beacon-WPx where "x" is the relevant work package number followed by a three digit number. Example: "EU-Beacon-WP4-001"*
Datatype	*Choose one or more of the relevant data types: Experimental Data, Observational Data, Raw Data, Derived Data, Physical Data (samples), Models, Images, Protocols, Input Data and Output Data. Data types are further described in Appendix 3.*
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 information about the data	*Provides a brief description of the relevant standards used and list relevant metadata (information about the data) in accordance with the description in Appendix 3.* *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 Beacon 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.*

APPENDIX 2: Data set template

Metadata (information about the data)
Dataset reference and/or name
Source
Dataset description
Standards and metadata (information about the data)
Data sharing
Archiving and preservation

APPENDIX 3: Data types

1 Experimental Data

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.

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.

Standards and information about the data
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.

Summary information about the data (metadata) 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

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

Standards and information about the data
The information (metadata) regarding 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

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

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.

Standards and information about the data
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 (information about the data) should at least include standards, techniques and devices used. It 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

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.

Standards and information about the data
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.

Information about the data (metadata) should include any standard or method or best practice used in the analysis. It 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)

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.

Standards and information about the data
When sampling an environment or experiment, blank samples are taken as well, as a reference. Information that should be included about the data (metadata) are 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 Images

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.

Standards and information about the data
Advanced imaging devices are calibrated to ensure prospering visualization.

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

7 Protocols

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.

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

Information (metadata) that should be included for Protocols are the purpose of the protocols, references to standards and literature.

8 Models, Input Data and Output Data

It is the responsibility of the Party performing the modelling that the models and codes used are quality assured, validated and traceably described.
Modelling files will be zipped and saved, stored in a safe and findable way, and backed up, together with a description of which version of the model/calculation code has been used for which modelling/calculation, as well as Input Data and Output Data.

APPENDIX 4:

FAIR Data Management at a glance: issues to cover in your Horizon 2020 DMP

This table provides a summary of the Data Management Plan (DMP) issues to be addressed, as outlined above.

DMP component	Issues to be addressed, can be very short. State N/A when relevant
1. Data summary	• 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
2. FAIR Data 2.1. Making data findable, including provisions for metadata (information about the data)	• Outline the discoverability of data (metadata (information about the data) 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 (information about the data) creation (if any). If there are no standards in your discipline describe what type of metadata (information about the data) will be created and how
2.2 Making data openly accessible	• 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 (information about the data), documentation and code are deposited • Specify how access will be provided in case there are any restrictions

2.3. Making data interoperable	• Assess the interoperability of your data. Specify what data and metadata (information about the data) 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?
2.4. Increase data re-use (through clarifying licences)	• 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
3. Allocation of resources	• Estimate the costs for making your data FAIR. Describe how you intend to cover these costs • Clearly identify responsibilities for data management in your experiment • Describe costs and potential value of long term preservation
4. Data security	• Address data recovery as well as secure storage and transfer of sensitive data
5. Ethical aspects	• 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
6. Other	• Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)