
# D6.1 – Data Management Plan

| | |
|---|---|
| Deliverable ID | D6.1 |
| Deliverable Title | Data Management Plan |
| Work Package | WP6 |
| | |
| Dissemination Level | PUBLIC |
| | |
| Version | 1.0 |
| Date | 2018-06-18 |
| Status | Final |
| | |
| Lead Editor | ISMB |
| Main Contributors | Xu Tao (ISMB), Enrico Ferrera (ISMB) |

**Published by the BRAIN-IoT Consortium**

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 780089.

## Document History

| Version | Date | Author(s) | Description |
|---|---|---|---|
| 0.1 | 2018-05-02 | Xu Tao (ISMB) | First Draft with TOC and initial contents |
| 0.2 | 2018-06-04 | Enrico Ferrera (ISMB) | Document reorganization and finalization. |
| 0.3 | 2018-06-10 | Enrico Ferrera (ISMB) | Document modified according to the internal reviews. Addes new section on "Data Management and the GDPR". |
| 1.0 | 2018-06-18 | Enrico Ferrera (ISMB) | Document ready to be submitted. |

## Review History

| Version | Review Date | Reviewer | Summary of Comments |
|---|---|---|---|
| 0.2 | 2018-06-06 | Etienne Gandrille (CEA) | Major comments. Sections to be reviewed. |
| 0.2 | 2018-06-06 | Miquel Cantero (ROB) | Approved with minor comments. |
| 0.3 | 2018-06-12 | Miquel Cantero (ROB) | Approved. |
| 0.3 | 2018-06-15 | Etienne Gandrille (CEA) | Approved. |

## Table of Contents

## Executive summary

This document provides the description of what data BRAIN-IoT project will generate, how it will be stored and managed and how it will be preserved after the end of the project. ISMB as main responsible of this deliverable, prepared this document with inputs and contributions from other Consortium partners and with the review from CEA and Robotnik.

This Data Management Plan complies with H2020 requirements and is based on the Data Management Plan guidelines [1]. The plan will be updated whenever changes to the project are made due to inclusion of new data sets, changes in consortium policies or other external factors.

# 1 Introduction

The purpose of this document is to present the initial Data Management Plan (DMP) of the BRAIN-IoT project and to provide the guidelines for maintaining the DMP during the project.

The Data Management Plan methodology approach adopted for the compilation of D6.1 has been based on the updated version of the "Guidelines on FAIR Data Management in Horizon 2020 version 3.0 released on 26 July 2016 by the European Commission Directorate – General for Research & Innovation" [1]. It defines how data in general and research data in particular will be handled during the research project and will make suggestions for the after-project time. It describes what data will be collected, processed or generated within the scope of the project, what methodologies and standards shall be followed during the collection process, whether and how these data shall be shared and/or made open for the evaluation needs, and how they shall be curated and preserved.

All BRAIN-IoT data will be handled according to EU Data protection and Privacy regulation and the General Data Protection Regulation (GDPR) [2].
The BRAIN-IoT DMP addresses the following issues:

- Data Summary
- FAIR data
  - Making data findable, including provisions for metadata
  - Making data openly accessible
  - Making data interoperable
  - Increase data re-use
- Allocation of resources
- Data security
- Ethical aspects
- Other issues

According to EU's guidelines regarding the DMP, the document will be updated – whenever the Project Board considers it necessary to be updated - during the project lifetime (in the form of deliverables).

BRAIN-IoT will be deployed in two pilot sites in Coruna and Valencia, Spain, with the aim to be replicated in several other places and domains. More specifically, BRAIN-IoT is envisaged to be evaluated also within the context of use cases identified in running Large Scale Pilot (LSP).
Currently (M5 of the project), the exact definition, deployment and usage of BRAIN-IoT functionalities are not yet completely defined. Therefore, we will need to update the DMP with the data that is being collected/created at each pilot site according to their usage and whether they can be published as Open Data.

## 1.1 Scope

This document is generated by WP6 "Test, Demonstration and Evaluation", and more specifically by task T6.1 "Integration and Lab-scale Evaluation".
The scope of the DMP is to describe the data management life cycle for all data sets to be collected, processed or generated in all Work Packages during the 36 months of the Brain-IoT project. FAIR Data Management is highly promoted by the Commission and since Brain-IoT deals with several kinds of data, relevant attention has been given to this task.
However, the Data Management Plan is going to be updated throughout the course of the project and more specifically, extended information on data and data management will be included in the upcoming deliverables D6.3 – "*Phase 1 Integration and Evaluation Framework*", due on M16, and D6.5 - "*Phase 2 Integration and Evaluation Framework*", due on M28.

## 1.2 Methodology

The DMP [1] concerns all the data sets that will be collected, processed and/or generated, shared, and deleted when not needed anymore, within the project.
The methodology the consortium follows to create and maintain the project DMP is hereafter outlined:

1. Create a data management policy.
    a. Using the elements that the EC guidelines [1] proposes to address for each data set.
    b. Adding the strategy that the consortium uses to address each of the elements.
2. Create a DMP template that will be used in the project for each of the collected data sets, see Section 5 - DMP dataset description template.
3. Creating and maintaining DMPs
    a. If a data set is collected, processed and/or generated within a work package, a DMP should be filled in. For instance, training data sets, example collections etc.
    b. For each of the pilots, when it is known which data will be collected, the DMP for that pilot should be filled in.
4. The filled DMPs should be added to the upcoming D6.3 and D6.5, describing which data are collected within the project as well as how it is managed.
5. Towards the end of the project, an assessment will be made about which data is valuable to be kept as Open Data after the end of the project.
    a. For the data that is considered to be valuable an assessment of how the data can be maintained and the cost involved will be made. The Consortium will also evaluate the possibility to share data, or a subset, under an Open Data Commons Open Database License (ODbL).

The deliverable is organized as following:
**Chapter 2** outlines a data overview in the BRAIN-IoT project. It details BRAIN-IoT data categories, data types and metadata.
**Chapter 3** outlines the data management policy in BRAIN-IoT about dataset naming and collection, giving also an insight about the Open Research Data Pilot under H2020 guidelines and FAIR Data principle, as well as how to achieve it.
**Chapter 4** presents the identified approach to be used in order to describe the set of data generated and collected by the project.

## 1.3 Related documents

| ID | Title | Reference | Version | Date |
|---|---|---|---|---|
| DoA | Description of Action/ Grant Agreement | ISMB | 1.0 | 2017-10-09 |
| D1.1 | Project Handbook, Quality & Risk Management Plan | IM | 1.1 | 2018-02-19 |
| D2.1 | Initial Visions, Scenarios and Use Cases | UGA | 1.0 | 2018-06-23 |

# 2 Data Management and the GDPR

The EU General Data Protection Regulation (GDPR) brings revolutionary changes to European data protection laws. Some principles found from the GDPR are defined to correspond to both the technological developments happened in recent years, and to better answer the requirements for privacy protection in the digitized world of today and tomorrow.
The principles relating to the personal data management are set out in GDPR's Article 5(1)

- Lawfulness, fairness and transparency: personal data shall be processed lawfully, fairly and in a transparent manner in relation to the data subject. To be more transparent while managing and processing data, making privacy policies more user friendly and promoting the rights of users could be considered.
- Purpose limitation: personal data shall be collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes.
- Data minimization: personal data shall be adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed. Considering the purpose, only necessary data is managed and processed. Data minimization is strongly related to purpose limitation, since enough data should be collected to achieve the purpose, but only the strictly amount needed.
- Accuracy: personal data shall be accurate and, where necessary, kept up to date. The erasure or rectification of inaccurate personal data must be implemented without delay.
- Storage limitation: personal data shall be kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed.
- Integrity and confidentiality: personal data shall be processed in a manner that ensures appropriate security of the personal data, including protection against unauthorized or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organizational measures.

Article 5(2) provides for the perhaps most important principle of all: the principle of accountability, which sets an obligation on data controllers to be responsible for and to be able to demonstrate compliance with the GDPR. It complements the GDPR's transparency requirements; data controllers must not only comply with the GDPR but must also be able to demonstrate it by e.g. documenting their decisions while managing and processing data.

In May 2018 the GDPR has been officially released. This means all partners within the consortium have to follow the new rules and principles. The novelty of the new regulamentation implies the consortium tools and partner specific guidelines for data management are not yet fully available.
This chapter addresses how the founding principles of the GDPR will be followed in the BRAIN-IoT project.

## 2.1 Lawfulness, fairness and transparency

BRAIN-IoT project describes all handling of personal data in its Data Management Plan. Some of the answers requested cannot be provided at the moment of writing this report. Therefore, updates to the plan will be provided in the next deliverables. Meanwhile, the project Wiki tool (see D1.1 – "*Project Handbook, Quality & Risk Management Plan*"), used as a logger of all the ongoing activities related to the project, will be used as a working tool to log also information related to DMP and will be updated accordingly as soon as new information about data sets become available. The collected information will be lately afterwards officially reported within upcoming deliverables D6.3 – "*Phase 1 Integration and Evaluation Framework*", due on M16, and D6.5 - "*Phase 2 Integration and Evaluation Framework"*, due on M28.
All data gathering from individuals will require informed consent of the subjects who are engaged in the project. Informed consent requests will consist of an information letter and a consent form. This will state the specific causes for the experiment, or other activity, how the data will be handled, stored, and shared. The request will also inform the subjects of their rights to have data updated or removed, and the project's policies on how these rights are managed.

As far as possible, BRAIN-IoT project will anonymise the personal data. Whenever considered necessary, further consent will be asked to use the data for open research purposes, this includes presentations at conferences, publications in journals as well as depositing a data set in an open repository at the end of the project.
The consortium tries to be as transparent as possible in their collection of personal data: while collecting data, information leaflet and consent form will describe the kind of information, the manner in which it will be collected and processed, if, how, and for which purpose it will be disseminated and if and how it will be made open access. Finally, the subjects will have the possibility to request what kind of information has been stored about them and they can request to be removed from the results.

### 2.2 Purpose limitation

BRAIN-IoT project will not collect any data that is outside the scope of the project. Each partner will only collect data which is needed within the scope of their specific work package.

### 2.3 Data minimisation

BRAIN-IoT will collect only data that is relevant for the project's research questions and demonstration. However, while testing the system in an environment including the interaction with human beings, it could be possible to collect indirect data related to the personal behaviours of the involved individuals. Since this data can be highly personal, it will be treated according to all guidelines on personal data and won't be shared without anonymization or explicit consent of the involved persons.

### 2.4 Accuracy

All data collected will be checked for consistency.
Since all data is gathered within a specific timeframe, we chose not to keep the data up to date, since it would hinder our research. However, we will try to capture the data as accurately as possible, for example "warehouse map" could be stored as "warehouse map in June 2018". This will remove the necessity of keeping this information up to date.

### 2.5 Storage limitation

All data that will no longer be used for research purposes will be deleted as soon as possible. All personal data or data that can be reconducted to personal information or behaviours will be made anonymous as soon as possible. At the end of the project, if the data has been anonymised, the data set could be considered to be released as open dataset. If data cannot be made anonymous, it will be pseudonymised as much as possible and stored according the archiving rules of the partner institutions who was responsible for the management of the specific data to be stored.

### 2.6 Integrity and confidentiality

All personal data will be handled with appropriate security measures applied. Each partner who is responsible for the management of specific data will store or share data through means and channels that comply with the GDPR.

### 2.7 Accountability

Within the scope of the project, the project and quality management is responsible for the correct data management within the project. Whether the partners follow the GDPR principles will be regularly checked during the project the project lifetime. For each data set, a responsible person has been appointed at partner level, who will be held accountable for the specific data set.

# 3 Data in BRAIN-IoT: an Overview

BRAIN-IoT project will deal with a large amount of raw data to measure the benefit of IoT and federation of IoT platforms within the two selected scenarios, i.e. Service Robotics and Water Critical Infrastructure Management, and also in other scenarios to be selected from the ones identified in Large Scale Pilot (LSP) projects, i.e. AUTOPILOT, MONICA, ACTIVAGE, IoF2020 and SynchroniCity.
From raw data, a large amount of derived data can be produced to address multiple research needs and enable Smart Behaviours. Some processing, such as cleaning, verification, conversion, aggregation, summarization or reduction could also be applied to raw data according to specific needs derived from the use cases.
In any case, data must be well documented in order to facilitate and foster sharing, to enable validity assessments and to enable its usage in an efficient way.
Thus, each data must be described using additional information called metadata. The latter must provide information about the data source, the data transformation and the conditions in which the data has been produced.

## 3.1 Data sets Categories

The BRAIN-IoT project will produce different categories of data sets:

- Context data: data that describe the context of an experiment.
- Acquired and derived data: data that contain all the collected information related to an experiment.
- Aggregated data: data summary obtained by reduction of acquired data and generally used for data analysis.

### 3.1.1 Context Data

Context data is any information that helps to explain observation during a measurement campaign. Context data can be collected, generated or retrieved from existing data. For example, it contains information such as presence of humans or presence of obstacles in the robot-path, quality of water, etc.

### 3.1.2 Acquired and Derived Data

Acquired data are all data collected during the course of the study for the sole purpose of the analysis.. Derived data is created by different type of transformation including data fusion, filtering, and classification. Derived data are usually required according to specific needs from the use cases. Derived data may contain derived measures and performance indicators referring to a time period when specific conditions are met. This category includes measures from sensors coming from robotic platforms or IoT devices and subjective data collected from either the users or the environment.

The following list outlines the data types and sources that will be collected:

**Service Robotics:**

The service robotics scenario identified three different interactions of robots with external world (see D2.1):

- Robot-thing interaction (e.g. robot needs of crossing door or using lifts, interactions with conveyor belts etc.)
- Robot-environment interaction in order to have environment context information (e.g. alarm system failure/errors, obstacles or humans in the way, detection of beacons etc.)
- Robot-robot interaction to enable self-organization and collaborative features (such as map generation and shared resources)

These use cases allow to roughly outline an initial set of types of data to be dealt with:

- robots involved in the scenario will be endowed with capabilities to scan and navigate the entire area of the warehouse and share the acquired information to update the knowledge base (e.g., map) on the go.
- robots will also be used for collecting additional information implicitly e.g. room temperature, presence of humans, also paying special attention to their privacy and any image recorded during operation, detection of in-path obstacles, other IoT devices etc.
- robots can also collect context information such as the presence of alarm system, layout, number of items to interact with, loads per day, which may also be an indicator of the performance, productivity of the factory , detection of beacons etc.

**Critical Water Management Infrastructure scenario:**

Currently a part of the systems that helps us develop the business processes are implemented in a platform called SICA. Relevant data for this scenario concern the urban water domain.
More specifically, in D2.1 the following domains have been identified, which outline an initial set of types of data to be dealt with:

RESOURCE

- Connection of a multiparameter probe to measure the water quality control parameters in reserve water (surface waters).
- Connection with the gaugin station to measure the circulating flow of the river (entry and exists of the reserve).
- Pluviometry and temperature.

TREATMENT

- Headstock deposits levels.
  - Volume
  - Cl/pH levels
- Pump systems at the plant.
  - Pump from the headstock to treatment.
  - Pump of treated water.

DISTRIBUTION

- Distribution deposits levels.
  - Volume
  - Cl/pH/turbidity
- Pump and repump systems.
- Section control systems.
  - Flows
  - Cl/pH/turbidity
- Tele-read meters:
  - Domestic. Control sections: ABERING platform.
  - Commercial. Sectors. iEcoCity platform.
  - Large clients. Complex multi-sensor remote systems.
- Control of green zone irrigation.
  - Irrigation programme. Connection with automatons.
  - Meteorological control: rain forecast.

### 3.1.3 Aggregated data

Aggregated data contains a specific part of the acquired or derived data (raw data). Its smaller size allows a simple storage in e.g. database tables and an easy usage suitable for data analysis. To obtain aggregated data, several data reduction processes are performed. The reduction process summarizes the most important aspects in the data into a list of relevant parameters or events, through one or all of the following processes: validation, curation, conversion, annotation. Aggregated data is generally created in order to answer different research question. They are supposed to be verified and cleaned, thus facilitating their usage for analysis purposes.

## 3.2 Metadata

This section provides the first recommendations regarding the description of the data provided by BRAIN-IoT project. As the project will collect several data categories and several data types, several metadata descriptions must be provided to describe the characteristics of each measure or component and also the origin on how the data was produced and collected.

BRAIN-IoT project will follow and adapt the metadata type recommendation provided by the FOT-Net Data project (http://fot-net.eu/ ). This project identifies in its Data Sharing Framework several metadata types that can be applied to BRAIN-IoT. The following list provides a first version of the metadata that may be managed by the project and their content in the upcoming months. A more detailed version will be provided in the upcoming deliverables D6.3 – "*Phase 1 Integration and Evaluation Framework*", due on M16, and D6.5 - "*Phase 2 Integration and Evaluation Framework"*, due on M28.

### 3.2.1 Metadata attributes of time-history data:

Time-history data corresponds to the history of a measurement over the time. Time-history data can be collected by legacy instrument, by IoT devices or IoT platforms.

Time-history data stores a variation over the time of single or complex physical value. To enable their re-use, each dataset provides a metadata description that includes the following descriptive attributes:

- Precision (accuracy)
- Unit of measure
- Sample rate (frequency of the measure)
- Filtering (low-pass, interpolation, etc.)
- Origin (data source)
- Type (Integer, Float, String)
- Error codes (full description of error codes)
- Quality (Quality measure related to this measure)
- Enumeration specification (Defines how to convert constant to correct value, e.g.: 1 means Left, 2 means Right)
-

### 3.2.2 Metadata attributes of aggregated data

As aggregated data varies depending on the purpose of the experiment, it can be described as time history measures or as time segment. Time segment is a sub-set of data parameters or measures generated by data summarization or data reduction.

This metadata type should include the following descriptive attributes:

- Description (Purpose of the aggregated data)
- Definition (Algorithm applied on the aggregated measures)

- Origin (Measures used to calculate the aggregated data)
- Unit (Unit of output value)

### 3.2.3 Metadata attributes of self-reported data

Self-reported data corresponds to interviews, surveys or questionnaires. This metadata type should include the following descriptive attributes:

- Description (Purpose of the questionnaire)
- Instructions (way how the collection process was executed)
- Type (Free text, single or multiple choices, etc.)
- Options (description of possible alternatives)

# 4 BRAIN-IoT Data Management Policy

The responsible party for creating and maintaining the DMP for a data set is the partner that creates/collects such data. If a data set is collected, processed and/or generated within a work package, a DMP should be created.
Before each pilot execution, it should be clear which data set is collected/created in the pilot and how the data will be managed, i.e. the DMPs for the pilot data must be ready and accepted. This will be done individually for each of the pilots because of the difference between the pilots being in different domains and of different types of data and events.

## 4.1 Naming and identification of the Data set

To have a mechanism for easily identifying the different collected/generated data, we will use a naming scheme. The naming scheme for BRAIN-IoT datasets will be a simple hierarchical scheme including country, pilot, creating or collecting partner and a describing data set name. This name should be used as the identification of the data set when it is published as Open Data in different open data portals. The structure of the naming of the dataset will be as follows:

BRAINIOT_{Country+Area Code or WP}_{Pilot Site or WP}_{Responsible Partner}_{Description}_{Data Set Sub Index}

**Figure 1: BRAIN-IoT Data Set Naming Scheme**

The parts are defined as follows:

- BRAINIOT: Static for all data sets and is used for identifying the project.
- Country+Area Code: The two letter ISO 3166-1 country code for the pilot where data has been collected or generated plus the numeric routing code that identifies each geographic area in the telephone numbering plan, e.g. ES96.
- WP: the work package label along with the work package number, e.g., WP6.
- Pilot Site: The name of the pilot site where the data was collected, without spaces with CamelCaps in case of multiple words, e.g. ServiceRobotics etc.
- Responsible Partner: The partner that is responsible for managing the collected data, i.e. creates and maintains the Data Management plan for the data set. Using the acronyms from D1.1, e.g. ISMB
- Description: Short name for the data set, without spaces with CamelCaps in case of multiple words, e.g., WarehouseMap, WaterPollution, etc.
- Data Set Sub Index: Optional numerical index starting from 1. The purpose of the dataset sub index is that data sets created/collected at different times can be distinguished and have their individual meta data.

BRAINIOT_ES96_ServiceRobotics_ROB_Warehouse_1

**Figure 2: BRAIN-IoT Data Set Naming Example**

In the example shown in Figure 2, the Data set is created within BRAIN-IoT project in Valencia city, Spain, at Service Robotics pilot site. Robotnik is responsible for the relevant Data Management plan for the dataset. The dataset contains location data and it is the first of a series of data sets collected at different times.
There can be situations where the data needs to be anonymised with regards to the location the data has been collected, for instance at some pilots it might not be allowed to publish people count data with the actual event location for security reasons. In these cases, the Country and Pilot Site will be replaced by string UNKNOWN when it is made available as Open Data.

For data sets that are not connected to a specific pilot site the Pilot Site should be replaced with the prefix WP followed by the Work Package number that creates and maintains the Data Management plan for the dataset, e.g., WP6. The same applies to the Country part which also should be replaced with the prefix WP followed by the Work Package number in the cases where the data set is not geographically dependent, such as pure simulations or statistics.

## 4.2 Data Summary / Data set description

The data collected/created needs to be described including the following information:

- 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)
    - Provide the identification of the re-used data, i.e. BRAIN-IoT identifier or pointer to external data, if possible.
- Specify the origin of the data
- State the expected data size (if known)
- Outline the data utility: to whom will it be useful

## 4.3 Fair Data

FAIR data management means in general terms, that research data should be "FAIR" (Findable, Accessible, Interoperable and Re-usable). These principles precede implementation choices and do not necessarily suggest any specific technology, standard, or implementation solution.

### 4.3.1 Making data findable, including provisions for metadata

This point addresses the following issues:

- Outline the discoverability of data (metadata provision)
- Outline the identifiability of data and refer to standard identification mechanism.
- Outline the naming conventions used.
- Outline the approach towards search keywords.
- Outline the approach for clear versioning.
- Specify standards for metadata creation (if any).

As far as the metadata are concerned, the way the consortium will capture and store information should be described. For instance, for data records stored in a database with links to each item, metadata can pinpoint their description and location.

There are various disciplinary metadata standards, however the BRAIN-IoT consortium has identified a number of available best practices and guidelines for working with Open Data, mostly by organisations or institutions that support and promote Open Data initiatives, and will be taken into account. These include:

- FOT-Net Data project
- Open Data Foundation
- Open Knowledge Foundation
- Open Government Standards

Furthermore, data should be interoperable and compliant with respect to data annotation and data exchange.

### 4.3.2 Making data openly accessible

The objectives of this aspect address the following issues:

- Specify which data will be made openly available and, in case some data is kept closed, explain the reason why.
- Specify how data will be made available.
- Will the data be added to any Open Data registries?

- Specify what methods or software tools are needed to access such data, if a documentation is necessary about the software and if it is possible to include the relevant software (e.g. in open source code).
- Specify where data and associated metadata, documentation and code are deposited.
- Data that will be considered safe in terms of privacy, and useful for release, could be made available for download under the ODbL License.
- Specify how access will be provided in case there are restrictions.

### 4.3.3 Making data interoperable

This aspect refers to the assessment of the data interoperability specifying which data and metadata vocabularies, standards or methodologies will be followed in order to facilitate interoperability. Moreover, it will address whether standard vocabulary will be used for all data types present in the data set in order to allow inter-disciplinary interoperability.
In the framework of the BRAIN-IoT project, we will deal with many different types of data coming from very different sources, but in order to promote interoperability we use of the following guidelines:

- OGC SensorThings API model for time series data [4], such as environmental readings etc.
- If the data is part of a domain with well-known open formats that are in common use, this should be selected.
- If the data does not fall in the previous categories, an open and easily machine-readable format should be selected.

### 4.3.4 Increase Data Re-use

This aspect addresses the following issues:

- Specify how the data will be licensed to permit the widest reuse possible.
    - Tool to help selecting license: https://www.europeandataportal.eu/en/content/show-license
    - If a restrictive license has been selected, explain the reasons behind it.
- Specify when data will be made available for re-use.
- Specify if the data produced and/or used in the project is useable by third parties, especially, after the end of the project.
- Provide a data quality assurance process description, if any.
- Specify the length of time for which the data will remain re-usable.

In order to maximize the reusability of data, ODbL licence could be considered in some cases as a good candidate to distribute datasets. ODbL allows to:

- to copy, distribute and use the database;
- to produce works from the database;
- to modify, transform and build upon the database;

as long as you:

- must attribute any public use of the database, or works produced from the database, in the manner specified in the ODbL. For any use or redistribution of the database, or works produced from it, you must make clear to others the license of the database and keep intact any notices on the original database.
- publicly use any adapted version of this database, or works produced from an adapted database, you must also offer that adapted database under the ODbL.
- redistribute the database, or an adapted version of it, then you may use technological measures that restrict the work (such as DRM) as long as you also redistribute a version without such measures.

## 4.4 Allocation of Resources

This aspect addresses the following issues:

- Estimate the costs for making the data FAIR and describe the method of covering these costs.

  - This includes, if applicable, the cost for anonymising data.
- Identify responsibilities for data management in the project.
- Describe costs and potential value of long-term preservation.

### 4.5 Data security and Privacy

Based on the self-assessment performed by the BRAIN-IoT consortium, no major ethics issues are foreseen to be relevant for project activities.

Nevertheless, the consortium recognizes the potential risks that the deployment of IoT technology developed in BRAIN IoT could generate. In fact, the project has a dedicated WP i.e. WP5 "End-to-end Security, Privacy and Trust Enablers" that is specifically conceived to mitigate these risks and focuses on:

- Threat Modelling and Assessment (Task 5.1);
- Decentralized Authorization, Authentication and Trust (Task 5.2);
- Privacy awareness and control (Task 5.3);
- End-to-end data security and provenance (Task 5.4).

The project consortium is committed to conducting responsible research and innovation and will respect careful experimentation methodologies whenever end users are present in experimentations:

- end users will get a complete briefing on the project, the experimentation and any potential risks as part of their training.
- the project will ensure that any end user involved understands and consent to the experiment.

In addition to the above approach that will be adopted during the project implementation and beyond, BRAIN-IoT has realized at proposal stage an ethic self-assessment of risks and identified two main points that can be of concerns:

- the involvements of end users in the experiments run on the test-sites.
- the potential collection and handling of personal data.

Such evaluation has also been performed taking into consideration possible links of BRAIN-IoT project to IoT Large Scale Pilots. However, it is worth observing that, in those cases, BRAIN-IoT will act as technology solution provider and will not take care of the data and user involvement aspects (that will remain within the scope of the Large Scale Pilots projects).

In the following, rules defined for handling data are presented. More specifically, the data collected will be treated as confidential and security processes and techniques will be applied to ensure their confidentiality. Overall the following general principles will be used regarding any data collection:

- Transparency of usage of the data: User – data subject in the European Union (EU) parlance - shall give explicit consent of usage of data.
- Collected Data shall be adequate, relevant and not excessive: The data shall be collected on "need to know" principle. This principle is also known as "Data Minimization". The principle also helps to setup the user contract, to fulfil the data storage regulation and enhance the "Trust" paradigm.
- Collector shall use data for explicit purpose: Data shall be collected for legitimate reasons and shall be deleted (or anonymize) as soon as data is no longer relevant.
- Collector shall protect data at communication level: The Integrity of the information is important because modification of received information could have serious consequence for the overall system availability. User has accepted to disclose information to a specific system, not all the systems. The required level of protection depends on the data to be protected according the cost of the protection and the consequence of data disclosure to unauthorized systems.

- Collector shall protect collected data at data storage: User has accepted to disclose information to a specific system, not all the systems. It also could be mandatory to get infrastructure certification. The required level of protection depends on the data to be protected according the cost of the protection and the consequence of data disclosure to unauthorized systems. As example, user financial information can be used to perform automatic billing. Such data shall be carefully protected. Security keys at device side and server side are very exposed and shall be properly protected against hardware attacks.
- Collector shall allow user to access / remove Personal Data: Personal Data may be considered as a property of the user. User shall be able to verify correctness of the data and ask – if necessary – correction. Dynamic Personal Data – for instance home electricity consumption – shall also be available to the user for consultation. For static user identity, this principle is simply the application of current European regulations according access to user profile.

## 4.6 Ethical aspects

Some of the most mature tests and demonstrations of the project will be run in "live" environment (city) where ordinary citizens are present.

The development of new human behaviours is an important impact of ICT that can be felt by end users as positive, neutral or negative. It is indeed a task of the project to perform a user-center evaluation of the BRAIN-IoT solution (task 6.2).

An additional task that could be possibly performed beyond the duration of the project, aims to ensure that end user involvement is done in condition as good as possible for the end users. This will involve the following activities:

- Engaging with end user only on an informed way: making sure they are aware of the presence of experiments and that relevant documentation, in understandable format (language and avoidance of technical jargon) is available.
- Gathering end user consent as a prerequisite for interaction and any data collection
- Providing a complaint procedure with a neutral third party
- Ensuring that end-users are free to refuse the experiment at any moment, including after it is started, without any prejudice or disadvantage.

## 4.7 Other issues

Other issues will refer to other national/ funder/ sectorial/ departmental procedures for data management that are used.

# 5 DMP dataset description template

During the course of the project, each work package will analyse which DMP components are relevant for its activities. When the pilots definitions will be ready with regards to which data is collected and how data is used, DMPs for the pilots need to be created.

This table is a template that shall be used to describe the datasets.

**Table 1: BRAIN-IoT Template for DMP**

| DMP Element | Issues to be addressed |
|---|---|
| **Identifier** | **Brain-IoT_WPX_TX.X_{Responsible Partner}_{Description}_{Data Set Sub Index }** |
| **Revision History** | Partner / Name / Description of change<br>ISMB / Xu Tao / Created initial DMP |
| **Dataset Description** | Each data set will have a full data description explaining the data provenance, origin and usefulness. Reference may be made to existing data that could be reused. |
| **Findability** | 1. Outline the discoverability of data (metadata provision).<br>2. Outline the identifiability of data and refer to standard identification mechanism.<br>3. Outline the naming conventions used.<br>4. Outline the approach towards search keywords.<br>5. Outline the approach for clear versioning.<br>6. Specify standards for metadata creation (if any). |
| **Accessibility** | 1. Specify which data will be made openly available? If some data is kept closed provide rationale for doing so.<br>2. Specify how the data will be made available.<br>3. 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)?<br>4. Specify where the data and associated metadata, documentation and code are deposited.<br>5. Specify how access will be provided in case there are any restrictions. |
| **Interoperability** | 1. Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.<br>2. Specify whether you will be using standard vocabulary for all data types present in your dataset, to allow inter disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies? |
| **Reusability** | 1. Specify how the data will be licenced to permit the widest reuse possible. |

| | |
|---|---|
| | 2. Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed.<br>3. 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.<br>4. Describe data quality assurance processes<br>5. Specify the length of time for which the data will remain re-usable. |
| **Data Sharing** | 1. Explanation of the sharing policies related to the data set between the next options:<br>2. **Open:** Open for public disposal<br>3. **Embargo**: It will become public when the embargo period applied by the publisher is over. In case it is categorized as embargo the end date of the embargo period must be written in DD/MM/YYYY format.<br>Restricted: Only for project internal use.<br>4. Each data set must have its distribution license.<br><br>5. Provide information about personal data and mention if the data is anonymized or not. Tell if the dataset entails personal data and how this issue is taken into account. |
| **Archiving and Preservation** | The preservation guarantee and the data storage during and after the project (for example: databases, institutional repositories, public repositories …) |

# 6 Resource allocation

Costs for establishing and maintaining the HBM4EU data repository are covered by the financial budget of BRAIN-IoT.

While the repository in itself is not maintained after the end of the project, all files stored within the BRAIN-IoT 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.

The responsibility for data management during and after the end of the project is up to the owners of the scenarios which are also the providers of the data sources and the organizations who are mainly interested to the semantic value of the data itself.

# 7 Conclusions

This deliverable provides a planning overview of the data that BRAIN-IoT project is going to deal with, together with related data processes and requirements that need to be taken into consideration.

The descriptions of the data sets will be incrementally enriched along the project lifetime. These descriptions include a detailed description, standards, methodologies, sharing and storage methods.

The Data Management Plan has been outlined within this deliverable and is going to be updated and further detailed in the upcoming deliverables D6.3 – "*Phase 1 Integration and Evaluation Framework*", due on M16, and D6.5 - "*Phase 2 Integration and Evaluation Framework"*, due on M28.

## Acronyms

| Acronym | Explanation |
|---|---|
| DMP | Data Management Plan |
| WP | Work Package |
| IoT | Internet of Things |
| WBS | Work Breakdown Structure |
| API | Application Programming Interface |
| OGC | Open Geospatial Consortium |
| ICT | Information and Communication Technology |
| FAIR data | Findable, accessible, interoperable and re-usable data |
| GDPR | General Data Protection Regulation |
| POPD | Protection of Personal Data |

# List of figures

# List of tables

## References

[1]. Guidelines on Fair Data Management in Horizon 2020, Version 3.0 26 July 2016; http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf (Accessed 03 June 2018)

[2]. Official PDF of the Regulation (EU) 2016/679 (General Data Protection Regulation), https://gdpr-info.eu/ (Accessed 03 June 2018)

[3]. 2018 reform of EU data protection rules, https://ec.europa.eu/commission/priorities/justice-and-fundamental-rights/data-protection/2018-reform-eu-data-protection-rules_en, (Accessed 03 June 2018)

[4]. Open Geospatial Consortium (OGC) SensorThings API. https://github.com/opengeospatial/sensorthings (Accessed 03 June 2018)