Title:	Document Version:
D6.2 Data Management Plan	2.3

Project Number:	Project Acronym:	Project Title:
649849	ENTROPY	Design of an Innovative Energy-Aware IT Ecosystem for Motivating Behavioural Changes Towards the Adoption of Energy Efficient Lifestyles

Contractual Delivery Date:	Actual Delivery Date:	Deliverable Type* - Security**:
29/02/2016	14/03/2016	R – PU

* Type: P – Prototype, R – Report, D – Demonstrator, O – Other
** Security Class: PU- Public, PP – Restricted to other programme participants (including the Commission), RE – Restricted to a group defined by the consortium (including the Commission), CO – Confidential, only for members of the consortium (including the Commission)

Responsible and Editor/Author:	Organization:	Contributing WP:
Cleopatra Bardaki (AUEBELTR)	AUEBELTR	WP6

Authors (organizations): Cleopatra Bardaki, Vasilis Stavrou (AUEBELTR)

Abstract:

This document includes the main elements of the data management policy adopted by the partners with regard to all the datasets that will be used during the project. In other words, the purpose of the DMP is to support the data management life cycle for all the data that will be collected, processed or generated by the project. Since the DMP is expected to mature during the project, this is a living document and more developed versions of it will be created at later stages of the project.

Keywords:

Data management, data sources, policy, meta-data, standards, data sharing, data archiving, data access, data preservation

Revision History

The following table describes the main changes done in the document since created.

Revision	Date	Description	Author (Organization)
v.0.1	04/01/2016	TOC of the document	Cleopatra Bardaki (AUEBELTR)
v.0.2	15/02/2016	First version of the document	Vasilis Stavrou (AUEBELTR)
v.0.3	17/02/2016	Major revision of the document	C. Bardaki (AUEBELTR)
v.0.4	22/02/2016	Inclusion of section 6	C. Bardaki (AUEBELTR)
v.0.5	03/03/2016	Inclusion of section 4 and minor revision of section 5	Vasilis Stavrou (AUEBELTR)
v.1.0	04/03/2016	Revision of section 3- inclusion of semantic model created by	C. Bardaki (AUEBELTR), A. Zafeiropoulos et al. (UBITECH), Fernando Terroso, Aurora Gonzalez, Antonio Skarmeta (UMU)
v.2.0	07/03/2016	Final revision of the document	C. Bardaki (AUEBELTR)
v.2.1	08/03/2016	Internal Review	Anna Fensel, Umutcan Simsek (UIBK)
v.2.1	08/03/2016	1st revision HESSO	Aristotelis Agianniotis, Mariam Barque, Dominique Genoud (HESSO)
v.2.2	10/03/2016	Revision of the document	Vasilis Stavrou (AUEBELTR)
v.2.3	11/03/2016	Final Revision of the document	Cleopatra Bardaki (AUEBELTR)

Executive Summary

This deliverable is the first version of ENTROPY's Data Management Plan (DMP). It includes the main elements foreseen in the European Guidelines for H2020 and the data management policy that will be used for all the datasets generated by the project. ENTROPY's DMP is driven by the project's three pilots. Specifically, this document describes the minimum datasets related to the three ENTROPY pilots.

Per dataset, the document presents a unified approach of the name and the description to be used. Furthermore, the standards and metadata are presented, as well as data sharing options along with archiving and preservation details.

Table of Contents

DEFINITIONS, ACRONYMS AND ABBREVIATIONS

Acronym	Title
CO	Confidential Only
DMP	Data Management Plan
DoW	Description of Work
Dx	Deliverable (where x defines the deliverable identification number e.g. D1.1)
PP	Program Participants
PU	Public Usage
R	Report

1. INTRODUCTION

The main objective of this deliverable is to provide the data management policy with regard to the data sources that the ENTROPY project collects, processes, generates and makes available. This deliverable is a 'living document' which will be further developed within the course of the project so as to define the research data that will be collected and specify how these data collections will be processed and managed in accordance with relevant standards and methodologies. Finally, it will also incorporate the decisions for data sharing and archiving.

Document Composition

This document is composed of the following six (6) chapters, each one covering a specific area as specified by the EC guidelines in its Data Management Plan template:

- Chapter 1: Introduction
- Chapter 2: Initial naming of the datasets
- Chapter 3: Description of the minimum datasets to be collected for each pilot
- Chapter 4: Standards and metadata
- Chapter 5: Data access and sharing mechanisms
- Chapter 6: Archiving and preservation of the data

2. DATASETS: REFERENCE AND NAME

ENTROPY project is still identifying a set of heterogeneous data sources through a series of interviews with project end-users. Following an iterative process, data sources are being established, encapsulating that way the project end-user requirements. A refinement process will continuously take place throughout the lifetime of the project, as new data sources are becoming available to the consortium.

ENTROPY project is driven by three different pilots in three different sites:

1) Pilot A: Navacchio Technology Park (NTP)

2) Pilot B: University of Murcia Campus (UMU)

3) Pilot C: Technopole in Sierre (HES-SO)

The teams working under these pilots have formed users' interaction scenarios with the system, which include data for overall assessment and behavioural analysis, providing a minimum set of data.

The data sets required by each pilot differ to each other, since the pilots have different types of data sources. However, similar naming methodology will be followed. The partners will receive one or more files (.xls/.csv) containing data. The name of the file should follow a specific structure, such as: PL_DS_FT_ND_V_D.

- PL: PiLot, the name of the pilot, the first letters (three max) of the pilot's responsible partner (UM for UMU, NT for NTP, HE for HES-SO)
- DS: DataSet, the set of data related to the pilot. It may take the value "ALL", if the file contains all the sets of data.
- FT: FormaT, the format of the file of the data
- ND: The name of the original document
- V: The version of the document
- D: The date of receiving the document or the date of creating this document (dd-mm-yyyy).

The respective folders used may follow similar structure: PL_FT_ND_D

- PL: PiLot, the first letters (three max) of the pilot's responsible partner (UMU, NTP, HES-SO)
- FT: FormaT, the format of the file of the data
- ND: The name of the original document
- D: The reception or creation date of the document.

Additionally, some of the pilots could expose their datasets by means of public web services to make the data access easier and more reliable.

3. DATASET DESCRIPTION

For each of these target groups, several parameters have been identified by each pilot. The data to be received by external sources will fill in these parameters. The pilots have identified some identical parameters, however some differ. For Pilot 1 and Pilot 2, there are five basic types of parameters, which are also shared by Pilot 3:

- Demographics: Data concerning the demographics of the users
- Building Data: Data describing the buildings' characteristics
- Psychographics: Data concerning the personality of the users
- Room Sensor Data: Data concerning the sensors' technical characteristics and measurements per room
- Building Sensor Data: Data concerning energy consumption in the building

Table 1 Variables corresponding to the common parameter types of all the three pilots

Parameter	Type	Unit	Mandatory
DEMOGRAPHICS			
User ID	String	-	N(NO)
Age	Numeric	Years	N
Gender	String	-	N
Ethnicity	String	-	N
Function (ex. Manager, professor, student etc.)	String	-	N
Educational level	String	-	N
Studies	String	-	N
Hours at university/campus	Numeric	hours	N
Kids at home (yes/no)	Boolean	-	N
Working hours	Numeric	hours	N
Reported health issues	String	-	N
BUILDING DATA			
Date	String	dd-mm-yyyy HH:mm:ss	
Building ID	String	-	Y(YES)
Construction year	Numeric		Y
Building type	String	-	Y
Building size	Numeric	m (meters)	Y
Windows percentage	Numeric		Y
Building regulations	String	-	Y
Consumption baseline	Numeric	kWh	Y
Sensor ID (link with sensor data)	String	-	Y
Total number of sensors	Numeric	-	Y
Internal temperature	Numeric	°C	Y
Internal humidity level	Numeric	%	Y
Occupants per room/building	Numeric		N
PSYCHOGRAPHICS			
Personality test	String	-	N
Curtailment behaviour	String	-	N
“Hassle factor”	String	-	N

Comfort level	String	-	N
The impact of incentives (Questionnaire)	String	-	N
Interest in energy renewable sources	String	-	N
Intrinsic interest in efficiency (Questionnaire)	String	-	N
ROOM SENSOR DATA			
HVAC			
Sensor ID	String	-	Y
Location	String	-	Y
Automated system (Yes/No)	Boolean	-	Y
State (ON/FF)	Boolean	-	Y
Operation mode (heating/cooling)	String	-	Y
Fan speed	String	-	Y
Nominal power	String	-	Y
Energy efficiency label	String	-	Y
Energy (Electricity, Gas, Fuel oil)	String	-	Y
Energy Meter			
Date (timestamp)	Date	dd-mm-yyyy HH:mm:ss	Y
Meter ID	String	-	Y
Energy consumption	Numeric	KWh	Y
Energy from renewable sources	String	-	Y
Type of energy source	String		Y
Building/Room ID (link with building/room data)	String	-	Y
Indoor Lighting System Management/Luminosity Sensors			
Sensor ID	String	-	Y
Location	String	-	Y
Automated system (Yes/No)	Boolean	-	Y
Light status (ON/OFF)	Boolean	-	Y
Light regulation (0-100%)	Numeric	%	Y
Hours of lighting per day	Numeric	Hours	Y
Type of lighting (ex. CFL, LED etc.)	String	-	Y
Number of lights on	Numeric	-	Y
Luminous flux	Numeric	lm(lumen)	Y
Nominal power	Numeric	W	Y

Humidity Sensors			
Sensor ID	String	-	Y
Location	String	-	Y
Humidity level (internal)	Numeric	%	Y
Presence sensor			
Sensor ID	String	-	Y
Location	String	-	Y
Number of attendees	Numeric	-	N
User ID	String	-	N
Enter timestamp	Date	-	N
Exit timestamp	Date	-	N
BUILDING SENSOR DATA			
Energy Meter			
Date (timestamp)	String	dd-mm-yyyy HH:mm:ss	Y
Meter ID	String	-	Y
Energy consumption (KWh)	Numeric	-	Y
Electrical consumption (Active and reactive power)	Numeric	-	Y
Energy from renewable sources	String	-	Y
Type of energy source	String	-	Y
Water Meter			
Meter ID	String	-	Y
Water meter type (Mass/Volumetric)	Boolean		Y
Water consumption	Numeric	-	Y
Environmental conditions monitoring (Weather station)			
Weather station ID	String	-	Y
Temperature (external)	Numeric	°C	Y
Barometric pressure	Numeric	Hpa	Y
Humidity (external)	Numeric	%	Y
Wind speed	Numeric	$m.s^{-1}$	Y
Wind direction	Numeric	°	Y
Precipitation	String	Mm	Y
Outside sun duration (luminosity)	Numeric	h/day (hours per day)	Y
Outside radiation	Numeric	$W/m^2/day$	N

		(daily radiation average)	

Further, Pilot 3 concerns three more types of parameters:

- Environment: Data concerning environmental characteristics of the building
- Energy: Data describing the building's characteristics per energy type used
- Price: Data concerning the price and its updates for gas and fuel oil, along with the consumption characteristics of the city

Table 2 Variables corresponding to three more parameter types of Pilot 3

Parameter	Type	Unit	Mandatory
ENVIRONMENT			
User ID	String	-	Y
Age	Numeric	Years	Y
Number of people at work	Numeric	-	Y
Week-end work	Numeric	-	Y
Planning production	String	-	Y
Studied surface	String	m^2	Y
Location	String	-	Y
Mountain mask/Building environment	String	-	Y
ENERGY			
Heating System			
Technology	String	-	Y
Nominal power	Numeric	W	Y
Output/energy efficiency	Numeric	%	Y
Energy (Electricity, Gas, Fuel oil)	String	-	Y
Energy (Electricity, Gas, Fuel oil)	Numeric	kWh	
Local energy control command	String	-	Y
Devices system			
Clusters definition (Cold, Heating,…)	String	-	Y
Nominal power	Numeric	-	Y
Hours at use per day (or per period)	Numeric	-	Y
Number	Numeric	-	Y
Lighting System			

Type of lighting (ex. CFL, LED etc.)	String	-	Y
Nominal power	Numeric	W	Y
Luminous flux	Numeric	lm	Y
Number	Numeric	-	Y
Hot Water			
Technology	String	-	Y
Nominal power	Numeric	W	Y
Output/energy efficiency	Numeric	%	Y
Energy (Electricity, Gas, Fuel oil)	String	-	Y
Energy (Electricity, Gas, Fuel oil)	Numeric	kWh	N
Ventilation			
Technology	String		Y
Nominal power	Numeric		Y
Output/energy efficiency	Numeric		Y
PRICE			
Gas			
Taxes percentage	Numeric		Y
Price evolution year 2015 - n	Numeric		Y
Fuel oil			
Taxes percent	Numeric		Y
Price evolution year 2015 - n	Numeric		Y
Electricity			
Name of supplier	String	-	Y
Taxes percentage	Numeric		Y
Power peak value for price shifting	Numeric	W	Y
Power peak percent	Numeric		Y
Reactive power percentage	Numeric		Y
Mean power peak percentage	Numeric		Y
Price evolution year 2015 - n	Numeric		Y

4. STANDARDS AND METADATA

ENTROPY project is related to different pillars, e.g., green energy, environment, etc. Several existing standards, addressing interoperability, adaptability and dynamicity issues of data on each of these specific fields. This section presents the required standards, along with the methodologies and technical documents that will be taken into account in order for the project to produce aligned data structures and data services.

4.1 ISO/TR 16344:2012

ISO/TR 16344:2012 provides a coherent set of terms, definitions and symbols for concepts and physical quantities related to the overall energy performance of buildings and their components, including definitions of system boundaries, to be used in all standards elaborated within ISO on energy performance of buildings.

These terms and definitions are applicable to energy calculations in accordance with the Technical Report and standards on the overall energy performance of buildings and their components, to provide input to the Technical Report or using output from the Technical Report. They are based on existing terms and definitions from standards and other documents referenced in the bibliography.

4.2 ISO 16346:2013

ISO 16346:2013 defines the general procedures to assess the energy performance of buildings, including technical building systems, and defines the different types of ratings, and the building boundaries. The purpose of ISO 16346:2013 is to (a) collate results from other international standards that calculate energy use for specific services within a building, (b) account for energy generated in the building, some of which may be exported for use elsewhere, (c) present a summary of the overall energy use of the building in tabular form, (d) provide energy ratings based on primary energy, carbon dioxide emission, or other parameters defined by a national energy policy, and (e) establish general principles for the calculation of primary energy factors and carbon dioxide emission coefficients.

ISO 16346:2013 defines the energy services to be taken into account for setting energy performance ratings for planned and existing buildings and provides (1) a method to compute the standard calculated energy rating, a standard energy use that does not depend on occupant behaviour, actual weather, and other actual (environment or indoor) conditions, (2) a method to assess the measured energy rating, based on the delivered and exported energy, (3) a method to improve confidence in the building calculation model by comparison with actual energy use, and (4) a method to assess the energy effectiveness of possible improvements. ISO 16346:2013 is applicable to a part of a building (e.g. flat), a whole building, or several buildings.

4.3 Methods to assess environmental impacts of ICT

ETSI Working Group DTR/EE-00008: This work defines the methods to assess the environmental impact Assessment of ICTs including the Positive Impact by using ICT Services. This work will define the methods to assess the environmental impacts of ICT, which have two aspects. (a) Negative impact caused by the energy consumptions or CO2 emissions of operators of ICT equipment and sites including telecom network, users' terminals and

datacentres for residential and business services. (b) Positive impact caused by energy saving or CO2 emission saving by using ICT services. We propose the method of how to quantify these impacts at national level.

4.4 Technical documents

IETF: IETF's mission is to improve the Internet by producing high quality technical documents that influence the way people design, use, and manage the Internet.

4.5 Metadata

Each data file will be accompanied by unique specified metadata, in order to allow their ease of access and re-usability. Below, we present the metadata form we will adopt.

Table 3: Metadata form

Parameter		
Document version	The version of this document	
Document format	The format of this document	
Description	A description of the data included in the document	
Date	The date of the creation of the document (yyyy-mm-dd)	
Keywords	Some keywords describing the content	
Subject	Small description of the data source	
Creator (Name of the creator of the data source)		
Sector of the provider	Information on the sector that this provider belongs to	
Permissions	The permission of this document are mandatory to be mentioned here	
Name of the Partner (The name of the partner that collected the data and is responsible for)		
Responsible person	The name of the person within the partner, who is responsible for the data	
Pilot	For which pilot the data will be used	
Scenario of data usage	How the data are going to be used in this scenario	
Description of the Data Source		
File format	The format of the data source provided	
File name/path	The name of the file	
Storage location	In case a URI/URL exists for the data provider	
Data type	Data type and extension of the file; e.g. Excel Sheet, .xlsx; Standard if possible	
Standard	Data standard, if existent	
Data size	Total data size, if possible	
Time references of data	Start date	End date
Availability	Start date	End date
Data collection frequency	The time frequency in which the data is collected; e.g. hourly, every 15 minutes, on demand, etc.	
Data quality	The quality of the data; is it complete, does it have the right collection frequency, is it available, etc.	
Raw data sample		
Textual copy of data sample		

Number of Parameters included:			
Parameter #1:			
Variables	**Name**	**Type**	**Mandatory**
	…	…	…
Parameter #2:			
Variables	**Name**	**Type**	**Mandatory**
	…	**…**	**…**

The Energy-Infrastructure Monitoring Parameters' Semantic Model and Citizens Environment Friendly Behavioural Semantic Model form a semantic metadata in an integrated manner. This document contains the initial versions of the both models. More detailed information about the models will be presented in their own deliverables.

Figure 1 represents the concepts and parameters - along with their relationships- we will adopt for collecting information from the different type of sensors from an energy efficiency perspective. The types of sensors refer mainly to energy consumption, production and storage meters. This semantic model is fully extensible, it's evolution with the addition of new concepts -or even the refinement of part of the existing concepts- will be an evolving process.

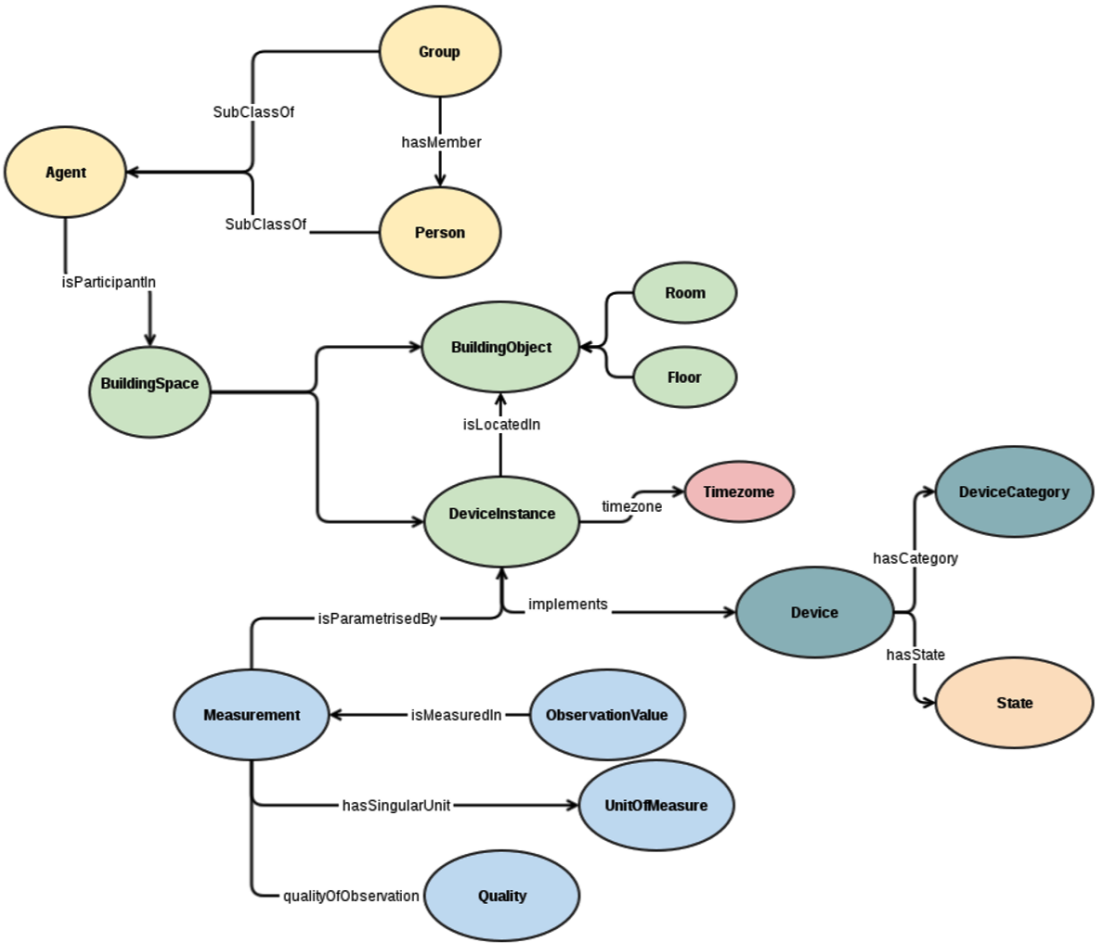

Figure 1: ENTROPY Energy Efficiency Semantic Model

A building space (BuildingSpace) in ENTROPY Energy Model defines the physical spaces of the building. A building space contains devices (DeviceInstance) or building objects (BuildingObject). A building object is an object in the building such as a room (Room) of a floor (Floor) that can contain one or more devices. A Device instance (DeviceInstance) implements a Device (Device), has a unique id and geolocation coordinates (latitude, longitude) and is located in a Specific Building Object. Each Device Instance (DeviceInstance) supports a set of measurements (Measurement) that are also associated with a set of quantitative and qualitative (Quality) characteristics (e.g. frequency, accuracy) and units of measurements (e.g. KWh, bar, m, ^{o}C etc.) (UnitofMeasure). The values of the monitored parameters are included in the Observation Values (ObservationValue). Each observed value is also associated with a timestamp (Timezone). A building space may be frequented by individuals (Agent) which could be one person (Person) or group of people (Group).

Figure 2 shows the basic concepts and their relationships in the behavioural model. This model will be utilized for representing extracted behavioural data of users. It includes concepts representing demographic data, as well as activity data and user's context information.

The ontology borrows concepts from external ontologies and vocabularies when appropriate, in order to increase reusability. For instance, initial model uses Agent, Person and OnlineAccount concepts from the FOAF[1] vocabulary. As stated in deliverable D1.1, reuse of other existing ontologies also will be considered in the later versions of this model.

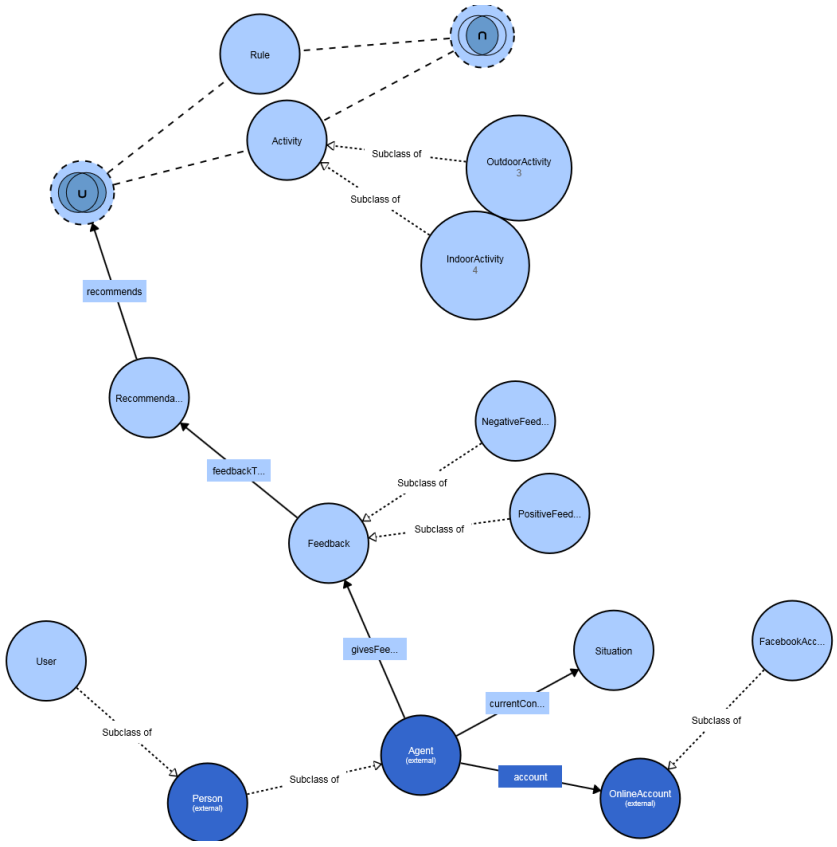

Figure 2: Initial concepts of the behavioural model

[1] *http://xmlns.com/foaf/spec/*

5. DATA ACCESS AND SHARING

Data access and sharing plan include several aspects that have to be identified regarding the data resulted from the project. Bellow the issues regarding the data access and sharing plan are presented in a more detailed manner.

5.1 IPRs and Privacy Issues

Data access and sharing activities will be implemented in compliance with the privacy and data collection rules and regulations, as they are applied nationally and in the EU, as well as with the H2020 rules. Concerning the results of the project, these will become publicly available based on the IPRs as described in the Consortium Agreement.

Due to the nature of the data involved, some of the results that will be generated by each project phase will be restricted to authorized users, while other results will be publicly available. Data access and sharing activities will be rigorously implemented in compliance with the privacy and data collection rules and regulations, as they are applied nationally and in the EU, as well as with the H2020 rules. One possibility would be to ask users to pre-register for the purpose of using the system and will then need to authenticate them against a user database. If successful, the users will have roles associated with them. These roles will determine the level of access that a user will be given and what they will be permitted to do.

5.2 Methods for Data Sharing

As the raw data included in the data sources, will be gathered from sensor nodes and information management systems, those could be seen as highly commercially-sensitive. Therefore, access to raw data can only take place between the specific end users and the partners involved in the analysis of the data. For the models to function correctly, the data will have to be included into the ENTROPY repository. The results of the data analytics in the orient phase are set to be anonymised and made available to the subsequent layers of the framework, which will then allow the possibility for external industry stakeholders to use the results of the project for their own purposes. Publications will be released and disseminated through the project dissemination and exploitation channels to make these parties aware of the project as well as appropriate access to the data. Additionally, data that are eligible for public distribution may be disseminated through:

- Scientific papers
- Lectureships in case of Universities
- Dissemination via the appropriate channels of the project
- Interest groups created by the project's partners

Rather than the raw data used, there will be knowledge obtained by applying analytics processes to low level information in order to extract behavioural information about users. Such behavioural data that will be collected throughout the project, or a fragment of it, will be published by following the linked data principles[2]:

- Use URIs to name (identify) things.

[2] *https://www.w3.org/DesignIssues/LinkedData.html*

- Use HTTP URIs so that these things can be looked up (interpreted, "dereferenced").
- Provide useful information about what a name identifies when it's looked up, using open standards such as RDF, SPARQL, etc.
- Refer to other things using their HTTP URI-based names when publishing data on the Web.

Open access to the anonymized behavioural data will be provided by means of periodic data dumps and read-only SPARQL endpoints.

6. ARCHIVING AND PRESERVATION

Short Term

All original raw data files and respective processing programs will be versioned over time and maintained in a date-stamped file structure. Access to the datasets will be given only after request and during the design phases of the project to the responsible person. These datasets will be automatically backed up on a nightly and monthly basis.

Respectively, the data generated by the system during the pilots of the project will be stored to the database of ENTROPY platform, whose DB schema will reflect the aforementioned pilot parameters. Back-ups of the DB will be performed and stored on a monthly-basis. Also, the datasets will be automatically backed up on a nightly and monthly basis.

Long Term

The project consortium is committed to make the high quality final data generated by ENTROPY available for use by the research community, as well as industry peers. We will identify appropriate platform solutions (e.g. https://joinup.ec.europa.eu/ and http://ckan.org/) that will allow the sustainable archiving of all the ENTROPY datasets after the life span of the project.