GOEASY
GalileO-based trustEd Applications for health and SustainabilitY

D6.1 – Data Management Plan

Deliverable ID	**D6.1**
Deliverable Title	**Data Management Plan**
Work Package	**WP6**
Dissemination Level	**PUBLIC**
Version	**1.0**
Date	**2018-04-03**
Status	**final**
Lead Editor	**FIT**
Main Contributors	**Sisay Adugna Chala (FIT),**
	Karl Catewitz (FIT),
	Yannick Bacheler (FIT),
	Anja Linnemann (FIT)

Published by the GOEASY Consortium

Document History

Version	Date	Author(s)	Description
0.1	2018-02-14	FIT	First Draft with TOC
0.2	2018-02-15	FIT	Introduction added
0.3	2018-02-23	FIT	Data naming & identification added
0.4	2018-02-28	FIT	Chapter 2-3 added, Appendix2 added
0.4_YB	2018-03-15	FIT	WP2 DMPs filled in
0.5	2018-03-23	FIT	Comments and adjustments from partners incorporated. References updated, TOC updated, modification Annex 2, minor corrections
0.5_ISMB	2018-03-28	ISMB	Review with minor corrections
1.0	2018-04-03	FIT	TOC updated, correction by ISMB incorporated, Final version for submission

Deliverable no. D0.0
Deliverable Title Data Management Plan
Version 1.0 - 03/04/2018

Page 2 of 17

Table of Contents

1   Introduction

The purpose of this document is to present the initial Data Management Plan (DMP) of the GOEASY 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"[*]. All GOEASY data will be handled according to EU Data protection and Privacy regulation and the upcoming General Data Protection Regulation (GDPR)[†].

The GOEASY 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 - if appropriate - during the project lifetime (in the form of deliverables). GOEASY will be deployed in two pilot sites in different countries: (I) Stockholm, Sweden and (II) Turin, Italy. Currently, the deployment and usage of the deployed GOEASY functionalities is not yet defined. Therefore, we will need to update the DMP with the data that is being collected/created at each pilot site according to its usage and whether it can be published as Open Data.

1.1   Scope

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 GOEASY project. FAIR Data Management is highly promoted by the Commission and since GOEASY deals with several kind of data, relevant attention has been given to this task. However, the DMP is a living document in which information will be made available on a more detailed level through updates and additions as the GOEASY project progresses.

1.2   Methodology

The DMP concerns all the data sets that will be collected, processed and/or generated 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 Appendix 1 GOEASY Template for DMP.
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 this document as updates in section 3.
   a) This document is the living document describing which data is 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.

---

[*] http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf (Accessed 15 February 2018)
[†] https://gdpr-info.eu/ (Accessed 22 March 2018)

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. We expect that in the GOEASY project, the partners can share most of such data under an Open Data Commons Open Database License (ODbL).

1.3 Related documents

ID	Title	Reference	Version	Date
[RD.1]	Description of Action/ Grant Agreement			

2 GOEASY 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 countries and of different types of events, i.e. closed, open etc.

2.1 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 GOEASY 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:

> GOEASY_{Country or WP}_{Pilot Site or WP}_{Responsible Partner}_{Description}_{Data Set Sub Index}

Figure 1: GOEASY Data Set Naming Scheme

The parts are defined as follows:
- GOEASY: Static for all data sets and is used for identifying the project.
- Country: The two letter ISO 3166-1 country code for the pilot where data has been collected or generated.
- WP: the work package together with 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., AsthmaWatch etc.
- Responsible Partner: The partner that is responsible for managing the collected data, i.e. creates and maintains the Open Data Management plan for the data set. Using the acronyms from D1.1.
- Description: Short name for the data set, without spaces with CamelCaps in case of multiple words, e.g., Location, Pollution level 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.

> GOEASY_IT_Turin_GAPES_Location_1

Figure 2: GOEASY Data Set Naming Example

In the example shown in Figure 2, the Data set is created within GOEASY project in Italy at Turin pilot site. GAPES is responsible for Open 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 Open 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.

2.2   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. GOEASY 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

2.3   Fair Data

FAIR data management means in general terms, that research data should be "FAIR" (**F**indable, **A**ccessible, **I**nteroperable and **R**e-usable). These principles precede implementation choices and do not necessarily suggest any specific technology, standard, or implementation solution.

2.3.1   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 GOEASY 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:
- Open Data Foundation
- Open Knowledge Foundation
- Open Government Standards

Furthermore, data should be interoperable, adhering for data annotation and data exchange, compliant with available software applications related to LBS.

2.3.2   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, will be made available for download under the ODbL License.
- Specify how access will be provided in case there are restrictions.

2.3.3   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 GOEASY 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[‡], 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.

2.3.4   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.

2.4   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.

2.5   Data security

The provisions for data security and recovery are taken care by the partners running their respective databases. The responsible partners will use the feedback generated by the security related tasks, when setting up their back ends. The security related feedback comes from T3.1, T3.3 and T4.5.

The security and integrity of the data transfer is guaranteed by the applied state-of-the-art software frameworks or libraries. At the end of the project, the consortium will decide if a long term preservation of the data is needed. EU's OpenAIRE suggestions for selecting a proper repository will be taken into account.

2.6   Ethical aspects

The informed consent covers the intended use of the data including long term preservation, in accordance with EU regulation (see Appendix 2). Further, Deliverable 8.1 "Ethics requirements" (including the updated Data Management Plan) refers to ethical aspects, with special focus on PODP (Protection of Personal Data).

2.7   Other issues

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

[‡] Open Geospatial Consortium (OGC) SensorThings API. https://github.com/opengeospatial/sensorthings (Accessed 5 August 2017)

3 Initial DMP Components in GOEASY

During third and fourth quarter of the project, each work package will analyse which DMP components are relevant in 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 definition will follow the template in Annex 1. Here below we present a first set of initial generic DMP components.

3.1 WP2 – User Scenarios (ApesMobility)

Table 1: DMP for WP2 – User Scenarios (ApesMobility)

DMP Element	Issues to be addressed			
Identifier	GOEASY_Italy_WP2_FIT_UserScenarios_ ApesMobility			
DMP Responsible Partner	FIT			
Revision History	**Date**	**Partner**	**Name**	**Description of change**
	2018-03-15	FIT	Yannick Bachteler	Created initial DMP
Data Summary	Definition of scenarios for scoping of the initial requirements (D2.1 Initial Visions, Scenarios and Use Cases; updated in D2.4 Updated Visions, Scenarios, Use Cases and Innovation; and in D2.6 Final Visions, Scenarios, Use Cases and Innovations) is based on brainstorming, focus group and discussions with pilot partners. Talking to and gathering data from end users is an integral part of the GOEASY project and will help to ensure that a useful product is created. Evaluated data is presented in a graphical way within the deliverables, e.g. as mind maps.			
Making data findable, including provisions for metadata	It will become both discoverable and accessible to the public when the consortium decides to do so. D2.1 and its updated (D2.4) and final version (D2.6) contain a table stating all versions of the document, along with who contributed to each version, what the changes where as well as the date the new version was created.			
Making data openly accessible	Data is available in deliverables (D2.1, D2.4, D2.6). The dissemination level of D2.1 is public. It is available through a document sharing system (BSCW) for the members of the consortium. As soon as deliverables will be publicized, it will be uploaded along with the other public deliverables to the project website or anywhere else the consortium decides.			
Making data interoperable	Raw data (e.g. audio recording of focus group) cannot be made freely available because it contains sensitive information.			
Increase Data Re-use	Engineers, who want to build similar systems, could use it as a foundation.			
Allocation of Resources	N/A			
Data security	Documentation of scenarios will be securely saved on the FITs premises and will be shared with the rest of the partners through the GOEASY wiki (Confluence) and document sharing system.			
Ethical aspects	N/A			
Other Issues	N/A			

3.2 WP2 – User Scenarios (AsthmaWatch)

Table 2: DMP for WP2 – User Scenarios (AsthmaWatch)

DMP Element	Issues to be addressed
Identifier	GOEASY_Sweden_WP2_FIT_UserScenarios_AsthmaWatch
DMP Responsible Partner	FIT

Revision History	**Date**	**Partner**	**Name**	**Description of change**
	2018-03-15	FIT	Yannick Bachteler	Created initial DMP

Data Summary	Definition of user scenarios for scoping of the initial requirements (D2.1 Initial Visions, Scenarios and Use Cases; updated in D2.4 Updated Visions, Scenarios, Use Cases and Innovation; and in D2.6 Final Visions, Scenarios, Use Cases and Innovations) is based on brainstorming, interviews and discussions with pilot partners. Talking to and gathering data from end users is an integral part of the GOEASY project and will help to ensure that a useful product is created. Evaluated data is presented in a graphical way within the deliverables, e.g. as mind maps.
Making data findable, including provisions for metadata	It will become both discoverable and accessible to the public when the consortium decides to do so. The D2.1 and its updated (D2.4) and final version (D2.6) contain a table stating all versions of the document, along with who contributed to each version, what the changes where as well as the date the new version was created.
Making data openly accessible	Data is available in deliverables (D2.1, D2.4, D2.6). The dissemination level of D2.1 is public. It is available through the document sharing system (BSCW) for the members of the consortium. As soon as deliverables will be publicized, it will be uploaded along with the other public deliverables to the project website or anywhere else the consortium decides.
Making data interoperable	Raw data (e.g. interview protocol) cannot be made freely available because it contains sensitive information.
Increase Data Re-use	Engineers who want to build similar systems, could use it as a foundation.
Allocation of Resources	N/A
Data security	Documentation of scenarios will be securely saved on the Fraunhofer premises and will be shared with the rest of the partners through the GOEASY wiki and document sharing system.
Ethical aspects	N/A
Other Issues	N/A

3.3 **WP2 – User Requirements**

Table 3: DMP for WP2 – User Requirements

DMP Element	**Issues to be addressed**
Identifier	GOEASY_WP2_WP2_FIT_UserRequirements_1
DMP Responsible Partner	FIT

Revision History	**Date**	**Partner**	**Name**	**Description of change**
	2018-03-15	FIT	Yannick Bachteler	Created initial DMP

Data Summary	Analysis and definition of user requirements for scoping of the initial requirements (D2.1 Initial Visions, Scenarios and Use Cases; and updated versions) are based on brainstorming, interviews, focus group and discussions with pilot partners (see previous DMP). The data is essential for the technical team to develop the GOEASY platform; other partner teams throughout the project, as well as the wider research community will benefit when results are published.
Making data findable, including provisions for metadata	It will become both discoverable and accessible to the public when the consortium decides to do so. The D2.1 and its updated (D2.4) and final version (D2.6) contain a table stating all versions of the document, along with who contributed to each version, what the changes where as well as the date the new version was created.
Making data openly accessible	Data is/ will be available in deliverables (D2.1, D2.4, D2.6). The dissemination level of D2.1 is public. It is available through the document sharing system

	(BSCW) for the members of the consortium. As soon as deliverables will be publicized, it will be uploaded along with the other public deliverables to the project website or anywhere else the consortium decides.
Making data interoperable	Raw data is recorded and formatted as user stories in the JIRA Issue tracker hosted at Fraunhofer premises.
Increase Data Re-use	Engineers who want to build similar systems, could use this as an example.
Allocation of Resources	N/A
Data security	Documentation of requirements will be securely saved on the Fraunhofer premises and will be shared with the rest of the partners through the GOEASY wiki and document sharing system.
Ethical aspects	N/A
Other Issues	N/A

3.4   WP5 – Scalability and e-Security stress-tests

Table 4: DMP-template for WP5 – Scalability and e-Security stress-tests

DMP Element	Issues to be addressed			
Identifier	GOEASY_WP5_WP5_ISMB_Stresstesting			
DMP Responsible Partner	ISMB			
Revision History	**Date**	**Partner**	**Name**	**Description of change**
	2018-xx-xx	ISMB		Created initial DMP
Data Summary				
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				

3.5   WP6 – Citizens Engagement, Recruitment and Support

Table 5: DMP-template for WP6 – Citizens Engagement, Recruitment and Support

DMP Element	Issues to be addressed			
Identifier	GOEASY_WP6_WP6_COT_CitizensEngagement			
DMP Responsible Partner	COT			
Revision History	**Date**	**Partner**	**Name**	**Description of change**
	2018-xx-xx	COT		Created initial DMP
Data Summary				
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	

3.6 WP6 – ApesMobility Pilot

Table 6: DMP-template for WP6 – ApesMobility Pilot

DMP Element	Issues to be addressed			
Identifier	GOEASY_WP6_WP6_GAPES_ApesMobility			
DMP Responsible Partner	ISMB			
Revision History	**Date**	**Partner**	**Name**	**Description of change**
	2018-xx-xx	GAPES		Created initial DMP
Data Summary				
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				

3.7 WP6 – Citizens Engagement, Recruitment and Support

Table 7: DMP-template for WP6 – Citizens Engagement, Recruitment and Support

DMP Element	Issues to be addressed			
Identifier	GOEASY_WP6_WP6_CNET_AsthmaWatchPilot			
DMP Responsible Partner	CNET			
Revision History	**Date**	**Partner**	**Name**	**Description of change**
	2018-xx-xx	CNET		Created initial DMP
Data Summary				
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				

3.8 WP6 – Holistic GOEASY Platforms and Applications Evaluation

Table 8: DMP-template for WP6 – Holistic GOEASY Platforms and Applications Evaluation

DMP Element	Issues to be addressed			
Identifier	GOEASY_WP6_WP6_FIT_PlatformEvaluation			
DMP Responsible Partner	FIT			
Revision History	**Date**	**Partner**	**Name**	**Description of change**
	2018-xx-xx	FIT		Created initial DMP
Data Summary				
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				

3.9 WP7 – Dissemination

Table 9: DMP-template for WP7 – Dissemination

DMP Element	Issues to be addressed			
Identifier	GOEASY_WP7_WP7_GAPES_Dissemination			
DMP Responsible Partner	GAPES			
Revision History	**Date**	**Partner**	**Name**	**Description of change**
	2018-xx-xx	GAPES		Created initial DMP
Data Summary				
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				

Abbreviation

Abbreviation	Explanation
DMP	Data Management Plan
WP	Work Package
IoT	Internet of Things
LBS	Location-based Service
WBS	Work Breakdown Structure
GNSS	Global Navigation Satellite System
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
GAPES	greenApes Srl SB
FIT	Fraunhofer Institute for Applied Information Technology
CNET	CNet Svenska AB
COT	Città di Torino
ISMB	Istituto Superiore Mario Boella sulle Tecnologie dell' Informazione e delle Telecomunicazioni
OGC	Open Geospatial Consortium
POPD	Protection of Personal Data

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 15 February 2018)

[2]   Official PDF of the Regulation (EU) 2016/679 (General Data Protection Regulation), https://gdpr-info.eu/ (Accessed 22 March 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 22 March 2018)

[4]   Open Geospatial Consortium (OGC) SensorThings API. https://github.com/opengeospatial/sensorthings (Accessed 5 August 2017)

DMP Element	Issues to be addressed			
Identifier	The identifier of the data set following the GOEASY naming principles, see 2.1			
DMP Responsible Partner	The Partner that is responsible for creating and maintaining the DMP			
Revision History	**Date**	**Partner**	**Name**	**Description of change**
	2018-xx-xx	xxx	xxx	Created initial DMP
Data Summary	Guidelines in 2.2			
Making data findable, including provisions for metadata	Guidelines in 2.3.1			
Making data openly accessible	Guidelines in 2.3.2			
Making data interoperable	Guidelines in 2.3.3			
Increase Data Re-use	Guidelines in 2.3.4			
Allocation of Resources	Guidelines in 2.4			
Data security	Guidelines in 2.5			
Ethical aspects	Guidelines in 2.6			
Other Issues	Guidelines in 2.7			

Appendix 2 GALILEO Informed Consent Form

Users will be required to read, understand and sign the Informed Consent Sheet.

Generic Consent Form Template[§]

I understand that my participation in the GOEASY project will involve [provide brief description of what is required, e.g. ...completing two questionnaires about my attitudes toward controversial issues which will require approximately 20 minutes of my time.].

I understand that participation in this project is entirely voluntary and that I can withdraw from the project at any time without giving a reason.

I understand that I am free to ask any questions at any time. I am free to withdraw or discuss my concerns with [name].

I understand that the information provided by me will be held totally anonymously, so that it is impossible to trace this information back to me individually. I understand that this information may be retained indefinitely.

I also understand that at the end of the project I will be provided with additional information and feedback about the purpose of the project.

I, _____(NAME) consent to participate in the project conducted by [name]

Signed:

Date:

[§] If required, the Information Consent Sheet can be extended e.g. to match the upcoming GDPR Guidlines.