ICT-25-2016-2017: Advanced robot capabilities research and take-up

Project Title:

Robot for Autonomous Underground Trenchless Operations, Mapping and Navigation

BADGER

Grant Agreement No: 731968

Research and Innovation Action (RIA)

Deliverable

D7.4. Data Management Plan

Deliverable No.		**D7.2**	
Workpackage No.	**WP7**	Workpackage Title and task type	**Exploitation and Dissemination**
Task No.	**T7.4**	Task Title	**Dissemination plan and communication tools**
Lead beneficiary		**SILO**	
Dissemination level		**PU – Public**	
Nature of Deliverable		**Report**	
Delivery date		**26 June 2018**	
Status		**Draft**	
File Name:		**BADGER_Deliverable_7.4.doc**	
Project start date, duration		**01 January 2017, 36 Months**	

This project has received funding from the European Union's Horizon 2020 Research and innovation programme under Grant Agreement n°731968

Authors List

Leading Author (Editor)				
Surname		*Initials*	*Beneficiary Name*	*Contact email*
Vartholomeos		P.	SILO	pvartholomeos@singularlogic.eu
Co-authors (in alphabetic order)				
#	*Surname*	*Initials*	*Beneficiary Name*	*Contact email*
1	*Giakoumis*	*D.*	*CERTH*	*dgiakoum@iti.gr*

Reviewers List

List of Reviewers (in alphabetic order)				
#	*Surname*	*Initials*	*Beneficiary Name*	*Contact email*
1	*Fischer*	*S.*	*TT*	*Sebastian.fischer@tracto-technik.de*
2	*Manacorda*	*G.*	*IDS*	*g.manacorda@idsgeoradar.com*

Document history			
Version	Date	Status	Modifications made by
1.0	24/06/2018	First version	SILO
2.0	26/06/2018	Second version	SILO, All partners
3.0	26/06/2018	Final version	SILO

Executive Summary

BADGER is a Horizon 2020 project participating in the Open Research Data Pilot. This pilot is part of the Open Access to Scientific Publication and Research Data programme in H2020. The goal of the program is to foster access to data generated in H2020 projects.

The current document provides detailed information about the datasets that are planned to be captured by the partners of the BADGER project. The foreseen datasets are those agreed by the partners as of month 18 of the project. A more complete list of datasets will be included in the future, as the project progresses. The DMP will be updated every semester. The report also presents a set of data security measures that pertain to the Badger applications and collected data.

BADGER will develop a data management portal as part of its website. The initial version of the data management portal will become available during the second half of the 2nd year of the project, in parallel to the establishment of the first versions of project datasets that can be made publicly available.

Table of Contents

1. Introduction

BADGER is a Horizon 2020 project participating in the Open Research Data Pilot. This pilot is part of the Open Access to Scientific Publication and Research Data programme in H2020. The goal of the program is to foster access to data generated in H2020 projects.

Open Access refers to a practice of giving online access to all scholarly disciplines in formation that is free of charge to the end-user. In this way data becomes re-usable, and the benefit of public investment in the research will be improved.

The EC has provided a document with guidelines [1] and a template (see Annex I) for projects participants in the pilot. The guidelines address aspects like research data quality, sharing and security. According to the guidelines, projects participating will need to develop a Data Management Plan (DMP).

The DMP describes the types of data that will be generated or gathered during the project, the standards that will be used, the ways how the data will be exploited and shared for verification or reuse, and how the data will be preserved.

This document has been produced following these guidelines and aims to provide a consolidated plan for BADGER partners in the data management plan policy that the project will follow. The document is the second version of the DMP, delivered in M18 (June 2017) of the project. The DMP will be updated periodically, every 6 months, during the lifecycle of the project.

1.1 Background of the BADGER DMP

The BADGER DMP will be written in reference to the Article 29.3. in the Model Grant Agreement called “the open access to research data” (research data management). Project participants must deposit their data in a research data repository and take measures to make the data available to third parties. The third parties should be able to access, mine, exploit, reproduce and disseminate the data. This should also help to validate the results presented in scientific publications. In addition, Article 29.3 suggests that participants will have to provide information, via the repository, about tools and instruments needed for the validation of project outcomes.

The DMP will be important for tracking all data produced during the BADGER project. Article 29.3 states that project beneficiaries do not have to ensure access to parts of research data if such access would lead to a risk for the project’s goals. In such cases, the DMP must contain the reasons for not providing access. According to the aforementioned DMP Guidelines it is planned that research data management projects funded under H2020 will receive support though the research Infrastructures Work Programme 2014-15 (call 3 e-Infrastructures). Full support services are expected to be available only to research projects funded under H2020, with preference to those participating in the Open Research Data Pilot.

2. BADGER Data Management Plan

2.1 The BADGER Data Management portal

BADGER will develop a data management portal as part of its website. This portal will provide to the public, for each dataset that will become publicly available, a description of the dataset along with a link to a download section. The portal will be updated each time a new dataset has been collected and is ready for public distribution. However the portal shall not contain any datasets that should not become publicly available.

The initial version of the portal will become available during the second half of the 2nd year of the project, in parallel to the establishment of the first versions of project datasets that can be made publicly available. The BADGER data management portal will enable project partners to manage and distribute their public datasets through a common infrastructure.

2.2 Datasets naming conventions

Concerning the convention followed for naming the BADGER datasets, it should be noted that the name of each dataset comprises: (a) a prefix 'DS' indicating a dataset, along with its unique identification number, e.g. 'DS1', (b) the name(s) of the partner(s) responsible to collect it, e.g. SILO, along with an identifier denoting the internal numbering of the dataset concerning the specific partner, e.g. -01 and (c) a short title of the dataset summarizing its content and purpose, e.g. Underground Object Recognition Dataset.

2.3 Summary of foreseen BADGER datasets

In the following, Table 1 provides a list of the expected datasets, whereas the detailed description of each dataset, in accordance to the H2020 DMP template is provided in the following sections. At this stage (M18) there are five datasets foreseen in the project, covering a series of research dimensions on the skills the BADGER robot should develop. In the course of the project more Datasets will be added in the Data Management Plan.

Table 1. Summary of foreseen BADGER datasets (as of month 18)

No	Dataset name
1	DS1_CERTH_GPR_Measurements
2	DS2_TT_01_System_Requirements
3	DS3_TT_02_Pilot_Experiments
4	DS4_UoG_Ultrasonic_System
5	DS5_UC3M_Control_system
6	DS6_CERTH_Surface-Subsurface_Mapping

3. Datasets Description

In this paragraph, we provide detailed information about the datasets that are planned to be captured by the partners of the BADGER project. These are the foreseen datasets as of month 18 of the project. More datasets will be included in the course of the project and the DMP will be updated every semester.

In order to meet the requirements of the DMP according to the Pilot Open Access of the Horizon 2020, each partner provided the description of their datasets using the template given in Annex I, which was formed by following the EC guidelines of the dataset aspects that should be reported in DMPs of the H2020 projects. Based on this information, partner SILO compiled the following tables.

DS1_GPR_Measurements	
Data description	
The GPR measurements dataset is collected due to the need for formulation of consistent radar images during the surface rover navigation. The data to be collected will be used for the detection of buried objects, such as pipes. Also, a synthetic dataset is generated for training of the related algorithms. The data will allow the training of the object detection algorithms to be developed.	
Partners activities and responsibilities	
Partner owner of the data; copyright holder (if applicable)	IDSGEO and CERTH
Partner in charge of the data collection	CERTH
Partner in charge of the data analysis	CERTH
Related WP(s) and task(s)	WP4
Standards	
Info about metadata (production and storage dates, places) and documentation.	tbd
Standards, format	The data collected consist of raw GPR measurements in plain text (ASCII) format. Other formats will be available, such as images (.png, .jpeg). Also, data in .hdf5 and .out formats are generated by gprMax software.
Estimated data size	The size of the files varies between 200KB – 20 MB.
Data exploitation and sharing	
Purpose use of the data analysis	Existing data are utilised for initial calibration of the GPR processing algorithms. These data are provided by IDS Georadar. The data will be useful to geoscience engineers by allowing further research regarding GPR signal processing. These datasets will enable training of object detection algorithms, as well as extracting of quantitative results

	regarding methods accuracy and robustness.
Data access policy/ dissemination level	The data produced and/or used in the project are useable by any interested party aiming to use them for research and development, especially by geoscience engineers. Part of the dataset could be made publicly available.
Embargo periods (if any)	tbd
Archiving and preservation (including storage and backup)	
Data storage. Where? For how long?	The data will be stored and be accessible at the project open-data database that will be linked to the project webpage.

DS2_System_Requirements	
Data description	
Data is generated and collected in order to identify relevant information to define use cases, user requirements, geological conditions, and system functionalities. Further data in terms of technical design are generated in order to be able to produce mechanical components. The generation / collection of these data is necessary to fulfil the tasks of the WP in which TT is to collaborate.	
Partners activities and responsibilities	
Partner owner of the data; copyright holder (if applicable)	TT
Partner in charge of the data collection	TT
Partner in charge of the data analysis	TT, CERTH, UC3M
Related WP(s) and task(s)	WP1, WP3, WP6
Standards	
Info about metadata (production and storage dates, places) and documentation.	tbd
Standards, format	Text (e.g. docx, pdf), calculation files (e.g. xlsx), presentations (e.g. pptx), photos (e.g. jpg), videos (e.g. mov), CAD files (e.g. dxf)
Estimated data size	tbd
Data exploitation and sharing	
Purpose use of the data analysis	Data will be useful to consortium members, dissemination and exploitation partners

Data access policy/ dissemination level	TT will make openly available use case information, end user requirements, geological information, prototype information as well as test requirement / condition data. Information about test results and test validation as well as market and business information will not be made available so competitors cannot make use of it. This information forms the basis of the future commercial exploitation of the BADGER technology. Data will be available on the BADGER website as reports and during dissemination as presentation or articles in magazines.
Embargo periods (if any)	There are no restrictions for data meant to be published. Confidential data will not be made available for public in a foreseeable period of time.
Archiving and preservation (including storage and backup)	
Data storage. Where? For how long?	The data will also be stored and be accessible at the project open-data database that will be linked to the project webpage.

DS3_Pilot_Experiments	
Data description	
Data will be generated during pilot experiments at TT industrial premisses in Lennestadt, Germany. The generation / collection of these data is necessary to fulfil the tasks of the WP in which TT is to collaborate.	
Partners activities and responsibilities	
Partner owner of the data; copyright holder (if applicable)	TT
Partner in charge of the data collection	TT
Partner in charge of the data analysis	TT, CERTH, UC3M
Related WP(s) and task(s)	WP3, WP6
Standards	
Info about metadata (production and storage dates, places) and documentation.	tbd
Standards, format	Text (e.g. docx, pdf), calculation files (e.g. xlsx), photos (e.g. jpg), videos (e.g. mov), sensor and instrumentation

	data (blob, .txt, .csv).
Estimated data size	tbd
Data exploitation and sharing	
Purpose use of the data analysis	Data will be useful to consortium members, dissemination and exploitation partners
Data access policy/ dissemination level	TT will make openly available use case information, end user requirements, geological information, prototype information as well as test requirement / condition data. Information about test results and test validation as well as market and business information will not be made available so competitors cannot make use of it. This information forms the basis of the future commercial exploitation of the BADGER technology. Data will be available on the BADGER website as reports and during dissemination as presentation or articles in magazines.
Embargo periods (if any)	There are no restrictions for data meant to be published. Confidential data will not be made available for public in a foreseeable period of time.
Archiving and preservation (including storage and backup)	
Data storage. Where? For how long?	The data will be stored at the project open-data database that will be linked to the project webpage.

DS4_Ultrasonic_System	
Data description	
Data will be collected to assess the performance of the ultrasonic systems, actuation devices, and trajectory-following performance. Simulation data will also be collected. These outputs are required to validate the objectives of the project.	
Partners activities and responsibilities	
Partner owner of the data; copyright holder (if applicable)	UoG
Partner in charge of the data collection	UoG
Partner in charge of the data analysis	UoG
Related WP(s) and task(s)	WP2
Standards	
Info about metadata (production and storage dates, places) and	The data will be stored on UoG's own repository using a unique filename and indexing system that will provide all

documentation.	appropriate test metadata. The definition of appropriate metadata will be such that all experimental conditions can be recreated at a later date. As this is a new project there is no standard as yet, but key parameters will include all internal gain settings, the power settings for the ultrasonics, actuator position histories, trajectories, and substrate parameters including sand type, depth, and compaction.
Standards, format	This data will be stored as numerical values in time domain, in a .csv or similar format, and will be required on all appropriate test runs. The data we envisage will be, for the post part, time domain logs and will not result in complex interoperability problems.
Estimated data size	The file size will depend on the length of the experimental runs carried out.
Data exploitation and sharing	
Purpose use of the data analysis	The data will be useful for UoG as we improve our systems, and to all partners for validation purposes.
Data access policy/ dissemination level	All data will be available upon reasonable request, but only those sections required to disseminate our work will be actively published. This is because not all runs will be successful and, although this will not be hidden, confusion can easily result if external partners focus on off-optimal runs. Publication of these datasets will generally be made through supporting files in association with our publications. Many publishers, and indeed our own library, can support this approach. The files will not require specialist software to open.
Embargo periods (if any)	There are no restrictions for data meant to be published. Confidential data will not be made available for public in a foreseeable period of time.
Archiving and preservation (including storage and backup)	
Data storage. Where? For how long?	The data will be stored and be accessible at the project open-data database that will be linked to the project webpage.

	Dataset will also be preserved in UoG infrastructure. The University of Glasgow library can support persistent, accessible, and curated data storage as part of their centrally-funded role within the university. The data is not expected to be sensitive, but will be securely backed-up by our library.

DS5_Control_System	
Data description	
Input-output control data will be stored, aiming at benchmarking motion control strategies (inverse simulation or other) for underground robot control.	
Partners activities and responsibilities	
Partner owner of the data; copyright holder (if applicable)	UoG, UC3M
Partner in charge of the data collection	UoG, UC3M
Partner in charge of the data analysis	UoG, UC3M
Related WP(s) and task(s)	WP2
Standards	
Info about metadata (production and storage dates, places) and documentation.	ROSbag added metadata
Standards, format	ROSbag format
Estimated data size	The file size will depend on the length of the experimental runs carried out.
Data exploitation and sharing	
Purpose use of the data analysis	The data will be useful for UoG and UC3M as we improve our systems, and to all partners for validation purposes.
Data access policy/ dissemination level	All data will be available upon reasonable request, but only those sections required to disseminate our work will be actively published.
Embargo periods (if any)	There are no restrictions for data meant to be published.
Archiving and preservation (including storage and backup)	
Data storage. Where? For how long?	The data will be stored and be accessible at the project open-data database that will be linked to the

	project webpage. Dataset will be preserved in UoG and UC3M infrastructure. The University of Glasgow library can support persistent, accessible, and curated data storage as part of their centrally-funded role within the university. The data is not expected to be sensitive, but will be securely backed-up by our library.

DS6_Surface-Subsurface_Mapping	
Data description	
The Surface-Subsurface_Mapping measurements will be collected from the integrated surface rover - GPR unit. The data will be utilized for the metric mapping of the surface rover and the subsurface mapping required for the navigation of the underground robot. Moreover, the subsurface data will be used for the utility mapping of the subsurface. Data collection will be performed at the specific test site constructed at CERTH premises, necessary for the developing and testing of all the mapping and surface rover navigation methods. For this purpose pipes of different type and material have been placed on the subsurface. The electromechanically integrated surface rover / GPR unit will be used for the data collection of stereo images, B-Scan, along with ground-truth measurements by performing multiple traverses of the field. The data will contribute to the 3D reconstruction of an unknown environment, which is one of the main objectives of the project.	
Partners activities and responsibilities	
Partner owner of the data; copyright holder (if applicable)	CERTH
Partner in charge of the data collection	CERTH
Partner in charge of the data analysis	CERTH
Related WP(s) and task(s)	WP4
Standards	
Info about metadata (production and storage dates, places) and documentation.	tbd
Standards, format	The data collected consist of raw GPR measurements in plain text (ASCII) format. Radargrams of images format (.png, jpg) Stereo Images with Calibration data (.bag) Timestamp synchronizaiton data in plain text (ASCII)

Estimated data size	The size of the files varies between 200KB – 1 GB. Total size of dataset expected to be ~100GB
Data exploitation and sharing	
Purpose use of the data analysis	The data to be produced can be utilized for the development of: • Outdoors mapping methods • GPR data processing and A-Assembly methods in B-Scans • Registered surface/subsurface recordings • Coupled surface/subsurface mapping methods • Further processing of subsurface data for utility mapping based on the extraction o semantics.
Data access policy/ dissemination level	The data produced and/or used in the project are useable by any interested party aiming to use them for research and development, especially by robotics and geoscience engineers. Part of the dataset could be made publicly available.
Embargo periods (if any)	tbd
Archiving and preservation (including storage and backup)	
Data storage. Where? For how long?	The data will be stored and be accessible at the project open-data database that will be linked to the project webpage.

4. Timeline of release of datasets

Most of the datasets will be generated and released by the end of the second year of the project. An indicative timeline for the release of the datasets that are foreseen in the current version of the BADGER DMP is provided in Table 2 below.

Table 2. Timeline of the release of the datasets

	Y1												Y2												Y3											
	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	M16	M17	M18	M19	M20	M21	M22	M23	M24	M25	M26	M27	M28	M29	M30	M31	M32	M33	M34	M35	M36
WP1. Robotic system specifications, tunneling specifications and system architecture (UC3M)																					DS2															
WP2. Robotic system design and hardware development and overall system integration (UC3M)																						DS5														
WP 3. Robot control system (UoG)																											DS4									
WP4. Robot cognition (CERTH)																											DS1			DS6						
WP5. Software development for data management, GUI, and process visualization (SILO)																																				
WP6. Validation experiments and application domains (TT)																																				DS3

This timeplan a preliminary estimation of dataset release dates. The actual timetable of the datasets expected to be released will be formulated progressively, as the project evolves.

5. Data set security measures

The data set that will be generated during the Badger project does not contain any personal data and therefore does not raise any issues with respect to the GDPR regulation.

All data collected are strictly related to the underground environment. Data generated by the BADGER robot during the project will not contain information of any public infrastructure, because all tests will be carried out at the premises of TT partner, as well as in the CERTH test site. Hence, during the 3- year period of the project there are no issues raised with respect to sensitive infrastructure information. However, in the future, the commercial Badger system will operate on construction sites and the collected information might include public infrastructure (utilities etc.) which might be considered sensitive and even raise national security issues. Therefore, appropriate security measures should be put into effect. These measures should handle Internet, Linux-machine and Wi-Fi communication security. Also, they should address local database storage and data export security issues. These measures are grouped and listed below:

The following simplified architecture in **Figure 1** depicts the main communication lines and software modules that need to be secured. All measures explained below refer to this figure. These measures will be implemented and demonstrated on the badger software architecture in the context of Task 5.4 in WP5.

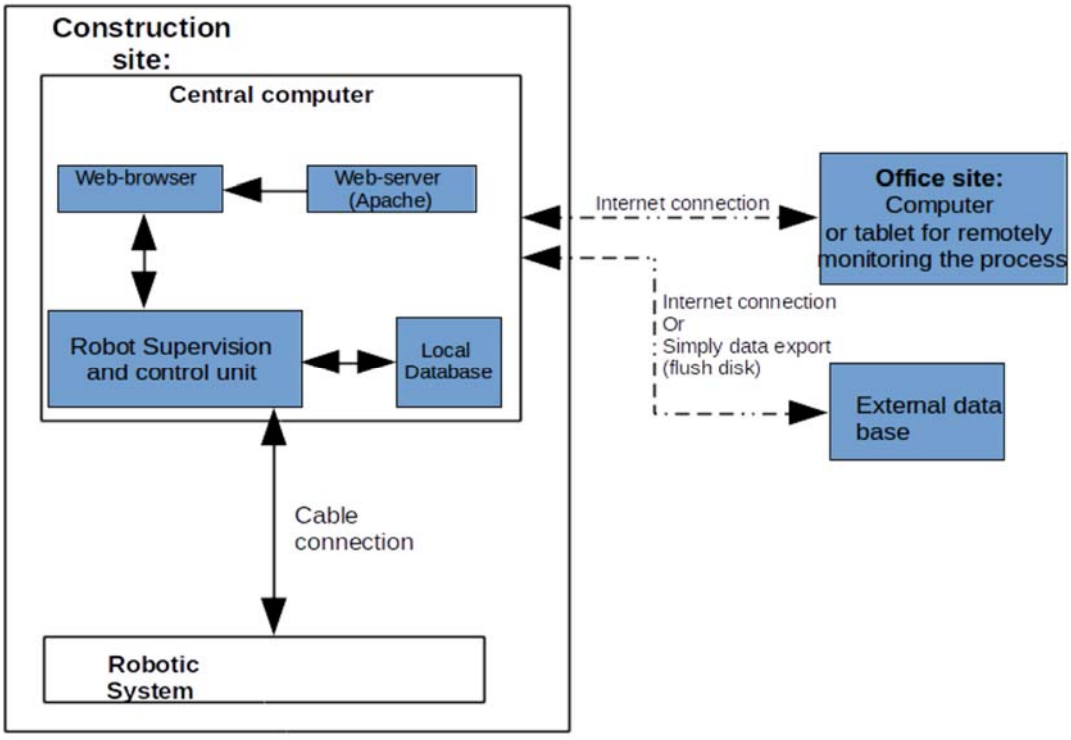

Figure 1: Main Communication architecture

Internet communication security measures include:

- Use VPN connection with key pairs for authentication instead of password. The end points are the construction site PC and a Tablet or a remote PC (at a remote office).
- Configure the Linux environment at the construction site so that Root access is granted only locally. This way no remote user can access Root privileges over the Internet

Linux machine security measures include:

- Harden the Linux at the construction site. This means:
  - Make sure Web browser uses https
  - Follow strict password guidelines (do not allow for default passwords, no weak passwords, change of password on a regular basis, etc.)
  - Configure Linux for minimum port definition (
  - Use only SSH
- Harden the Apache server
  - Allow communication only through 443 port (installation of SSL certificate)
  - Disable directory listing
  - Disable not necessary modules
  - Restrict access to directories
  - Use mod_evasive and mod_security modules
  - Limit the default size of requests

Wi-Fi security measures include:

- Use of WPA2 protocol

Database security include:

- Encryption of the data which is exported. Encrypted data will be exported using a flush disk (hence data is encrypted on transit). When data is copied to an external database it will be decrypted.
- In case data from the local database is transmitted over the Internet it should be encrypted and the transmission should be done over a secure channel such as VPN.

6. Conclusions

The current document provides preliminary, yet detailed information about the datasets that are planned to be captured by the partners of the BADGER project. These are the foreseen datasets as of month 18 of the project. A more complete list of datasets will be included in the future, as the project progresses. The DMP will be updated every semester. The report also presents a set of data security measures that pertain to the Badger applications and collected data.

BADGER will develop a data management portal as part of its website. The initial version of the data management portal will become available during the second half of the 2nd year of the project, in parallel to the establishment of the first versions of project datasets that can be made publicly available.

References

[1] /http://www.gsrt.gr/EOX/files/h2020-hi-oa-data-mgt_en.pdf

Annex I:

The following is a template provided by the EU commission to assist in the authoring of the Data Management Plan.

At this stage partners will be in position to fill-in only part of the requested information, i.e. the most basic, such as type of data generated, public/confidential, purpose of data etc. The table will be updated regularly (every 3 months). Hence, partners will have the opportunity, at a later stage, to add/modify information.

DMP component	Issues to be addressed
1. Data summary	• State the purpose of the data collection/generation • Explain the relation to the objectives of the project • Specify the types and formats of data generated/collected • Specify if existing data is being re-used (if any) • Specify the origin of the data • State the expected size of the data (if known) • Outline the data utility: to whom will it be useful
2. FAIR Data 2.1. Making data findable, including provisions for metadata	• Outline the discoverability of data (metadata provision) • Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers? • Outline naming conventions used • Outline the approach towards search keyword • Outline the approach for clear versioning • Specify standards for metadata creation (if any). If there are no standards in your discipline describe what type of metadata will be created and how

2.2 Making data openly accessible	• Specify which data will be made openly available? If some data is kept closed provide rationale for doing so • Specify how the data will be made available • Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)? • Specify where the data and associated metadata, documentation and code are deposited • Specify how access will be provided in case there are any restrictions
2.3. Making data interoperable	• Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability. • Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?
2.4. Increase data re-use (through clarifying licences)	• Specify how the data will be licenced to permit the widest reuse possible • Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed • Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why • Describe data quality assurance processes • Specify the length of time for which the data will remain re-usable
3. Allocation of resources	• Estimate the costs for making your data FAIR. Describe how you intend to cover these costs • Clearly identify responsibilities for data management in your project • Describe costs and potential value of long term preservation
4. Data security	• Address data recovery as well as secure storage and transfer of sensitive data
5. Ethical aspects	• To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former
6. Other	• Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if

	any)

HISTORY OF CHANGES		
Version	**Publication date**	**Change**
1.0	13.10.2016	▪ Initial version