Information and Communication Technologies (ICT) Programme
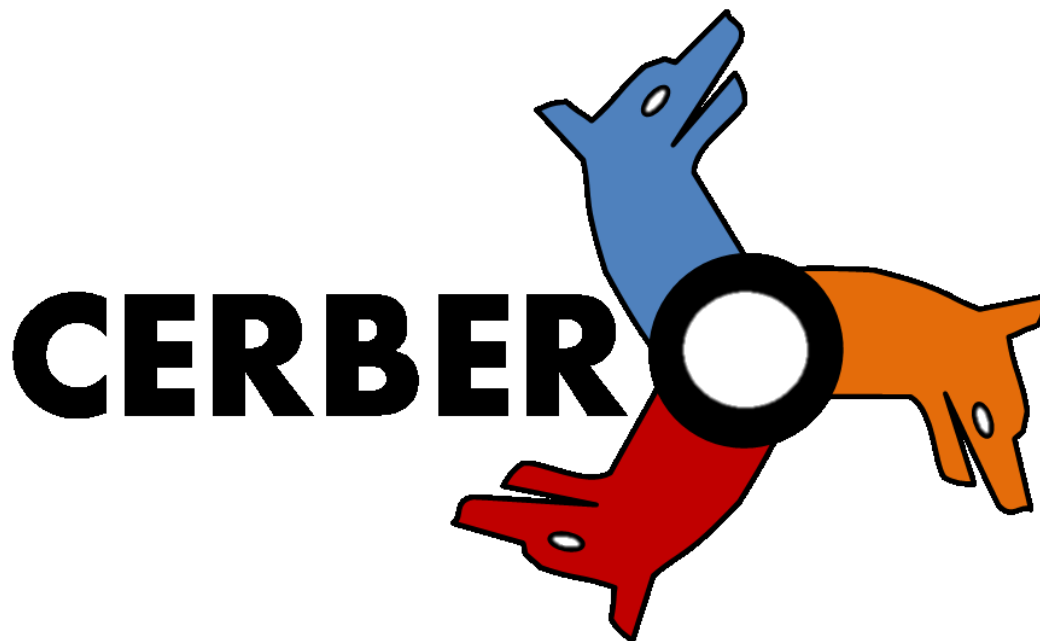
Project Nº: H2020-ICT-2016-1-732105

D1.7: Open Data Management Plan

Lead Beneficiary:	TNO
Workpackage:	WP1
Date:	
Distribution - Confidentiality:	Public

Abstract:

This document is the Open Data Management plan that describes which data will be made available to third parties and in which way.

WP1 – D1.7: Open Data Management Plan

Disclaimer

This document may contain material that is copyright of certain CERBERO beneficiaries, and may not be reproduced or copied without permission. All CERBERO consortium partners have agreed to the full publication of this document. The commercial use of any information contained in this document may require a license from the proprietor of that information.

The CERBERO Consortium is the following:

Num.	Beneficiary name	Acronym	Country
1 (Coord.)	IBM Israel – Science and Technology LTD	IBM	IL
2	Università degli Studi di Sassari	UniSS	IT
3	Thales Alenia Space España, SA	TASE	ES
4	Università degli Studi di Cagliari	UniCA	IT
5	Institut National des Sciences Appliquées de Rennes	INSA	FR
6	Universidad Politecnica de Madrid	UPM	ES
7	Università della Svizzera Italiana	USI	CH
8	Abinsula SRL	AI	IT
9	AmbieSense LTD	AS	UK
10	Nederlandse Organisatie Voor Toegepast Natuurwetenschappelijk Ondeerzoek TNO	TNO	NL
11	Science and Technology	S&T	NL
12	Centro Ricerche FIAT	CRF	IT

For the CERBERO Consortium, please see the http://cerbero-h2020.eu web-site.

Except as otherwise expressly provided, the information in this document is provided by CERBERO to members "as is" without warranty of any kind, expressed, implied or statutory, including but not limited to any implied warranties of merchantability, fitness for a particular purpose and non infringement of third party's rights.

CERBERO shall not be liable for any direct, indirect, incidental, special or consequential damages of any kind or nature whatsoever (including, without limitation, any damages arising from loss of use or lost business, revenue, profits, data or goodwill) arising in connection with any infringement claims by third parties or the specification, whether in an action in contract, tort, strict liability, negligence, or any other theory, even if advised of the possibility of such damages.

The technology disclosed herein may be protected by one or more patents, copyrights, trademarks and/or trade secrets owned by or licensed to CERBERO Partners. The partners reserve all rights with respect to such technology and related materials. Any use of the protected

Document Authors

The following list of authors reflects the major contribution to the writing of the document.

Name(s)	Organization Acronym
Nederlandse Organisatie Voor Toegepast Natuurwetenschappelijk Onderzoek TNO	TNO

The list of authors does not imply any claim of ownership on the Intellectual Properties described in this document. The authors and the publishers make no expressed or implied warranty of any kind and assume no responsibilities for errors or omissions. No liability is assumed for incidental or consequential damages in connection with or arising out of the use of the information contained in this document.

Document Revision History

Date	Ver.	Contributor (Beneficiary)	Summary of main changes
23.05.2018	0.1	TNO	Initial draft
24.05.2018	0.1	UNISS	Draft review
31.05.2018	0.2	TNO	Second draft
01.06.2018	0.3	TNO	Third draft
15.06.2018	0.4	TNO	Fourth draft, open data added
25.06.2018	0.5	TNO, UNISS, TASE, IBM	Final draft for review, comments UNISS, TASE and IBM incorporated

Table of contents

1. Executive Summary

This document is the Open Data Management Plan that describes which CERBERO use case data will be made available as open data to third parties and in which way. The used general approach and methodology (based on [BOMOD 2015]) for providing open data is described in deliverable D1.6 (version 1 of the document). This 2nd version of the of the document contains new insight and choices on how and which data of the use case demonstrators can and will be made available as open data. This includes adaptation of FAIR principals (Findable, Accessible, Interoperable and Re-usable) for the opening of CERBERO data, as defined in [FAIR] and [H2020 DMP]. The focus of this version of the open data plan is the data that have been made available along this year up until the M18 milestone, when first use case demonstrators are developed.

1.1. Structure of Document

The qualitative open data produced by each of the three CERBERO use cases is described in chapter 2. In chapter 3 the quantitative open data produced by the three use cases is described. Chapter 4 describes the selected process used to provide the open data. Chapter 5 describes the quality and security aspects involved.

1.2. Related Documents

[D1.6]

[BOMOD 2015]

[D2.4]

[D2.7]

[D8.3]

[FAIR]

[H2020]

2. Qualitative Open Data

The qualitative open data of the project can generically be classified in User Requirements, Technical Requirements and Results. Depending on the use case, different sets of these quantitative data sets will be provided.

By adapting the proposed method for open data publishing, all use cases benefit from an open data management policy of the project in which the published data generally is made available via:

- Project deliverables;
- Research publications;
- Social media;
- Project and partner web sites.

Deliverables

- The project deliverables provided as qualitative open data (see folder on CERBERO website: http://www.cerbero-h2020.eu/deliverables/) currently include:
    - D2.3 CERBERO Scenarios Description (Version 1)
    - D2.4 CERBERO Scenarios Description (Version 2)
    - D2.6 CERBERO Technical Requirements (Version 1)
    - D2.7 CERBERO Technical Requirements (Version 2)
    - D6.7 CERBERO Demonstration Skeleton (Ver 1)
    - D6.8 CERBERO Space Demonstrator (Ver 1)
    - D6.9 CERBERO Ocean Monitoring Demonstrator (Ver 1)
    - D6.10 CERBERO Smart Travelling Demonstrator (Ver 1)

Table 1 Deliverables as quantitative open data

	PE	ST	OM
Use Case Descriptions and Case-Dependent Requirements	@M9 D2.3 @M13 D2.4		
CERBERO Technical Requirements	@M9 D2.6 @M14 D2.7		
CERBERO demonstration skeleton	@M15 D6.7		
Demonstration Results	@M18 D6.8	@M18 D6.10	@M18 D6.9

-

Publications

Other publication provided by CERBERO as open data are also provided on the CERBERO website, see: http://www.cerbero-h2020.eu/publications/

Use case data

WP1 – D1.7: Open Data Management Plan

Links to already published use case specific qualitative open data are listed in Table 2.

Table 2 Qualitative use case specific open data (available at M18)2 Qualitative use case specific open data (available at M18)

PE	ST	OM
<not available yet>	CERBERO_Smart Travelling - Focus Group results V11 (final) – CRF and TNO: http://www.cerbero-h2020.eu/wp-content/uploads/2018/06/CERBERO_Smart-Travelling-Focus_-Group_results.pdf'http://www.cerbero-h2020.eu/wp-content/uploads/2018/06/CERBERO_Smart-Travelling-Focus_-Group_results.pdf'	YouTube movie "Enhancing underwater videos for reduced visibility situations": https://goo.gl/uj6QFr

2.1. *Self-Healing Systems for Planetary Exploration*

For the Space use case (see [D2.4]) different sets of qualitative data can be identified.

The possible qualitative data sets which could be provided as open data include:

- Technical requirements – Space use case:
  - List of technical requirements (robotic arm and motor control)
- Algorithm description to be used in space use case
  - Inverse kinematic algorithm
  - Motor control algorithm
  - Trajectory generation
- HW&SW partitioning

2.2. *Smart Traveling for Electric Vehicles*

For the Smart Traveling use case (see [D2.4]) different sets of qualitative data can be identified. Some of these data sets could be made available as open data for third parties.

The possible qualitative data sets which could be provided as open data include:

- User requirements – Smart Traveling:
  - The list of requirements from the users of the driving simulator, used for the Smart Traveling use case.
- Technical requirements – Smart Traveling:
  - List of technical requirements, derived from the user requirements, which will be used for the development of the demonstrator (by applying the tools from the CERBERO toolchain).
- Results of (Italian and Dutch) Focus Group sessions for Smart Traveling use case – *already provided at M18*:

- o Views of different types of drivers (both conventional, hybrid, hybrid plug-in and fully electric cars);
- o Usage of electric vehicles for different types of trips;
- o Positive aspects a driver is confronted with when driving an electric vehicle;
- o Negative issues / obstacles a driver is confronted with when driving an electric vehicle;
- o Driver views on foreseen functionalities of future electric vehicles.

2.3. Ocean Monitoring

The Ocean Monitoring use case (see [D2.4]) will also publish some open data for research and development purposes with focus on outreach to both research and broader audiences.

For the Ocean Monitoring, the quantitative open data will include:

- Technical requirements – Ocean Monitoring:
  - o List of technical requirements for the Ocean Monitoring robot;
- Videos and images taken both from both surface and sub-sea situations – *initial video already provided at M18, see Table 2.Table 2*.

Suitable open data of the Ocean Monitoring use case can be videos and images taken both from both surface and sub-sea situations. Some of this imagery can be beneficially provided as open data via social media, such as YouTube, but also via traditional web sites including partner and project web sites.

- In particular, machine-enhanced footage that help communicate our information fusion methods, such as for object/obstacle detection, edge enhancements, and so forth, can be of particular interest for research communities to see in order to better understand for instance how both information fusion algorithms and marine robots work.

3. Quantitative Open Data

Quantitative data are extremely use case dependent and a common classification is not possible. Therefore, in the following sections each use case providers is proving a list of possible collectable data, to be integrated and specified in details along the progress of the project. Also, here the FAIR [FAIR] principles are applied.

Links to already published quantitative open data from the different use cases are listed in the table below.

Table 3 Quantitative use case specific open data (available at M18)3 Quantitative use case specific open data (available at M18)

PE	ST	OM
<not data available yet>	Open data used for the development of the M18 demonstrator: • MECA-DynAA interface data; • Route data (send from MECA to DynAA for calculation); • Route evaluation request (using itinerary model); • Route evaluation response (using itinerary model). See open data set: http://www.cerbero-h2020.eu/wp-content/uploads/2018/06/Meca-DynAA_open_data.zip	<not data available yet>

3.1. Self-Healing Systems for Planetary Exploration

For the Space use case (see [D2.4]) the following quantitative data could be provided:
- Data on the robotic arm and servomotors;
- Data on the motor characteristics;
- Data on the hardware platform (FPGAs, drivers, etc.);
- VHDL & SW implementations of robotic arm algorithms;
- Simulation models for robotic arm:
  - Forward and Inverse Kinematic model;
  - Trajectory generation algorithm;
  - 3D representation of the arm trajectory;
- VHDL & SW implementations of motor control algorithms.

Some of them will be available as open data (obfuscated code) or under NDA according to the needs of the project. These decisions will be taken as soon as the development of the demonstrators (WP6) will progress and according to the exploitation of the different collected data.

3.2. Smart Traveling for Electric Vehicles

For the Smart Traveling use case (see [D2.4]) the following quantitative data could be provided:

- Data on the battery and electric engine models (provided by TNO) for different simulation scenarios;
- Collected data from the simulator on different components of the simulator (engine, battery, car, control data from the user);
- Data of the configured routes for the use case (e.g. geographic coordinates, road information (type, width, tracks), elevation of the road segments, GPS information, 3D data of surroundings/environment (trees, buildings, …etc.), whether data, data of other simulated traffic) - *a first set is already provided at M18, see Table 3*;
- Data on the charging infrastructure (geographic locations and other info on the loading poles like capacity, availability and price);
- Data on users (like preferences, type of driver and other characteristics);
- Data on route calculations (planned possible routes) - *a first set is already provided at M18, see Table 3*;
- Data on route predictions (including estimates for time and energy consumption*) - a first set is already provided at M18, see Table 3;*
- Data on user advice (based on available routes and user preferences);
- Data on simulated trips (simulated car behavior, energy consumption, followed routes, etc.);
- Registered user behavior (behavior during simulated trips, including control actions and possibly eye/face tracking).

It still needs to be decided by CRF (and TNO/S&T) which of this data can be provided as open data and which of this data is sensitive (e.g. because of privacy) or confidential (as it is for example related to actual design of the CRF system).

These decisions will be taken as soon as the development of the demonstrators (WP6) will progress and according to the exploitation of the different collected data.

3.3. Ocean Monitoring

As indicated in section 2.3, the Ocean Monitoring use case (see [D2.4]) will publish open data for research and development purposes with focus on outreach to both research and broader audiences.

The data generated or collected for Ocean Monitoring can be both qualitative and quantitative. The qualitative data are mainly individual user and business requirements and market needs, along with individual user and business feedback. Quantitative material includes the metrics and other key performance indicators defined in the evaluation of the Ocean monitoring data that

can be used for measuring the project success. Qualitative data may be more widely available as open data providing it is suitably anonymised and is part of highlighting wider trends as otherwise it is more likely to be considered commercially sensitive. It can be useful in providing background on why and how certain quantitative data emerge.

The quantitative open data in the Ocean Monitoring use case can include:

- Videos and images of a seabed and marine flora and fauna taken from both surface and sub-sea situations including 360 degrees videos and images. Selected subsets of data can be included for open data.
- Machine-enhanced footage (e.g. object / obstacle highlight, edge enhancement)
- Publicly available data collections for automatic evaluation of object detection and tracking methods, image retrieval models, data fusion techniques, etc. They can be used for general testing purposes as they come with the ground truth data. Use-case specific data collections can be also created for the context of marine environment from existing publicly available collections by filtering them based on text, specific visual characteristics such as colour, texture, etc., or a combination of text and visual features. Such data collections may include filtered **ImageClef** and **MirFlickr** image collections for example. Resulting data sourced from the publicly available data can be shared.
- Geo-data from GPS regarding location information.
- Sonar data of the marine environment including sonar images of the seabed, depending on availability.
- Infrared and thermal images of the surface around the marine robot, depending on availability.
- Battery data including the power consumption characteristics, performance, temperature of cells and external temperature, optimal battery topology.
- Laser and ultrasound data of the encountered obstacles, for example, depending on availability.
- Autopilot data such as list of destinations, starting locations, speed, weather, etc.
- Interactions with the user which may include user behavior when remotely operating the robot, user satisfaction levels on a specific scale, user feedback when augmenting the visual image and refining the search for similar objects, eye-tracking data for improving the user interface.
- User characteristics like preferences, types of users, usage patterns, etc. Suitable and data protection compliant subsets can be provided as open data.
- Calculated optimal paths to the destination which can be later reused.
- Visual and textual features (Vector Space Model) corresponding to information objects such as images and videos.

As indicated within relevant items above, there will be decisions with regard to availability of data from image and video collection subsets (due to commercial sensitivity) and user data (due to privacy and security).These decisions will be taken and reviewed during the course of project and the development of the demonstrators.

3.4. Other Relevant Data

CERBERO framework development and integration has brought to the creation of other relevant data, which can be used by third parties to reproduce our experiments, for comparison purposes, or to run new experiments.

In Table 4 we provide an overview of this data. Components description has already been provided in D5.6 and goes beyond the scope of this document

Table 4 Availability of open data on CERBERO tools4 Availability of open data on CERBERO tools

Component	Open Data Availability
Verification Suite	List of input requirements (https://github.com/SAGE-Lab/robot-arm-usecase).
DynAA	<no open data available at M18 yet> A short demonstration video on the purpose and use of the (TNO) DynAA tool is available at: http://youtu.be/ZP6q9J5wX4khttp://youtu.be/ZP6q9J5wX4k

4. Process to provide the Open Data

In order to provide the open data to third parties, the relevant CERBERO partners will provide their open data to the project coordinator. Open data have been made available on the CERBERO website:

http://www.cerbero-h2020.eu/open-data/
http://www.cerbero-h2020.eu/deliverables/

Data have been classified and collected, as you could see in the previous sections of this deliverable, and in the following months the coordinator will make sure that CERBERO open data are included in all relevant open data hubs / registries.

The actual publishing technologies (like storage mechanisms and of APIs for accessing the data) will depend on the capabilities of the sharing platform, the requirements and preferences of potential data users, available resources and the structure and size of the data to be stored. Currently, we just collected and classified the data, right after the review (as data from demonstrators and tools will be finalized) we intend to take a decision on the publishing technologies (beside the CERBERO website) and access to the open data sets of CERBERO. The CERBERO website does not have any limits of space and moreover, being an Aruba service, we do not expect issues related to backup, etc.

Actions for providing and managing the open data are summarized in Table 5.

Table 5 Process to provide open data5 Process to provide open data

Preferences and requirements	
Collect preferences and requirements of potential data users.	
Determine preferred publishing mechanisms (flat files (e.g. RDF/XML) and/or database query options / API (e.g. SPARQL or JSON) for the available (and foreseen) open data sets). Depending on the data set and use case, the mechanisms can differ.	
Initiation phase:	
Definition of the open data sets (and indication of possible updates of this data) of the CERBERO partners.	Achieved as described in Sections 2 and 3
Definition of (sufficient) meta data describing the different data sets.	
Definition of (ontology) linkage to other data sets	
Select hosting platform (data repository) for storage of the open CERBERO data (e.g. (via) [OpenAIRE], [Re3data] or [EUOPENDP]), taking into account available data sets and preferred publishing mechanisms.	Identification has been performed already and we will be able to finalize selection as demonstration and proof of concepts activities will be finalized after the review meeting.
Determine additional data hubs / data registries for reference towards the CERBERO open data (to increase visibility of the data sets).	

WP1 – D1.7: Open Data Management Plan

Development phase	
Agreement with open data hosting & sharing platform provider for hosting of the CERBERO open data (and ensure that the open data will also be available after completion of the project)	The CERBERO website already contains a section with open data: http://www.cerbero-h2020.eu/open-data/
Possibly adapt meta data and publishing format of the open CERBERO data, depending on requirements defined by hosting platform and the available publishing mechanisms	
Attach an appropriate license on the provided open data (e.g. Creative Commons License)	Qualitative data, being the result of CERBERO public deliverable, do not require any specific license. Nevertheless, people intending to use the provided documents should properly cite them, provided that for each deliverable (even public ones) all rights are reserved (see "© 2017 CERBERO Consortium, All Rights Reserved.")
Collect the CERBERO open data sets, APIs and relevant meta data from CERBERO partners to be uploaded to the hosting platform(s)	Data collection has started and will continue during the project lifetime.
Execution phase	
Upload of the collected open data on the data hosting & sharing platform (which will be accessible by third parties) and possibly also on one or more open data registries / hubs and/or websites of involved CERBERO partners	Already collected data are available and partly accessible from third parties.
Execute procedure for updates of already stored data (e.g. using a versioning mechanism), received from the CERBERO partners	Versioning will be put in place if needed during the second project iteration.
Create and maintain references to the open data on other sites/directories, like [Re3data] (http://www.re3data.org/)	References and notifications will be created/performed as soon as the first release of the data will be fully available after the review meeting.
Inform [OpenAIRE] (https://www.openaire.eu/) on the published CERBERO open data sets (in case data set was not published in OpenAIRE itself)	

It is assumed that the total size of the open CERBERO data to be stored on the sharing platform will not exceed 3 TB.

5.   Quality and Security aspects

5.1. Allocation of resources

Within the project resources will be arranged to publish the generated open data of the CERBERO project via the agreed channels (e.g. the CERBERO website).

5.2. Data security

The open data will be stored on the CERBERO website and will be equipped with same security measures implemented for the website.

5.3. Ethical aspects

Most of the open data within the CERBERO project is not linked to people and does not have any related privacy issues. An exception is the Smart Travelling use case where human drivers are used to execute the simulation. Also, human participants were used in the focus group sessions.

For the focus group sessions consent was given by the participants in the Netherlands and Italy (which resulted in data set on electric vehicle drivers) to use the data within the project and to publish results to the public (as open data). The collected data was anonymized and reference to actual persons were replaced by fake names to protect the participants and avoid any privacy issues.

6. References

[D1.6]   Open Data Management Plan (version 1):
                             http://www.cerbero-h2020.eu/wp-content/uploads/2017/11/D1.6.pdf

[CERBERO 2017]   http://www.cerbero-h2020.eu

[D2.4]   CERBERO Scenarios Description (Ver. 2)

[D2.7]   CERBERO Technical Requirements (Ver.2)

[D8.3]   Innovation, Standardization and Exploitation plan (Ver. 1)

[BOMOD 2015]   BOMOD, Management and development model for open data, Michael van Bekkeum and
                             Sandra Struijker Boudier, April 2015, TNO, ISBN: 978-90-5986-459-7,
                             http://publications.tno.nl/publication/34616703/ATAycW/eckartz-2015-bomod.pdf

[EUOPENDP]   EU Open data portal, http://data.europa.eu/euodp/en/home/

[OpenAIRE]   OpenAIRE: EU-funded Open Access portal, https://www.openaire.eu/

[Re3data]   Re3data is a registry of open data repositories, http://www.re3data.org/

[FAIR]   FAIR principles: http://www.nature.com/articles/sdata201618

[H2020 DMP]   H2020 templates: Data management plan v1.0 – 13.10.2016:
                             http://ec.europa.eu/research/participants/data/ref/h2020/gm/reporting/h2020-tpl-oa-data-
                             mgt-plan_en.docx