Information and Communication Technologies (ICT) Programme

Project N^o: H2020-ICT-2016-1-732105

D1.6: 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 in which way.

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 technology and related material beyond the terms of the License without the prior written consent of CERBERO is prohibited.

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 Ondeerzoek 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
11.07-2017	0.1	TNO	Initial draft
18.09.2017	0.2	TNO	Second draft
29.09.2017	0.3	TNO, AmbieSense, TASE	Input of TASE and AmbieSense
09.10.2017	1.0	TNO, IBM, UNISS	Final version

Table of contents

1. Executive Summary

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

1.1. Structure of Document

In the document the general approach and methodology for providing open data is described in Chapter 2.

Chapter 3 describes the qualitative open data produced by each of the three CERBERO use cases (Self-Healing Systems for Planetary Exploration, Smart Traveling for Electric Vehicles and Ocean Monitoring).

Chapter 4 describes the quantitative open data produced by the three use cases.

Chapter 5 finally describes the process to provide the open data.

1.2. Related Documents

[BOMOD 2015]

[D2.3]

[D2.6]

[D8.3]

2. Introduction

CERBERO Project is taking part to the Pilot on Open Research Data. The Data Management Plan presented in this document will naturally evolve and gain precision and substance over the project lifespan.

The data generated or collected by the project are both qualitative and quantitative. The qualitative data are mainly requirements and market needs, whereas quantitative material includes the metrics and other key performance indicators (KPIs) defined in the evaluation of the data that can be used for measuring the project success.

Quantitative data includes also the data collected from the use cases. Whereas some data are open as data from the use cases, the data collected from system design, security analysis and commercial metrics and KPI data shall remain confidential in order to protect results to be commercially exploited in the future (see exploitation plan [D8.3]). Therefore, the project provides this Open Data Management Plan that describes which data will be made available to third parties in which way, for example by attaching an appropriate Creative Commons License and registering data repositories at the re3data service.

The necessary IT platforms to host relevant research data will be provided. The consortium will make sure dedicated sharing platforms are set up, where the research data will reside and coordinate access to these platforms. The consortium will establish features like versioning of files, perform regular data backups and make sure the data will be available, also beyond the project lifetime.

For deciding if and how to provide open data, the TNO developed methodology BOMOD (see [BOMOD 2015]) can be used. In the BOMOD methodology a management and development model for open data is proposed, which can be used to consider all relevant aspects of opening data to thirds parties. For deciding on opening data, the open data decision tree of BOMOD can for example be applied (see Figure 1).
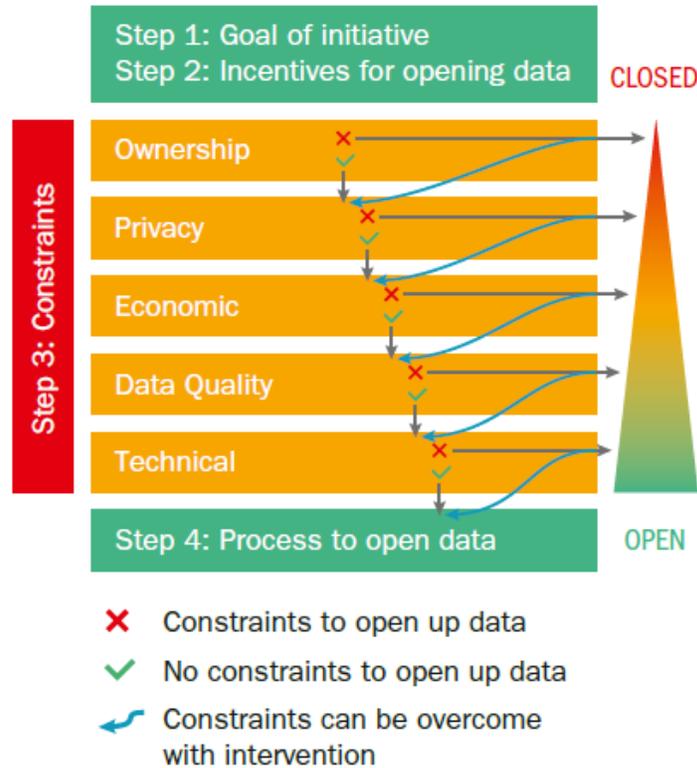
Figure 1: Open Data Decision Tree

When it is clear which data can be shared as open data, the project can follow the BOMOD methodology to open the data, which is composed of the following steps:

<u>Step 1: Goal of initiative</u>

Identification of the goal of data sharing.

<u>Step 2: Incentives to open data</u>

Identification of the incentives for individual stakeholders.

<u>Step 3: Constraints</u>

In this step five categories are used to structure barriers to open data:

- Ownership
- Privacy
- Economic
- Data quality
- Technical barriers

Once a constraint is identified in one of these categories, possible interventions need to be determined to overcome the identified barrier.

Examples of possible interventions are:

- Technical interventions:
    o Access control
    o Aggregation or anonymization of data

- o Improve data quality
- o Publish metadata
- o Support data standards
- Organizational interventions:
  - o Trust mechanisms
  - o Successful pilot projects
  - o Sharing of best practices
  - o Cultural change programs
  - o Collaboration mechanisms

Step 4: Process to open data

In step the data set is actually published and the identified technical and organizational interventions will be executed.

As defined in the BOMOD methodology, it will be important to provide sufficient meta data on all provided open data and define some quality aspects (like completeness and accuracy). This will help the third parties on finding and being able to actually use the open data.

Apart from the process on the steps how to open data, the BOMOD methodology also gives guidelines on the following aspects:

- **Governance** (which includes definition of roles and working group, decision making process and participations in working groups)
- **Finance** (to ensure continuity and realize value from the data, it is important that a sustainable financial model and sound business case is maintained).
- **Tactics** for opening data will include steps on following aspects:
  - o Community building (ensuring that all stakeholders are involved and take part in the data sharing)
  - o Data re-use (ensuring that the data can be found by potential users, using different data registries / repositories)
  - o Data model (to make data better understandable and usable, including the linkage of the data to other data sets)
- **Operations** (including the publishing of new data as well as changing already published data)
  - o Preferences and requirements (of the users of the open data)
  - o Initiation phase (based on preferences and requirements decide what data to publish or modify)
  - o Development phase (prepare selected data, including linkage, metadata, how to publish (support of queries or download only))
  - o Execution phase (actual publishing of the data)

These guidelines can help the project with the actual implementation of a facility for providing open data from the different use cases.

3. 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.

3.1. *Self-Healing Systems for Planetary Exploration*

For the Space use case (see [D2.3]) 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

3.2. *Smart Traveling for Electric Vehicles*

For the Smart Traveling use case (see [D2.3]) 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 form 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:
    - Views of different types of drivers (both conventional, hybrid, hybrid plug-in and fully electric cars);
    - Usage of electric vehicles for different types of trips;
    - Positive aspects a driver is confronted with when driving an electric vehicle;
    - Negative issues / obstacles a driver is confronted with when driving an electric vehicle;
    - Driver views on foreseen functionalities of future electric vehicles.

3.3. *Ocean Monitoring*

The Ocean Monitoring use case (see [D2.3]) 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:
    - List of technical requirements for the Ocean Monitoring robot;
- Videos and images taken both from both surface and sub-sea situations.

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. Thus, as per the proposed method for open data publishing above, the use case already benefits from an open data management policy in which the published data generally is made available via:

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

4. 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.

4.1. Self-Healing Systems for Planetary Exploration

For the Space use case (see [D2.3]) 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.

4.2. Smart Traveling for Electric Vehicles

For the Smart Traveling use case (see [D2.3]) 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);
- 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);
- Data on route predictions (including estimates for time and energy consumption);
- 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.

4.3. Ocean Monitoring

As indicated in section 3.3, the Ocean Monitoring use case (see [D2.3]) will publish open data for research and development purposes with focus on outreach to both research and broader audiences.

The quantitative open data will include:

- Machine-enhanced footage (e.g. object / obstacle detection, edge enhancements).

As soon as the development of the demonstrators (WP6) will progress, and according to the exploitation of the different collected data, a more detailed list of data will be provided.

5.   Process to provide the Open Data

In order to provide the open the data to third parties, the relevant CERBERO partners will provide their open data to the project coordinator. The coordinator will set up in the following months (by M18 when we are closing the first project iteration) a data hosting and sharing solution, accessible for thirds parties to host the open CERBERO data. The project coordinator will arrange that data can be hosted on the selected sharing platform and ensure that references to the open data are included in all relevant open data hubs / registries.

The actual publish technologies (like storage mechanisms and of APIs for accessing the data) will depend on the capabilities of the sharing platform and the requirements and preferences of potential data users and the structure and size of the data to be stored. During the project (when data sets become available) a decision will be made on the publishing technologies used for the storage of and access to the open data sets of CERBERO. One of the possibilities that we are considering is to host all the data on the CERBERO website. In the case of opting for such a solution, we will not have any limits of space and moreover, being an Aruba service, we will not have issues related to backup, etc.

The steps to be executed for providing and managing the open data sets will include:

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 n 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;
- 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;
- Determine additional data hubs / data registries for reference towards the CERBERO open data (to increase visibility of the data sets).

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);
- 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);
- Collect the CERBERO open data sets, APIs and relevant meta data from CERBERO partners to be uploaded to the hosting platform(s).

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;
- Execute procedure for updates of already stored data (e.g. using a versioning mechanism), received from the CERBERO partners;
- Create and maintain references to the open data on other sites/directories, like [Re3data] (http://www.re3data.org/);
- 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.

6.   References

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

[D2.3]   CERBERO Scenarios Description (Ver. 1)

[D2.6]   CERBERO Technical Requirements (Ver.1)

[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/