Information and Communication Technologies (ICT) Programme

Project Nº: H2020-ICT-2016-1-732105

D1.3: Open Data Management Plan (Final version)

Lead Beneficiary:	TNO
Workpackage:	WP1
Date:	24/12/2019
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.

Disclaimer

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 Onderzoek 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.

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

Document Revision History

Date	Ver.	Contributor (Beneficiary)	Summary of main changes
09.09.2019	0.1	TNO	Initial draft
05.12.2019	02	UNISS	Review and Contribution
09.12.2019	03	AS, TNO, TASE	Update
12.12.2019	04	TNO	Version for internal review
16.12.2019	04	AS	Update
17.12.2019	05	TNO, CRF	Final version for review
20.12.2019	06	TNO, UNISS, TASE	Final updates
23.12.2019	10	TNO	Final version 1.0
24.12.2019	11	IBM, TNO	Final version 1.1

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 final version of the of the document contains new insight and choices on how and which data of the use case demonstrators have been 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 until the M36 milestone, when the final use case demonstrators are developed.

1.1. Structure of Document

Here follows the structure of the deliverable.

- 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]

[D2.1]

[D2.2]

1.3. Requirements

As specified in D2.2, the project set a requirement (CERBERO-0010) to provide Open Data Management Plan. Additional requirements were set like “RC2: Open access to as much tools and data as possible”, “IC1: Dissemination of results in all relevant industry communities” and “IC2: Technical education” for which publication of open data are be beneficial.

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 are available on the website CERBERO website: http://www.cerbero-h2020.eu/deliverables/

  At M36 just already approved deliverables are included, final ones will be published after project review and deliverables approval.

Table 1 Deliverables as quantitative open data

	PE	OM	ST
CERBERO Scenario Description	@M25 D2.1		
CERBERO Technical Requirements	@M26 D2.2		
CERBERO demonstration skeleton	@M32 D6.1		
Demonstration Results	@M36 D6.2	@M36 D6.3	@M36 D6.4

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 and Tools

Links to qualitative open data from use cases and tools are listed in Table 2 and discussed in Sections 2.1, Sections 2.2 and Sections 2.3.

Table 2 Qualitative open data (available at M36)

Data set / Tool / Use Case	**Findable** (link / website, meta data)	**Accessible** (metadata and data understandable for human / machine, trusted repository)	**Interoperable** (data format used)	**Re-usable** (clear usage license)
Focus group results ST UC	CERBERO website: https://www.cerbero-h2020.eu/wp-content/uploads/2018/06/CERBERO_Smart-Travelling-Focus_-Group_results.pdf	Explanation in file	PDF file	Open to use
Video Smart Travelling use case	*(Will be provided via the CERBERO website after video is completed. Planned date: end of January 2020)*	Explanation in video	MPEG file	Open to use
PiSDF Dataflow test (Preesm & Spider) applications and datasets.	GitHub: https://github.com/preesm/preesm-apps	Trusted repository	Applications: XML, C code, Eclipse Projects, text files. Datasets: Standard Images and videos formats (raw yuv, ppm.)	Applications: CeCILL-C Datasets: Creative Common or MIT
Robotic Arm qualitative data	https://www.cerbero-h2020.eu/wp-content/uploads/2019/12/Open-Data-PE.zip	Explanation in file	DOCX files	Open to use
Trajectory Generator for Robotic Arm Applications	https://www.cerbero-h2020.eu/wp-content/uploads/2019/12/Trajectory-Generator.zip	Code compliant with *Robotic Arm Workspace Description*	MatLab file	3-Clause BSD Licence

DLS algorithm implementation	https://www.cerbero-h2020.eu/wp-content/uploads/2019/12/DLS-implementation.zip	Code of the DLS Algorithm and results exploration for different trajectories and iterations.	MatLab files	3-Clause BSD Licence
Video of the robotic arm showing the configuration and implementation of the use case in the lab.	https://www.cerbero-h2020.eu/wp-content/uploads/2019/12/VideoRobotArm.zip	Explanation in video	MPEG file	Open to use
Enhancing underwater videos for reduced visibility situations	YouTube movie: https://goo.gl/uj6QFr%20	Explanation in video	MPEG video	Open to use
Ocean Monitoring image filter optimization data set	CERBERO website: https://www.cerbero-h2020.eu/wp-content/uploads/2019/12/UseCaseOceanMonitoring.zip	Readme with description of data	CSV file (xz-compressed)	Creative Commons Attribution Share-Alike

2.1. Self-Healing Systems for Planetary Exploration

For the Planetary Exploration use case (see [D2.1]) different sets of qualitative data are identified. Some of these data sets are made available as open data for third parties.

The qualitative data sets which will be provided as open data include:

- Technical requirements – Planetary Exploration:
  - List of technical requirements (see [D2.2])
- Motion planning algorithms:
  - Nelder-Mead optimization;
  - Damped Least Squares method.
  - See:
    - Trajectory Generator: https://www.cerbero-h2020.eu/wp-content/uploads/2019/12/Trajectory-Generator.zip
    - DLS Matlab Implementation: https://www.cerbero-h2020.eu/wp-content/uploads/2019/12/DLS-implementation.zip
- Robotic arm: https://www.cerbero-h2020.eu/wp-content/uploads/2019/12/Open-Data-PE.zip
- Video of the robotic arm showing the configuration and implementation of the use case in the lab: https://www.cerbero-h2020.eu/wp-content/uploads/2019/12/VideoRobotArm.zip
- Story board with the definition of Planetary Exploration use case, see [D2.1].

2.2. Ocean Monitoring

The Ocean Monitoring use case (see [D2.1]) 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 provided quantitative open data includes:

- Technical requirements – Ocean Monitoring:
  - List of technical requirements for the Ocean Monitoring robot (see [D2.2]);
- Optimization data set for adaptive image enhancement
  - Training data for a range of image filtering techniques, to build a model for selecting the optimal combination of techniques based on quickly-assessed image characteristics, like brightness and contrast, see: https://www.cerbero-h2020.eu/wp-content/uploads/2019/12/UseCaseOceanMonitoring.zip;
- Videos and images taken both from both surface and sub-sea situations – *initial video already provided at M18,* see: https://goo.gl/uj6QFr.

Suitable open data of the Ocean Monitoring use case are videos and images taken both from both surface and sub-sea situations. Some of this imagery are 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.

2.3. Smart Traveling for Electric Vehicles

For the Smart Traveling use case (see [D2.1]) different sets of qualitative data are identified. Some of these data sets are made available as open data for third parties.

The qualitative data sets which will 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, see [D2.1];
- 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), see [D2.2]
- Results of (Italian and Dutch) Focus Group sessions for Smart Traveling use case – *already provided at M18*:
  - 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.

  See set at: http://www.cerbero-h2020.eu/wp-content/uploads/2018/06/CERBERO_Smart-Travelling-Focus_-Group_results.pdf%E2%80%99
- Video of the driver simulator set up showing the configuration and implementation of the use case at the CRF driving simulator in Turin. Completion of video is planned in January 2020. Video link will be included in CERBERO website.
- Story board with the definition of Smart Travelling use case, included in [D2.1].

3. Quantitative Open Data

During the development of the use cases extensive sets of quantative data were produced. In the following sections all quantative data provided as open data sets are listed. Furthermore data sets of specific tool integrations are listed. Also, here the FAIR [FAIR] principles are applied.

Adopting the FAIR means that providers of the open data should not only have thought about how people would be able to find the data but also if people will understand what exactly is provided (e.g. by added meta data), if data can be understood by a human or machine (meta data / descriptions) and what will be needed to interoperate with the data (e.g. what format is used for the provided data set). Furthermore, it should be clear under what conditions the data can be used.

Links to the quantitative open data from the different use cases and tools are listed in Table 3.

Table 3 Quantitative open data (available at M36)

Data set / Tool / Use Case	**Findable** (link / website, meta data)	**Accessible** (metadata and data understandable for human / machine, trusted repository)	**Interoperable** (data format used)	**Re-usable** (clear usage license)
Dyna & MECA data of the Smart Travelling use case	CERBERO website: https://www.cerbero-h2020.eu/wp-content/uploads/2019/12/MECA-DynAA_open-data.zip	Readme with description of data	• Json format (data shared between MECA, HMI and DynAA) • PBF format (used open street map data)	Open to use, no licence restrictions.
CRF (SCANeR) driving simulator Smart Travelling use cases driving session	CERBERO website: https://www.cerbero-h2020.eu/wp-content/uploads/2019/12/CRF_DRIVING_SIMULATOR_open_data.zip	Readme with description of data	CSV file	Open to use, no license restrictions
List of Sample Trajectories for Robotic Arm tests	CERBERO website: https://www.cerbero-h2020.eu/wp-content/uploads/2019/12/Trajectory-Generator.zip	Readme with description of data	CSV file	Open to use, no licence restrictions.
Ocean Monitoring image filter	CERBERO website: https://www.cerbero-h2020.eu/wp-	Readme with description of data	CSV file (xz-compressed)	Creative Commons Attribution Share-Alike

optimization data set	content/uploads/2019/12/UseCaseOceanMonitoring.zip			
PREESM and AOW data for optimal software scheduling	CERBERO website: https://www.cerbero-h2020.eu/wp-content/uploads/2019/12/AOW_Preesm.zip	Readme with description of data. JSON files formats are extended version of formats described in D5.3 section 3.4. PREESM – AOW – DynAA connection	• JSON files for AOW input • JSON files of AOW output • CPLEX dat files for alternative AOW/CPLEX optimization input • PREESM standard data files • Conversion scripts	Open to use, no licence restrictions.
MDC and AOW data for optimal data flow merging	CERBERO website: https://www.cerbero-h2020.eu/wp-content/uploads/2019/12/MDC_AOW.zip	Readme with description of data. Examples of input/output files for the MDC profiler. Examples of output files for the MDC profiler enhanced with AOW.	• XDF XML-based format of data flow definitions • Text files of optimization results (.out) • XDF files of optimization results	Open to use, no licence restrictions.
Robotic arm quantitative data	CERBERO website: https://www.cerbero-h2020.eu/wp-content/uploads/2019/12/Open-Data-PE.zip	Readme with description of data	• DOCX files for technical descriptions and models explanation • PREESM files for model-based implementation • Python files for GUI and simulator	Open to use, no licence restrictions.

3.1. Self-Healing Systems for Planetary Exploration

For the Space use case (see [D2.1]) the following quantitative data is provided:

- Data on the robotic arm;
- Data on the hardware platform (FPGAs, drivers, etc.);
- VHDL & SW implementations of robotic arm algorithms;
- Forward kinematic model;
- Energy model of the robotic arm in vacuum conditions;
- GUI for user command and simulation.

The data sets are available at:

- Robotic arm: https://www.cerbero-h2020.eu/wp-content/uploads/2019/12/Trajectory-Generator.zip and https://www.cerbero-h2020.eu/wp-content/uploads/2019/12/Open-Data-PE.zip
- Planetary exploration: https://www.cerbero-h2020.eu/wp-content/uploads/2019/12/Open-Data-PE.zip
- DLS implementation: https://www.cerbero-h2020.eu/wp-content/uploads/2019/12/DLS-implementation.zip

3.2. Ocean Monitoring

As indicated in section 2.2, the Ocean Monitoring use case (see [D2.1]) publishes 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.

The quantitative open data in the Ocean Monitoring use case includes:

- A large scale optimization data set with a total of over 84 million records. These are based on 4331 video frames taken from 128 videos. The videos/images of seabed and marine flora, fish, and surface images including ships and marine life, see: https://www.cerbero-h2020.eu/wp-content/uploads/2019/12/UseCaseOceanMonitoring.zip ;
- Machine-enhanced footage (e.g. object / obstacle highlight, edge enhancement) – this was already made available for viewing on YouTube, see: https://goo.gl/uj6QFr.
- 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. We have used such publicly available collections in developing and improving our methods. These

collections include **ImageClef** and **MirFlickr** image collections for example as well as more specific underwater image benchmarking sets. These are available at:

- MIRFlickr: https://press.liacs.nl/mirflickr/#sec_download
- ImageCLEF: https://www.imageclef.org/datasets
- Underwater image enhancement benchmark: https://li-chongyi.github.io/proj_benchmark.html

There were some other types of data that were originally also considered as possibilities for open data that would be reviewed during the course of the project and development of demonstrators. For example, these included geo-data, sonar data and infrared and thermal images. Geo data from GPS regarding location information or optimal navigational paths was not identified as an issue during customer-discovery interviews and therefore wasn't a priority task. Similarly, this was the case for the autopilot data (e.g. destinations, locations, speed, whether) as autonomous mode was not a priority issue along with sonar data, infrared and thermal images. Battery data such as power consumption characteristics, performance, cell temperature and battery topology are important and can sometimes be an issue. However, we were advised against focusing on this aspect as these were much more complicated and inter-related features than might have seemed earlier on. It would make more sense to focus on this aspect if an autonomous vehicle was a priority need. Laser and ultrasound data for potential obstacles is an important area, even if not an immediate priority, and we have noted it for subsequent future work. Temperature and humidity information was considered more important in some surface OM scenarios. These along with any corresponding image and videos are subject to commercial sensitivities. Similarly, in addition to commercial considerations, due data privacy and security concerns user data such as user characteristics, preferences, usage patterns, system interactions are omitted.

We contribute to Open Data in a manner that takes into consideration commercial sensitivities and user privacy/security and have made available a very large-scale data set for optimization use. The subsection below provides details on this.

3.2.1. Optimization data set

The optimization data set consists of open data that is used to build a model for runtime adaptation between a range of different video enhancement processes. This model is used to select weights to use a combination of enhancement filters, depending a range of fixed but measurable image characteristics of the source image.

The current optimization data set consists of:

- Images extracted from a range of 128 videos, at every 60 frames (i.e., approximately every other key frame)
- Total number of source images: 4,331

- For each image, an image quality assessment was conducted for every combination of seven different enhancement filters, plus the original source image, with weights at a resolution of 0.1. Therefore, there are eight possible weights, all between 0.0 and 1.0 at steps of 0.1, which sum to 1.0. There are 19,448 valid combinations of weights, and these were computed using a "stars and bars" method, see: https://en.wikipedia.org/wiki/Stars_and_bars_(combinatorics).
- Total data set size: 84,229,288 records (19,448 combinations for each of 4,331 images)
- Format: xz-compressed CSV
- Static values per-frame: pixels (width multiplied by height), brightness, and entropy (a standard proxy for contrast)
- Data file columns: filename, pixels, brightness, entropy, w1, w2, w3, w4, w5, w6, w7, w8, keypoints.

The goal is to provide an effective model for selecting w1 … w8, given the number of pixels, brightness, and entropy, to generate an optimal number of keypoints, where, for now, we are working on the optimal number of keypoints being the mean for that specific frame. The techniques used for w1 … w8, and which is the original image, are not currently public, but do not affect the nature of the optimization task.

3.3. Smart Traveling for Electric Vehicles

For the Smart Traveling use case (see [D2.1]) the quantitative data provided is data from scenario execution shared between MECA, DynAA and HMI:

- Data on route calculations (planned possible routes);
- Data on route predictions (including estimates for time and energy consumption);
- OpenStreetMap data used in SCANeR, HMI and MECA from Geofabrik (see: https://www.geofabrik.de/geofabrik/openstreetmap.html).
- Data on primary driving controls (pedals, steering wheel)
- Parameters of vehicle dynamics (e.g. torque, lateral acceleration, speed, …)
- Latitude and longitudinal positions
- Registered data of drowsiness flag provided by the simulator during the driving

The data sets are available at:

1. MECA and DynAA : https://www.cerbero-h2020.eu/wp-content/uploads/2019/12/MECA-DynAA_open-data.zip and
2. CRF (SCANeR) driving simulator: https://www.cerbero-h2020.eu/wp-content/uploads/2019/12/CRF_DRIVING_SIMULATOR_open_data.zip

3.4. Software Scheduling Optimization (PREESM <-> AOW)

Optimal software scheduling on different hardware platforms can be used in the context of all CERBERO use cases. During CERBERO project, PREESM <-> AOW tool connection has beenevaluated in order to perform optimal software scheduling. These results are not demonstrated on any particular use-cases, because from one hand such demonstration requires many additional preparations, such as description of applied algorithm as SDF graphs and timing evaluation of corresponding software on target hardware platform, and from other hand PREESM <-> AOW connection results are still being evaluated up to M36. The paper of results of this evaluation is still in preparation, however the input data used during this evaluation, as part of output data (optimal scheduling of different SDF graphs on different hardware platforms) are available.

The data includes:

- AOW input files in JSON format. Each JSON file includes:
    - SDF graph of particular application.
    - Target hardware architecture description.
    - Timing evaluation, that is description on how many times specific software actor requires to run on specific processing element of target hardware.

  JSON file formats described in json_aow_input section of Readme file and are extended version of formats described in D5.3 section 3.4. PREESM – AOW – DynAA connection.
- CPLEX .dat files for alternative AOW/CPLEX optimization input. For each specific optimization cases there are 3 .dat files:
    - xxx_app.dat – represent SDF graph
    - xxx_archi.dat – represent target hardware architecture description
    - xxx_app_srdag – represent list of actor precedencies for single rate SDF graph for the same application

  CPLEX .dat file contents described in dat_aow_input section of Readme file.
- AOW output files in JSON format. JSON file formats described in json_aow_output section of Readme file and are extended version of formats described in D5.3 section 3.4. PREESM – AOW – DynAA connection.
- All Preesm data required to generate the AOW input files:
    - A standard Preesm project with architecture descriptions in “Archi” folder, SDF graphs of sample applications in “Algo” folder, source code of these applications in “Code”, scenarios with timings in “Scenarios” folder, workflows to generate the intermediate files, the timings, and the code from AOW output in “Workflows” folder.
    - A “generatedData” folder containing intermediate files such as flat version of SDF graphs and single rate SDF graphs, the timings, and finally the schedules computed by preesm for comparison.
    - A “Readme” file and various scripts to generate all the aforementioned intermediate files and to execute the Preesm workflow from AOW output file.
- Scripts to convert the aforementioned Preesm data to AOW json and dat input files. These scripts are described in “preesm_to_aow_scripts” section of Readme file.

All inputs provided for two different hardware platforms:

- For regular x86 platform (regular file names)
- For embedded Odroid platform (xxx_odroid file names)

The sample applications are the following (with different scale options):

- Prediction
- Sift (sift_4K, sift_medium, sift_normal)
- Squeezenet
- Stereo (stereo, stereo_plarge)
- Training

These applications correspond to well-known image processing applications, standard or with neural networks.

Dummy applications are also used for scheduling compatibility testing and debugging (BroadcastTest14, BroadcastTest24, CacheTest14, CacheTest24).

For now, AOW output files are provided only for prediction, stereo and sift_normal applications.

See data set at: https://www.cerbero-h2020.eu/wp-content/uploads/2019/12/AOW_Preesm.zip

3.5. Data Flow Merging Optimization (MDC <-> AOW)

Preliminary analysis on the integration of AOW and MDC has been carried out. The integration will enable faster analysis of large design space to determine the optimal merging configuration. No use cases are taking benefits at the moment of this integration, which has still to be assessed on a large scale. Nevertheless, data used during these first integrations steps are available. Online you will find a directory with three folders

1. dataflows: contains input dataflow specifications described according to the RVC-CAL language:
    a. net - contains the .xdf (xml dataflow) description of networks implementing the different kernels;
    b. cal - contains the .cal description of the atomic actors.
2. characterization: contains data related to low level profiling (area, power, frequency) of actors and dataflow networks, data have been retrieved for ASIC implementations at 90nm.
    a. - area_per_actor.csv
    b. - pwr_per_actor.csv
    c. - timing_per_dataflow.csv
3. dse_example:
    a. - dse_profiled.csv - example of the design space exploration (DSE) performed with MDC in brute force mode. Data refer to the UC1 of the paper "DSE and Profiling of Multi-Context Coarse-Grained Reconfigurable Systems" [PALUMBO 2013].
    b. aow_output – contains the DSE outcomes obtained using AOW over the same example of dse_profiled.csv

See data set at: https://www.cerbero-h2020.eu/wp-content/uploads/2019/12/MDC_AOW.zip

4. Process to provide the Open Data

In order to provide the open data to third parties, the relevant CERBERO partners have 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 provided via the CERBERO website as open data.

The actual publishing technologies (like storage mechanisms and of APIs for accessing the data) depended on the requirements and preferences of potential data users, available resources and the structure and size of the data to be stored. Given the number of sets available and the limited size of the provided data sets, we have chosen to only publish the data sets via the CERBERO website. 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 4.

Table 4 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.	Given the limited number of sets available, the data sets themselves are provided as separate files.
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.	If needed explanation is added inside the provided data set.
Definition of (ontology) linkage to other data sets	No specific linkage is provided.
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 the CERBERO website is selected as the location to store and provide the open data sets.
Determine additional data hubs / data registries for reference towards the CERBERO open data (to increase visibility of the data sets).	No additional data registries were selected (given the limited number of sets available)
Development phase	
Agreement with open data hosting & sharing platform provider for hosting of the CERBERO open data (and ensure	The CERBERO website already contains a section with open data: http://www.cerbero-h2020.eu/open-data/

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	For publication on the CERBERO website there were no explicit requirements on meta data or formats.
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 was mainly collection from the final M36 demonstrators.
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	All collected data is available on the CERBERO website and accessible by third parties.
Execute procedure for updates of already stored data (e.g. using a versioning mechanism), received from the CERBERO partners	No versioning was required. Available sets are updated if needed until the final completion of all demonstrators.
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 this document is formally approved.
Inform [OpenAIRE] (https://www.openaire.eu/) on the published CERBERO open data sets (in case data set was not published in OpenAIRE itself)	

The total size of the open CERBERO data stored is well below the initial maximum set at 3 TB.

5. Quality and Security aspects

5.1. Allocation of resources

Within the project resources were 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 stored on the CERBERO website is 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 real 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]	http://www.cerbero-h2020.eu
[D2.1]	CERBERO Scenarios Description (Final version)
[D2.2]	CERBERO Technical Requirements (Final version)
[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
[PALUMBO 2013]	F. Palumbo, C. Sau and L. Raffo, "DSE and profiling of multi-context coarse-grained reconfigurable systems," *2013 8th International Symposium on Image and Signal Processing and Analysis (ISPA)*, Trieste, 2013, pp. 744-749. doi: 10.1109/ISPA.2013.6703836