software framework for runtime-Adaptive and secure deep Learning On Heterogeneous Architectures

Project Number 780788

Project Acronym ALOHA

D8.7	Open Data Management Plan – First update		
Work Package:	WP8	**Lead Beneficiary:**	UNICA
Type:	Report	**Dissemination level:**	Public
Due Date:	30/06/2019	**Delivery:**	25/06/2019
Version:	1.0		

Brief description:

This document presents the second version of the ALOHA Open Data Management Plan. It describes the policy adopted by the ALOHA Consortium regarding the management of data produced and/or used in the project, detailing what research data the project will collect/generate, whether and how they will be exploited or made accessible for verification and re-use, and how they will be curated and preserved.

This document is the first update of deliverable D8.6, submitted in month M6. Specifically, it adds some details to ALOHA research data, based on findings and decisions made by the Consortium in the period from June 2018 to June 2019.

Deliverable Author(s):

Name	Beneficiary
Daniela Loi	UNICA
Paolo Meloni	UNICA

Deliverable Revision History:

Reviewer Beneficiary	Issue Date	Version	Comments
UNICA	06/05/2019	V0.1	Released of the Table of Content and introduction
UNICA	31/05/2019	V0.2	Draft version including all contributions from partners
PLURIBUS ONE	05/06/2019	V0.3	Review on user community data and information
UNICA	13/06/2019	V0.4	Final draft
PLURIBUS	25/06/2019	V0.5	Review
IBM	25/06/2019	V0.6	Review
UNICA	25/06/2019	V1.0	Final version including all reviewers' suggestions

Disclaimer

This document may contain material that is copyright of certain ALOHA beneficiaries, and may not be reproduced, copied, or modified in whole or in part for any purpose without written permission from the ALOHA Consortium. The commercial use of any information contained in this document may require a license from the proprietor of that information. The information in this document is provided "as is" and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability.

The ALOHA Consortium is the following:

#	Participant Legal Name	Acronym	Country
1	STMICROELECTRONICS SRL	ST-I	Italy
2	UNIVERSITA' DEGLI STUDI DI CAGLIARI	UNICA	Italy
3	UNIVERSITEIT VAN AMSTERDAM	UVA	Netherlands
4	UNIVERSITEIT LEIDEN	UL	Netherlands
5	EIDGENOESSISCHE TECHNISCHE HOCHSCHULE ZUERICH	ETHZ	Switzerland
6	UNIVERSITA' DEGLI STUDI DI SASSARI	UNISS	Italy
7	PKE HOLDING AG	PKE	Austria
8[1]	**CA TECHNOLOGIES DEVELOPMENT SPAIN SA**	**CA**	**Spain**
9	SOFTWARE COMPETENCE CENTER HAGENBERG GMBH	SCCH	Austria
10	SANTER REPLY SPA	REPLY	Italy
11	IBM ISRAEL - SCIENCE AND TECHNOLOGY LTD	IBM	Israel
12	SYSTHMATA YPOLOGISTIKIS ORASHS IRIDA LABS AE	IL	Greece
13	PLURIBUS ONE SRL	PLURIBUS	Italy
14	MAXQ ARTIFICIAL INTELLIGENCE, LTD (formerly MEDYMATCH TECHNOLOGY, LTD)	MaxQ-AI (formerly MM)	Israel
15[2]	**UNIVERSIDAD POMPEU FABRA**	**UPF**	**Spain**

---

[1] The participation of CA TECHNOLOGIES DEVELOPMENT SPAIN SA (CA) has been terminated on May 9th, 2019.
[2] UNIVERSIDAD POMPEU FABRA (UPF) has been added as new beneficiary on May 29th, 2019.

Table of Contents

Figures

Tables

1   Executive Summary

As a project participating in the Open Research Data (ORD) pilot in Horizon 2020, ALOHA delivered a Data Management Plan in the first 6 months of its lifecycle (Deliverable D8.6), outlining what research data will be findable, accessible, interoperable and re-usable (FAIR) during and after the end of the project. In developing this plan, the Consortium takes into account the need to balance openness and protection of scientific information, commercialisation and Intellectual Property Rights (IPR), privacy concerns and security, as stated in Guidelines on FAIR Data Management in Horizon 2020 published by the European Commission[3].

This document provides **an updated version of the ALOHA Open Data Management Plan presented in deliverable D8.6, by introducing new data for internal use and for exploitation purposes**. It takes into account the experience gained so far regarding data and the progress of the work made by the Consortium in the period from June 2018 to June 2019.

The plan describes the open research data strategy to be implemented and specifies:

- what digital research data the project will collect, process and/or generate,
- which methodology and standards will be applied,
- which data will be kept confidential and which will be made openly available to Third Parties for verification and re-use and,
- how these data will be curated and preserved for future researchers.

The content of this document reflects the exploitation and Intellectual Property Rights (IPR) requirements as defined in the Consortium Agreement.

This Data Management Plan is not a fixed document. It will evolve and gain more precision and substance during the next months. The last version of this plan will be submitted at the end of the project:

- D8.8 Open Data Management Plan – Final update [M36]

1.1   Advances in comparison to previous version

The following updates exist in this deliverable in comparison to D8.6:

- Minor updates to the dataset for ALOHA toolflow development (section 4.2.1, Table 6). New datasets and information are highlighted in bold font;
- Minor updates to the Speech datasets for the Smart Industry Use Case (section 4.2.3, Table 9);
- Addition of User Community Information to the Data for dissemination, communication and exploitation purposes (section 4.3, Table 13).

---

[3] http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

1.2 Acronyms and abbreviations

Acronym	Meaning
CDNN	Convolutional Deep Neural Network
CNN	Convolutional Neural Network
CT	Computed Tomography
DICOM	Digital Imaging and COmmunications in Medicine
DNN	Deep Neural Network
EC	European Commission
ERC	European Research Council
FAIR	Findable, Accessible, Interoperable and Re-usable
IPR	Intellectual Property Rights
ONNX	Open Neural Network Exchange Format
ORD	Open Research Data
SoC	System on Chip

7

2   Open access policy to research results in H2020

The policy on open research data in H2020, also reported in Article 29.3 of the ALOHA Grant Agreement, requires projects participating in the ORD pilot to:

- deposit research data into a research data repository of their choice;
- take measures to enable third parties to access, mine, exploit, reproduce and disseminate the research data free of charge;
- provide information - via the repository - about tools and instruments for validating the results, e.g., specialised software or software code, algorithms and analysis protocols. Where possible, these tools or instruments should be provided.
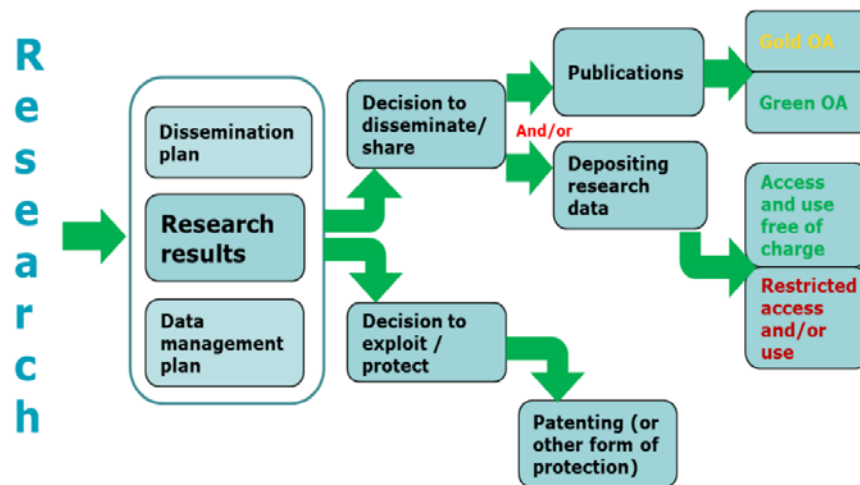

Figure 1: H2020 Open access policy to publications and open research data[4]

According to the European Commission guidelines, research data "refers to information, in particular facts or numbers, collected to be examined and considered as a basis for reasoning, discussion, or calculation. In a research context, examples of data include statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images." They can be underlying data from publications (i.e., data needed to validate the results presented in scientific publications), curated data and/or raw data. Research data produced during a H2020 project should be identified in the Data Management Plan and made as openly available as possible for verification and re-use. However, participating in the ORD pilot does not necessarily mean opening up all research data. As showed in Figure 1, beneficiaries can also choose to keep selected datasets (or even all data) closed if access to them would lead to a risk for the project's goals. In this case, the Data Management Plan should contain the reasons for not giving access to data or for opting out of the pilot.

The ALOHA Consortium will follow the strategy showed in Figure 1 and all the data will be managed in line with the Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020.

---

[4] Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020, Version 3.2, 21 March 2017. http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

3 Research results in ALOHA

The ALOHA Consortium will collect and generate mainly three categories of data:

- data for project management purposes,
- data collected and used to conduct internal development, evaluation and analysis of the ALOHA toolflow,
- data for knowledge transfer such as scientific publications, public project deliverables, communication and dissemination material.

A complete list of all data to be collected and created during the project is shown in the tables below. Additional information about each set of data will be provided in the following sections.

Table 1: Project management data

Type of data	Format
Agenda and meeting minutes and presentations	.doc, .ppt
Templates of deliverables, presentations, posters, minutes	.doc, .ppt
Project Logo	vector files, jpg, png
Deliverables and reports to be uploaded in the Participant Portal	.pdf
Financial information to be provided to the EC	.xls, .pdf

Table 2: Data for internal use

Type of data	WPs	Partners involved
Datasets to be used during the ALOHA toolflow development	WP2	ALL
Target architecture description	ALL	UNICA, ETHZ, ST-I
Datasets for the implementation of the three use cases	WP5	PKE, REPLY, MAXQ, SCCH
Deep Neural Network (DNN) topologies	WP2-4	ALL

Table 3: Data for dissemination, communication and exploitation purposes

Type of data	Format
Dissemination and communication material (i.e. flyers and brochures, project public presentations and posters, press releases, interviews, video)	.pdf, .mp3, .mpeg, .avi
Public deliverables	.doc, .pdf
Open-access papers on journals/conferences describing project results	.pdf, .tex
Exploitation plans and activities	.doc, .pdf
User community information (collection and use)	**.xls, .xlsx**

4 Data description

To give a detailed description of the ALOHA data, the template provided by the European Research Council (ERC) for a data management plan[5], shown in Table 4, has been used. Since many questions concerning the FAIR principles cannot be answered at this moment, this section of the Open Data Management Plan will be reviewed and updated as the project proceeds. Each dataset used in the ALOHA project is presented in a separate subsection that is structured according to the ERC template.

Table 4: Template used for data description

Data summary	• State the purpose of the data generation and explain 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 and the expected size (if known) • Outline the data utility (to whom will it be useful)
Making data findable	• 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?
Making data openly accessible	• Specify which data will be made openly available. If some data is kept closed provide rationale for doing so (e.g. ethical, rules of personal data, intellectual property, commercial, privacy-related, security-related). • 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 (e.g. online repository)
Making data interoperable	• Specify what field-specific data and metadata vocabularies, standards or methodologies will be used to facilitate interoperability.
Increase data re-use	• Specify how the data is licenced to permit the widest reuse possible • Specify what data will be made available for re-use and for how long. If applicable, specify why and for what period a data embargo is needed
Allocation of resources and data security	• Estimate the costs for making data FAIR and potential value of long term preservation. • Describe how these costs are planned to be covered • Describe procedures that will be put in place for long-term preservation and curation of data

---

[5] ERC Data Management Plan Template, version 1.0, 21 April 2017, available at
http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination_en.htm

10

4.1 Data for project management

Most of the project management data will be generated within WP8. In Table 5, a description of these data is provided.

Table 5: Data for project management

Data summary	These data will be generated to ensure an effective and efficient project management, and to achieve global quality of the project and timely completion within budget. The types and formats of project management data are listed in Table 1. No project management data is being re-used. The project management data will probably not exceed 1 GB. These data will be useful to all project partners and to the EC, for what concerns deliverables, reports and financial information.
Making data findable	The project management data will not be findable with metadata. Public deliverables will be findable on the project website. All project deliverables will follow naming conventions and versioning numbering schemes already reported in Deliverable 8.5 – Project Quality Handbook.
Making data openly accessible	Data generated for project management purposes will not be made publicly available, with the exception of public deliverables once approved. Confidential deliverables, periodic reports and financial information will be provided to the EC through the Participant Portal. All meeting presentations/minutes/agenda will be accessible to project partners through the project documents repository (i.e. IBM Box).
Making data interoperable	No project management data will be made interoperable.
Increase data re-use	No data re-use will be made possible for project management data.
Allocation of resources and data security	No additional costs will be made for making project management data FAIR. Data will be stored on the project documents repository on IBM Box. All project partners have a personal password-protected account to access the repository. The access is granted by the IPR Manager and can be revoked at any time.

4.2 Data for internal use

The data collected and used to conduct internal development, evaluation and analysis of the ALOHA toolflow will be used within WP2, WP3, WP4 and WP5 activities. A description of these data is provided in the following sections. These data are to be used within the Consortium throughout the duration of the project.

4.2.1 Datasets used during the project development

These data will mainly consist of datasets to be used for the training, test and validation of a design point during the development of the ALOHA toolflow.

Table 6: Dataset for ALOHA toolflow development

Data summary	These data will include ALOHA-specific state-of-the-art public datasets used for benchmarking deep learning algorithms, such as MNIST, CIFAR10, CIFAR100, ImageNet, MS-COCO, **Speech Commands**, VoxCelab, LibriSpeech, Urban Sound, TCIA, OASIS and **Virat**: http://www.viratdata.org/ **PETS**: http://www.cvg.reading.ac.uk/PETS2014/a.html

11

	Cityscapes: https://www.cityscapes-dataset.com/ **PASCAL VOC**: http://host.robots.ox.ac.uk/pascal/VOC/ They will be wrapped for ALOHA integration and used for early testing and tuning of the ALOHA toolflow, as well as for comparison with the state-of-the-art. Datasets for internal toolflow development will be released in terms of 1) their wrappers for ALOHA integration, 2) the scripts necessary to fetch the original public datasets from their source, where applicable. The wrappers will be released as ONNX descriptions. The scripts as open-source Python code. The aforementioned public datasets are re-used. Size of the public datasets will range between a few MB (e.g. MNIST) to hundreds of GB (ImageNet, MS-COCO). Size of the wrappers will be of a few hundred KB per wrapper at most. The data will be useful for the purpose of testing the ALOHA toolflow on well-known datasets where performance has been reported publicly.
Making data findable	All public datasets include metadata to make them findable. The wrappers constitute the only layer of additional metadata super-imposed to the publicly available datasets.
Making data openly accessible	All data is already openly accessible (public datasets) or will be made accessible within the Project (wrappers, script). Data that will be made accessible within the Project will be released in the form of open-source code published on GitHub or equivalent services. Usage of the data may depend on additional libraries that are already publicly available as open-source software (e.g. PyTorch). **We plan to release also some additional versions of the aforementioned publicly-available datasets, including the perturbed data used by the security evaluation tool to assess robustness of the deep learning algorithms developed in WP2.**
Making data interoperable	Data will be made partially interoperable by means of utilizing the ONNX format for wrappers.
Increase data re-use	Data produced within the project (wrappers, scripts) will be available publicly with no restriction on re-use. Case-by-case re-use restrictions may apply on public datasets.
Allocation of resources and data security	No additional costs are presumed for public datasets.

4.2.2   Target architecture description

These data will consist of a set of JSON-expressed files including the characteristics of the target hardware architectures to be used in ALOHA in terms of processing capabilities, processing units, memory, and data bandwith. A preliminary description of the file format has been proposed and will be validated in WP4.

Table 7: Target architecture description

Data summary	The platform description data file is used by the desing space exploration toolchain in order to understand the capabilities of the platform described and map the considered DNNs in an optimal way on the platform, or to decide if the network cannot be mapped on the given platform. The data is prepared by the platform owner. It is a very compact description and the size would be around tens of Kbytes.

12

	The data will be provided to WP3 and WP2 as abstract descriptions of the platforms in order to be used as input in the ALOHA toolchain.
Making data findable	The data can be loaded as local files or it can be found on the ALOHA website or on the platform owner's websites.
Making data openly accessible	Usually these data can be made public, because they are an abstraction that provides the capabilities of the platform but it does not expose reserved information on the platform. In any case, the pubic availability of the data will be decided by the owner of the platform. The toolchain does not require a specific Access Point for the data, that can be provided with a local file or a web link.
Making data interoperable	The definition of the various fields of the platform description will be done in the project and will be standard for all the platforms.
Increase data re-use	The license of the data depends on the platform owner choices. Prospectively, a repository of platform descriptions will be collected and made available for new users to reduce ALOHA adoption efforts.
Allocation of resources and data security	The amount of data is negligible in terms of size, so no specific costs can be considered for the preservation of the data.

4.2.2.1 Orlando platform

Orlando is the R&D System-On-Chip developed by STMicroelectronics that embeds a set of DSPs and hardware accelerators for the execution of CDNNs. The SoC is provided with a development board that allows the connection to sources of data (images, audio) and the connection to external systems for the transmission of the results. The architecture description data has already been detailed in Table 7, but some specific aspects of the data depend on the platform owner, especially regarding the availability of the data. Since the content of the platform description is not finalized, it is not possible to state if there are some confidential information related to Orlando platform. Considering the actual information contained in the description, it can be made publicly available, and the data can be eventually hosted on the ALOHA website.

4.2.3   Datasets for use-case implementation

These data will consist of speech and image datasets to be used as test data for the implementation of the three use cases. PKE will also use datasets composed of trajectories of tracked objects and their classification based on security relevant aspects.

Table 8: Image datasets for the Surveillance Use Case

Data summary	Data described here will be used in the implementation of the surveillance use case for model training and evaluation. The following types of data will be used.
	• Images/Videos will be obtained as follows:
	   o Open datasets: Please refer to Table 6 for more information. Other open datasets may be added following the same scheme presented in that table.
	   o Recorded Images: PKE will record image data from test installations in its premises. Data will be collected according to the privacy protection laws. For business confidentiality and particularly privacy protection reasons, these data will not be made public.
	   o Artificially generated images and videos: PKE will investigate methods for generating artificial data for training CNNs. Due to its nature, data

	generated by this method will not contain personal data. Due to business confidentiality, these data will not be made public. • Trajectories. Trajectory data will be derived from the image data by manual annotation, automated object tracking or artificially generated. The trajectories will not contain personal data. Due to business confidentiality, the trajectory data will not be made public.
Making data findable	Not applicable as data recorded will not be made public. Regarding open datasets, refer to Table 6 for more information.
Making data openly accessible	Not applicable as data recorded will not be made public. Regarding open datasets, refer to Table 6 for more information.
Making data interoperable	Not applicable as data recorded will not be made public. Regarding open datasets, refer to Table 6 for more information.
Increase data re-use	Not applicable as data recorded will not be made public. Regarding open datasets, refer to Table 6 for more information.
Allocation of resources and data security	Data recording, generation, storing and annotation will be done according to agreed procedures regarding data protection. For these tasks, resources will be allocated according to the needs of the Surveillance Use Case.

Table 9: Speech datasets for the Smart Industry Use Case

Data summary	For the implementation of the Smart Industry Use Case, the following free and openly available speech corpora will be used: • Google Speech Command Dataset (3 GB, License: CC-BY 4.0): http://download.tensorflow.org/data/speech_commands_v0.02.tar.gz (v2 become available in 2018 and includes more keywords and more records) • Mozilla Common Voice (12 GB, License: CC-0): https://voice.mozilla.org/ • LibriSpeech ASR corpus (23 GB, License: CC-BY-4.0): http://www.openslr.org/12/
Making data findable	Open datasets, refer to Table 6 for more information.
Making data openly accessible	Open datasets, refer to Table 6 for more information.
Making data interoperable	Open datasets, refer to Table 6 for more information.
Increase data re-use	Open datasets, refer to Table 6 for more information.
Allocation of resources and data security	Open datasets, refer to Table 6 for more information.

Table 10: Image datasets for the Medical Use Case

Data summary	For the implementation of the Medical Use Case we plan to use the following datasets: • MaxQAI OOS Tarining Dataset composed of 1700 non-contrast brain CT studies (~30GB, License: MaxQAI)) • MaxQAI OOS Dataset composed of 360 non-contrast brain CT studies (~30GB, License: MaxQAI)) • MaxQAI-Test Dataset composed of 180 non-contrast brain CT studies (~15GB, License: MaxQAI)) – shall be available by 31 of December 2018 All 3 datasets shall include the following: • non-contrast brain DICOM CT scan series • Mask Ground Truth in same size of CT scan indicating per every voxel if it is considered a bleed voxel or not.
Making data findable	MaxQAI-Test Dataset is currently being collected and the partner MaxQ-AI will update about means to detect data once they publish it. This dataset shall include 180 non-contrast CT scans, 90 with IC hemorhage and 90 normal scans. Each study shall be identified by its Study ID and Series ID according to DICOM standards. Each Study shall also include ground truth indication of Bleed/Normal truthed by 2 or 3 (in case of disagreements) neuroradiologists.
Making data openly accessible	Only MaxQAI-Test Dataset shall be openly available only to the ALOHA consortium partners. Data can be viewed using any DICOM viewer such as RadiAnt (https://www.radiantviewer.com/) MaxQ Training & OOS data set can not have open access since they were collected from hospitals with privacy agreements. Ground Truth mask shall be in a ".mat" format and a wrapper for presenting the data shall be available
Making data interoperable	Not applicable
Increase data re-use	Not applicable. MaxQAI-Test Dataset shall be openly available only to the ALOHA consortium partners.
Allocation of resources and data security	All datasets shall be stored in MaxQAI database on AWS. Data Curation cost shall be covered by MaxQAI

4.2.4 DNN topologies

These data will consist of publicly available DNN topologies to be used as starting point for the design space exploration process and for conducting transfer learning.

Table 11: DNN topologies

Data summary	These data will include publicly available DNN topologies, such as AlexNet, VGGNet, MobileNet, SqueezeNet, ResNet50, ResNet110, UNet, and others. DNN topologies will be used together with the corresponding pre-trained weights available from open-source software. Depending on the dataset size and task complexity expected weight of DNN model with pre-trained weights might reach several hundred Mbytes. The data will be useful to project partners both as starting point for the design space exploration process and for conducting transfer learning based on pre-trained DNN.

15

Making data findable	All publicly available DNN topologies and weights include metadata to make them findable. All generated by ALOHA tool DNN topologies and weights during runtime will be available if not otherwise specified by the use-case partners.
Making data openly accessible	All data is already openly accessible (public DNN topologies and weights) or will be made accessible within the Project (generated DNN topologies and weights) if not otherwise specified by the use-case partners. Data that will be made accessible within the Project will be released in the form of open-source code and/or utilizing the ONNX and published on GitHub or equivalent services.
Making data interoperable	Data will be made partially interoperable by means of utilizing the ONNX format for wrappers.
Increase data re-use	Data produced within the project (weights, topologies) will be available publicly with no restriction on re-use if not other specified by the use-case partners. Case-by-case re-use restrictions may apply on publicly provided weights.
Allocation of resources and data security	No additional costs are presumed for public datasets.

4.3 Data for dissemination, communication and exploitation purposes

Most of these data will be collected and generated within WP6 and WP7 to develop an effective communication, dissemination and exploitation strategy of ALOHA research results. In Table 12 and Table 13, a description of these data is provided.

Table 12: Data for knowledge transfer

Data summary	These data will be generated for dissemination, communication and exploitation purposes. They will be useful to ensure the project knowledge transfer to external parties and to maximize the impact of the research results. The types and formats of these data are listed in Table 3. The size of these data is not predictable at this moment. These data will be useful to all project partners, to research community involved in machine learning, artificial intelligence, embedded system development and parallel programming techniques, and to third parties (private and public companies) that could benefit from the exploitation of the project results.
Making data findable	All data for dissemination, communication and exploitation purposes will contain the project's acronym, reference to the words EU and Horizon 2020 ensuring the promotion of the funding scheme and the identification and accessibility of the work in the future. In social media, ALOHA posts will be findable and discoverable using the Twitter account @ALOHA_H2020, or the LinkedIn account or referring to the project website http://www.aloha-h2020.eu
Making data openly accessible	Most of the data for external parties will be made openly accessible. All dissemination and communication material will be available on the regular public reports on WP7 activities (D7.2, D7.3, D7.4, D7.5 and D7.6) and part of them will be also published on the project social media accounts. All public deliverables will be published on the project website (once approved by EC) and provided to the EC through the Participant Portal.

16

	Peer reviewed scientific papers that contain data and results from the project will be published using a gold or green open access model when possible. The ALOHA open access policy to publications has been already reported in Deliverable D7.2 – Plan for the Dissemination and Communication. It follows the Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020. Exploitation plans will be available on the regular public reports on WP6 activities (D6.1, D6.2, D6.3) and will be published on the project website once approved by EC. However, depending on the level of detail described in the plans, some information may be kept confidential within project partners.
Making data interoperable	Not applicable
Increase data re-use	These data will be allowed for reuse following standard digital practices, i.e. naming the source of the information (e.g. http://www.aloha-h2020.eu/index.php/news-events/xxx).
Allocation of resources and data security	Estimated cost for making these data FAIR is related to costs of 'gold' open access publications. Data will be stored on the project documents repository on Box and on the project website once approved.

Table 13: ALOHA User Community Information Collection and Use

Data summary	These data will be collected and generated for building, setting up, and monitoring the ALOHA User community. The types and formats of these data are listed in Table 3. The size of these data is not predictable at this moment. These data will be useful to all project partners to ensure the project knowledge transfer to the ALOHA User Community.
Making data findable	The ALOHA consortium has appointed an internal data protection officer (DPO), responsible for the retention of data (see details in the ALOHA Privacy Policy). Data will not findalble without explicit and valid request to the DPO. Under certain circumstances, ALOHA consortium may be required to disclose Personal Data if required to do so by law or in response to valid requests by public authorities (e.g. a court or a government agency).
Making data openly accessible	Data collected won't be openly accessible
Making data interoperable	The ALOHA consortium has appointed an internal data protection officer (DPO), responsible for the retention of data (see details in the ALOHA Privacy Policy). Under the DPO authorization, other ALOHA staff members, especially task leaders or work package leaders, responsible for research tasks (including European Commission project officers of reviewers), may access to user data, only for networking purposes. Personal Data won't be transferred for business purposes.
Increase data re-use	The consortium may use Personal Data to contact users with newsletters, marketing or promotional materials, to announce events and with other information that may be of interest to the user community. The user may opt out of receiving any, or all, of these

17

	communications from the consortium by contacting the data protection officer. ALOHA may also use the collected data for various purposes: • To allow users to participate in interactive features of the project • To provide customer support • To improve the user experience, gather analysis or valuable information • To detect, prevent and address technical issues • To report to the European Commission (H2020 Research and Innovation programme) the impact of the ALOHA research project.
Allocation of resources and data security	The ALOHA consortium has appointed an internal data protection officer (DPO), responsible for the retention of data (see details in the ALOHA Privacy Policy). The ALOHA consortium may use third-party Service Providers (Google) to collect and store the user community subscriptions. More information on the privacy practices of Google, are displayed here: Google Privacy & Terms web page