



D7.4 Data Management Plan (updated)

Project acronym: ADMORPH
Project full title: Towards Adaptively Morphing Embedded Systems
Grant agreement no.: 871259

Due Date:	30.06.2021
Delivery:	Month 18
Lead Partner:	UvA
Editor:	Juliane Steinhardt, UvA
Dissemination Level:	Public (P)
Status:	
Approved:	Internal review (UAU, FC.ID)
Version:	1.1

DOCUMENT INFO – Revision History

Date and version number	Author	Comments
12/02/2020, v.0.1	UvA	First draft
02/03/2020, v.0.2	UvA	Second draft
04/03/2020, v.0.3	UvA	Minor changes (input UAU)
12/04/2020, v.1.0	UvA	Minor changes (input FC.ID)
14/04/2021, v.1.1	UvA	Minor update D7.2 (with inputs from UAU and FC.ID)

List of Contributors

Date and version number	Beneficiary	Comments
12/02/2020, v.0.1	UvA	
02/03/2020, v.0.2	UvA	
04/03/2020, v.0.3	UvA, UAU	
12/04/2020, v.1.0	UvA, FC.ID	
14/04/2021, v.1.1	UvA, UAU, FC.ID	

GLOSSARY

[illegible]

TABLE OF CONTENTS

Brief description.....	4
Executive summary	4
1 Open access policy to research results in H2020	6
2 Research results of ADMORPH.....	6
3 Data description.....	7
3.1 Data for project management.....	9
3.2 Data used by / produced from ADMORPH research	10
3.3 Data for dissemination, communication and exploitation purposes	11
4 References	14

Brief description

This deliverable describes the policy adopted by the ADMORPH Consortium regarding the management of data produced and/or used in the project. It details 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.

Disclaimer



This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 871259 (ADMORPH project).

This deliverable reflects only the author's view and the European Commission is not responsible for any use that may be made of the information it contains.

This document may contain material that is copyright of certain ADMORPH beneficiaries, and may not be reproduced, copied, or modified in whole or in part for any purpose without written permission from the ADMORPH 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 ADMORPH Consortium is the following:

#	Participant organisation name	Short name	Country
1	Universiteit van Amsterdam (Coordinator)	UvA	The Netherlands
2	Thales Nederland B.V.	TNL	The Netherlands
3	SYSGO S.A.S.	SYS	France
4	Université du Luxembourg	UNILU	Luxembourg
5	Lunds Universiteit	ULUND	Sweden
6	United Technologies Research Center Ireland, Limited	UTRC-I	Republic of Ireland
7	Q-media, S.R.O.	QMA	Czech Republic
8	FCiências.ID - Associacao para a Investigacao e Desenvolvimento de Ciencias	FC.ID	Portugal
9	University of Augsburg	UAU	Germany

Executive summary

As a project participating in the Open Research Data (ORD) pilot in Horizon 2020, ADMORPH is expected to deliver regular updates of its initial Data Management Plan that was produced in the first 6 months of its lifecycle, 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 Commission¹.

This document provides the updated version of the ADMORPH Open Data Management Plan. It 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. It is important to note that beneficiaries participating in the ORD pilot are not obliged to make all datasets open (as described in their DMP, and in compliance with confidentiality, security, data protection, and other relevant considerations).

This Data Management Plan is not a fixed document. It will evolve and gain further precision and substance during the remaining lifespan of the project. A final version of this plan will be submitted as deliverable at the end of the project:

- D7.5 Data Management Plan – Final version [M42]

1 Open access policy to research results in H2020

The policy on open research data in H2020, also reported in Article 29.3 of the ADMORPH 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.

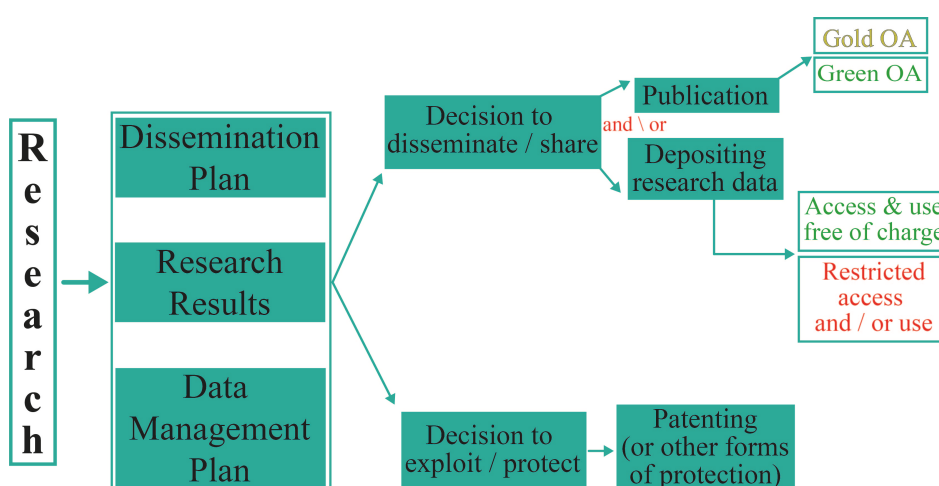


Figure 1: H2020 Open access policy to publications and open research data².

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 ADMORPH Consortium will follow the strategy shown 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.

2 Research results of ADMORPH

The ADMORPH Consortium will collect, use and generate mainly three categories of data:

1. data for project management purposes,
2. data used to conduct the ADMORPH research as well as data produced by ADMORPH experiments,
3. data for knowledge transfer such as scientific publications, public project deliverables, communication and dissemination material.

A complete list of all data to be collected, used, 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. The second category of data (i.e., used and/or produced by ADMORPH research tools) can be subdivided into several data types:

- Raw input data and (configuration) parameters for software (tools), like input to applications, simulators, analysis tools, etc.
- Meta-data describing input/output data, applications / application models, machine / system models, etc.
- Binary images of Operating Systems or VMs and bitstreams for FPGAs
- Experimental output data, which will mostly consist of data produced by simulators, analysis tools or directly measured from (operating systems running on) real systems. These data mostly involve the measurement of extra-functional properties of CPS: i.e. data on system performance (e.g., timing behavior), system reliability, etc.

The primary research data is expected to be in relatively small volume, on the order of magnitude of a few 100s Mb. The binary images/bitstreams can, however, take up several Gb's.

Table 1: Project management data

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

Table 2: Data used by / produced from ADMORPH research

Type of data	WPs	Partners involved	Format
Input data / (tool) parameters	WP1-5	ALL	.txt, XML, JSON, binary
Meta-data	WP1-5	ALL	.txt, XML, JSON
Images / bitstreams	WP4,5	ALL, but primarily SYS, UNILU and UTRC	binary
Experimental output data	WP1-5	ALL	.txt, XML, binary

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)	.tex, .pdf, .mp3, .mpeg, .avi
Public deliverables	.tex, .doc, .pdf
Open-access papers on journals/conferences describing project results	.pdf, .tex
Exploitation plans and activities	.doc, .pdf

3 Data description

To give a detailed description of the ADMORPH data, the DMPonline tool for an EC H2020 data management plan, shown in Table 4, has been used. Since not all questions concerning the FAIR principles can be answered at the current stage of the project, this section of the Open Data Management Plan will be reviewed and updated as the project proceeds. Each data category used in the ADMORPH project is presented in a separate subsection that is structured according to the EC H2020 template³.

Table 4: Template used for data description

<i>Data summary</i>	<ul style="list-style-type: none"> • State the purpose of the data generation and explain relation to the objectives
---------------------	---

	<p>of the project.</p> <ul style="list-style-type: none"> • Explain the relation to the objectives of the project • Specify the types and formats of data generated/collected • Specify if existing data is being re-used (if any) • Specify the origin of the data and the expected size (if known) • Outline the data utility (to whom will it be useful)
<i>Making data findable</i>	<ul style="list-style-type: none"> • Outline the discoverability of data (metadata provision) • Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers? • Outline naming conventions used • Outline the approach towards search keyword • Outline the approach for clear versioning • Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how
<i>Making data openly accessible</i>	<ul style="list-style-type: none"> • 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 how the data will be made available • Specify what methods or software tools are needed to access the data. Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)? • Specify where the data and associated metadata, documentation and code are deposited (e.g. online repository) • Specify how access will be provided in case there are any restrictions
<i>Making data interoperable</i>	<ul style="list-style-type: none"> • Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability. • Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?
<i>Increase data re-use</i>	<ul style="list-style-type: none"> • Specify how the data will be licenced to permit the widest reuse possible • Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed • Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why • Describe data quality assurance processes • Specify the length of time for which the data will remain re-usable
<i>Allocation of resources and data security</i>	<ul style="list-style-type: none"> • Estimate the costs for making your data FAIR. Describe how you intend to cover these costs • Clearly identify responsibilities for data management in your project • Describe costs and potential value of long-term preservation
<i>Data security</i>	Address data recovery as well as secure storage and transfer of sensitive data

<i>Ethical aspects</i>	To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former
<i>Other</i>	Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

3.1 Data for project management

Most of the project management data will be generated within WP7. In Table 5, a description of these data is provided.

Table 5: Data for project management

<i>Data summary</i>	<p>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.</p> <p>The types and formats of project management data are listed in Table 1.</p> <p>No project management data is going to be re-used.</p> <p>The project management data will probably not exceed 1 GB.</p> <p>These data will be useful to all project partners and to the EC, for what concerns deliverables, reports and financial information.</p>
<i>Making data findable</i>	The project management data will not be findable with metadata. Public deliverables will be findable on the project website and on the project's GitLab platform. All project deliverables will follow naming conventions and versioning numbering schemes already reported in Deliverable 7.1 – Project Handbook.
<i>Making data openly accessible</i>	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 (GitLab).
<i>Making data interoperable</i>	No project management data will be made interoperable.
<i>Increase data re-use</i>	No data re-use will be made possible for project management data.
<i>Allocation of resources and data security</i>	<p>No additional costs will be made for making project management data FAIR.</p> <p>Data will be stored on the project GitLab documents repository. All project partners have a personal password-protected account to access the repository. The access is granted by the manager of ADMORPH's GitLab infrastructure and can be revoked at any time.</p>
<i>Data security</i>	The GitLab ADMORPH document repository is periodically backed up, to ensure that data are recoverable should an accidental failure corrupt the main storage devices.
<i>Ethical aspects</i>	There are no ethical aspects to be considered concerning project management data.
<i>Other</i>	No.

3.3 Data used by / produced from ADMORPH research

Data used and produced by AMORPH involves WPs 1-5. In Table 6, a description of the data is provided.

Table 6: Data used by / produced from ADMORPH research

<i>Data summary</i>	<p>As mentioned before, the research data consists of i) raw input data and (configuration) parameters for software (tools), ii) meta-data describing input/output data, applications / application models, machine / system models, etc., ii) binary images and bitstreams for FPGAs, and iv) experimental output data. The formats of these data types are listed in Table 2. It is expected that the storage needs for the research data of types i, ii, and iv are relatively low, on the order of several 100s of MB. The storage needs for images/bitstreams will be several GB's. It is expected that (some of) this data will be relevant to other researchers in the domain, to use as a baseline for further research.</p>
<i>Making data findable</i>	<p>All publicly available research data will be accompanied by meta-data to make them findable. To this end, we are considering the use of the Zenodo platform⁴. This repository has been funded by the European Union with the specific objective of storing the research data that will be generated by Framework 7 and Horizon 2020 research projects. Using this repository has a number of advantages over the self-hosting or use of an institutional repository. In concrete:</p> <ul style="list-style-type: none"> • It provides a more visible and more centralised location for accessing research data, covering a number of EU projects and other sources of research data, rather than a single project and institution. This improves dissemination of the data. • Data in the repository can be accessed free of charge. A large number of standard licences are supported. • Data in the repository can easily be searched, mined and otherwise exploited. • The portal automatically generates a DOI and metadata for each dataset that is submitted. The DOI provides a location-independent reference point that is intended to ensure long-term accessibility of the data. This can be referred to directly from research publications. • Because it is dedicated to this purpose, the repository provides a long-term storage solution, ensuring the longevity and relevance of the research data. In particular, the data that is generated is guaranteed to be stored and curated long after the project has finished. Without continual intervention, this would be difficult to achieve using local and/or institutional repositories. <p>The metadata that is recorded for each dataset will include the date of submission, the owner of the data, a description of the data content, and a link to the ADMORPH project. Publications from ADMORPH will use Zenodo DOI references to link to relevant datasets, where possible. Any local storage of data will be according to the GDPR guidelines of the respective local entities. Currently, we have not yet used Zenodo to formally make any research data available and findable, but instead used self-hosting or institutional repositories where needed to make software and data available. This was due</p>

	to the fact that only relatively small amounts of, often preliminary, research data have been produced thus far. In the near future, it needs to be decided when and how we will make a transfer to Zenodo to make any research data more findable.
<i>Making data openly accessible</i>	The data provided and generated by the industrial partners in ADMORPH will not be made publicly available unless it is explicitly approved to do so by the partner(s) in question. All the data provided and generated by the academic partners in ADMORPH, which does not originate from any of the industrial partners, will be made public. Data that will be made accessible within the project will be released on the project's GitLab platform, which is hosted by one of our academic partners (FC.ID). For making data openly accessible, we are also considering the Zenodo platform (see above). Publicly available software tools developed in the scope of ADMORPH, consuming and producing project data, will be stored on the GitLab infrastructure. This software will also be made available on the ADMORPH website for public accessibility.
<i>Making data interoperable</i>	When applicable and needed, data will be made interoperable by utilizing common (meta-)data formats, such as XML and JSON, and by providing a (project-standardized) description of the fields within these data files.
<i>Increase data re-use</i>	We will ensure that publicly available project data will be made accessible, e.g. using the Zenodo platform (see above), via a standard license (e.g., a Creative Commons license). To further increase the re-use of ADMORPH's results, a substantial part of the software tools that will be used and developed in the project will be free, open source software. However, absolute care will be taken to avoid releasing proprietary software and to achieve the best possible commercial exploitation for tools that are developed and/or modified in the course of the project. Where tools are restricted, it will in some cases be possible to obtain these tools under some sufficiently liberal license agreement to allow full reproduction of the research experiments (possibly on payment of a fee). We feel that this constitutes a proper balance between obtaining the best possible and most convincing research results, and making those results as open as possible.
<i>Allocation of resources and data security</i>	No additional costs are foreseen for providing FAIR storage and access to ADMORPH datasets.
<i>Data security</i>	Within the project, research data will be made accessible via the ADMORPH GitLab platform, which will also provide daily data backup services.
<i>Ethical aspects</i>	There are no ethical aspects to be considered concerning the research data.
<i>Other</i>	No.

3.4 Data for dissemination, communication and exploitation purposes

Most of these data will be collected and generated within WP6 to develop an effective communication, dissemination and exploitation strategy of ADMORPH research results. In Table 7, a description of these data is provided.

Table 7: Data for knowledge transfer

<i>Data summary</i>	<p>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. More specifically, this data will be useful to all project partners, to the research community involved in the design of embedded systems and CPS, and to third parties (private and public companies) that could benefit from the exploitation of the project results.</p> <p>The types and formats of these data are listed in Table 3. The size of these data is not predictable at this moment.</p>
<i>Making data findable</i>	<p>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.</p> <p>In social media, ADMORPH posts will be findable and discoverable using the Twitter account @ADMORPH1, or the LinkedIn account (https://www.linkedin.com/in/admorph/) or referring to the project website http://www.admorph.eu/</p>
<i>Making data openly accessible</i>	<p>Most of the data for external parties will be made openly accessible.</p> <p>All dissemination and communication material will be available on the regular public reports on WP6 activities (D6.1, D6.2, D6.3) and part of them will be also published on the project social media accounts.</p> <p>All public deliverables will be published on the project website (once approved by EC) and provided to the EC through the Participant Portal.</p> <p>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 ADMORPH open access policy to publications has been already reported in Deliverable D6.1 – Plan for the Dissemination and Communication. It follows the Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020.</p> <p>Exploitation plans will be made available through the regular reports on WP6 activities (D6.2, D6.3). In principle, these reports are confidential, but if the project partners are in agreement on making (parts of) these deliverables publicly available (e.g., in the case this would be beneficial to the project and/or to some of its partners), we will publish this information on the project website once approved by the EC. However, depending on the level of detail described in the plans, some information may still be kept confidential and only accessible to project partners.</p>
<i>Making data interoperable</i>	Not applicable.
<i>Increase data re-use</i>	These data will be allowed for reuse following standard digital practices, i.e. naming the source of the information.
<i>Allocation of resources and data security</i>	<p>Estimated cost for making these data (scientific papers) FAIR is related to costs of 'gold' open access publications.</p> <p>No additional costs are foreseen for providing the remaining data FAIR, given that these data will be stored on the project documents repository on GitLab and on the project website once approved.</p>
<i>Data security</i>	Not applicable.
<i>Ethical aspects</i>	Not applicable.

<i>Other</i>	No.
--------------	-----

4 References

¹http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

²Adapted from: 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

³https://ec.europa.eu/research/participants/data/ref/h2020/other/gm/reporting/h2020-tpl-oa-data-mgt-plan-annotated_en.pdf

⁴www.zenodo.org