D1.2 DATA MANAGEMENT PLAN

Grant Agreement nr	856879
Project acronym	PRESENT
Project start date (duration)	September 1st 2019 (36 months)
Document due:	29/02/2020
Actual delivery date	25/02/2020
Leader	UPF
Reply to	josep.blat@upf.edu
Document status	Submission Version

Project funded by H2020 from the European Commission

Project ref. no.	856879
Project acronym	PRESENT
Project full title	**P**hotoreal **RE**altime **S**entient **ENT**ity
Document name	Data Management Plan
Security (distribution level)	PU
Contractual date of delivery	29/02/2020
Actual date of delivery	25/02/2020
Deliverable name	D1.2 Data Management Plan
Type	Report
Status & version	Submission Version
Number of pages	8
WP / Task responsible	UPF
Other contributors	All partners
Author(s)	Vanesa Daza, Guillem Antó, Josep Blat
EC Project Officer	Ms. Adelina Cornelia DINU - Adelina-Cornelia.DINU@ec.europa.eu
Abstract	This deliverable reports on the data plan strategy for the PRESENT project, summarises and discusses the type of data which the project will be dealing with, and especially in terms of its potential release.
Keywords	Data, publications, content, software, results
Sent to peer reviewer	Yes
Peer review completed	Yes
Circulated to partners	No
Read by partners	No
Mgt. Board approval	No

Document History

Version and date	**Reason for Change**
Shell 13-01-2020	shell created to request contributions from partners
Initial version 20.02.2020	first version of deliverable after the contributions period
Version 0.2 25.02.2020	Version after internal review, with the changes requested

Table of Contents

1. Abstract

This deliverable reports on the data plan strategy for the PRESENT project. The deliverable summarises and discusses the type of data which the project will be dealing with, and especially in terms of its potential release. The data is classified into different types, namely, publications, software, content and results. The strategy addresses data that can be made available for open research and measures to make it accessible. It outlines the access strategy for each type of data. It discusses resources, as well as some technical details such as security. In terms of making data accessible, there is an important component related to dissemination. The plan will be updated in the reports dealing with the capture of content and dissemination.

2. Data summary: several different types of data

PRESENT plans to generate several types of data where granting access is relevant: (academic) *Publications*, (professional) *Content*, *Software*, and other *Results*. Let us discuss briefly each of them, before going into more detail.

PRESENT is a Research and Innovation Action (RIA). As such, from the research component of the project, a number of academic *Publications* are expected to be produced. This type of data is well understood and does not need to be discussed further.

As PRESENT is heavily industrially oriented, most *Content* that the project will be working on is professional, protected by Intellectual Property rights (IP). A most relevant example of content generated within the project is the high quality capture of an actor, led by Framestore (FS) and Cubic Motion (CM), which was introduced in the amendment needed at the beginning of the project. Other relevant examples of content mentioned in the Description of Action (DoA) are related as well to machine learning / neural networks: data to be used to advance in body animation, rigging simplification, improved rendering, socially aware behaviour adaption, or the SSI framework; content for evaluation and demonstration will be generated as well; this is discussed later in relationship to *results*.
PRESENT intends to use mostly existing datasets, to focus on improvements rather than on generating data. On the other hand, most professional content of very high quality is IP protected - and, furthermore, it cannot be modified, for industrially related reasons -, and the high quality capture of an actor could be an exception to this situation. PRESENT will be exploring if it makes sense to make it widely accessible because of the very tight focus of the capture; and, in case it makes sense, when it could be made accessible and under which terms, both related to the contract to be signed with the actor, and to the competitive advantage of the industrial partners that needs to be strengthened.
On the other hand, data collected for later deployment of the agent in relationship to personalisation will remain on the users' private devices and is not relevant for this deliverable.

Software developed by the industry partners will mostly become part of new commercial products of some of the companies, or will reinforce their competitive advantage. This type of data is not relevant for this deliverable, oriented to the widest access. On the other hand, complementing the industrial strategy of partners, some of the academic ones have a policy of wide access to basic software libraries, as discussed in more detail later.

With respect to *Results*, an important aspect is to disseminate and promote project outcomes through demonstration, and PRESENT will discuss how to grant access to some results content, and demonstrate it widely.

3. FAIR Data

3.1 Making data openly accessible

We discuss next the accessibility strategy for each type of data mentioned in section 2.

As indicated earlier, *Publications* are the most discussed and understood data. PRESENT adheres to Open Access publication as a matter of principle, and compliance with Horizon 2020 rules for Open Access to scientific publications needs to be ensured. Each partner will choose the most suitable approach (either "green" OA or "gold" OA) for each publication concerned. The *e-repository*, which is the institutional repository of the UPF, the coordinator of the PRESENT project, is among the top 20 providers of publications and projects H2020 and ERC: it meets all the requirements established by the European Union within the framework of Open Access publishing, and will be used for UPF generated publications. Inria has its own repository (HAL, https://hal.inria.fr/), has been fully supporting Open Science strategies (including an 'obligation to deposit' since 2015), and it is a major source of publications, which includes both publications and software. UAu has an open repository as well (OPUS-Datenbank). UPF as coordinator will check the compliance with the H2020 rules, will make available the e-repository for the partners and that all publications resulting from the project are referenced on the web site, to ensure its widest dissemination, linking to the appropriate repositories.

With respect to *Software*, several academic partners promote Open Access. For instance,
UPF-GTI makes available web based 3D graphics software through GitHub (mostly through updated versions of *WebGLStudio* https://webglstudio.org/), as well as part of the results of EU supported current projects (HDR4EU, SAUCE). Continuing with this option seems a good alternative to make software as widely accessible as possible. A similar strategy is also used by UAu with respect to their *Social Signal Interpretation Framework* https://hcm-lab.de/projects/ssi/, which is going to be enhanced within PRESENT. We have already mentioned that HAL, Inria's repository, also includes software as well as publications.

As indicated earlier, professional *Data* of high quality not covered by IP is scarce; PRESENT partners will discuss whether it makes sense and is possible to make the high quality data character captured widely available. Other content generated, beyond using publicly available datasets, will be made available by the academic institutions as in the past.

With respect to *Results,* driven by dissemination and demonstration, PRESENT will discuss how to grant access to results content not protected by copyright, and demonstrate it widely.

The previous measures, which seem the most appropriate to ensure widest accessibility in the short and long term for each different type of data will be complemented with the efforts of the project to make all of them easily reachable from a single point. A data section will exist on PRESENT web, where the different types of data indicated earlier will be sufficiently documented and linked so that their wide accessibility is enhanced. For all the different types of data which we have discussed, exploiting what Zenodo linking could offer will be discussed.

3.2 Making data findable, including provisions of metadata

The use of established repositories for most data generated in PRESENT makes them most findable, as they include procedures which include persistent and unique identifiers (handle) which are automatically generated as well as metadata in standard formats.

In the case of professional content, this is not the case. A first concern of PRESENT will be that the capture of content, and the content itself gets appropriately documented, with as much detail as possible. How useful this documentation is for a wide variety of potential users will be initially tested by the partners themselves, as they will be the first users of this content. This documentation will be enhanced throughout the whole project life.

4. Allocation of resources

The strategy outlined in the previous paragraphs, based on existing institutional repositories, or well established ones such as GitHub seems to be minimalistic in terms of resources allocation, as they are part of the policy of institutions or GitHub. Let us remark that this has usually led to longer availability periods as has happened in the past.

Another resource for the accessibility of software is devoting sustained efforts to enhance and update the contributions, and document them. They mostly depend on funding for the projects being carried out, and appropriate efforts within the projects. PRESENT will be including these efforts as part of its dissemination and demonstration strategy.

As repeatedly mentioned, the situation is less standard for some of the professional content. The resources to support access during the life of the project exist. In the final stages of the project PRESENT will be able to estimate the interest of extending the access, and the appropriate resources to make it possible. In the past, UPF has been able to provide some support with respect to this in the framework of the resources in place through the *Maria de Maeztu grant for Excellence in Research*, awarded to the DTIC-UPF. Its renewal is under review at the moment of writing this deliverable. In the final stages of the project this should have become clear, and could be one of the options.

5. Data security

For the companies involved in PRESENT, most content is IP protected and their standards with respect to data security (including back-ups) are extremely high.

The public academic institutions involved in PRESENT are also committed to strong data security policies. CREW will make sure that appropriate data security is in place in tests involving users.

Within this section it seems convenient to discuss a security aspect related to the personalisation aspects mentioned above. It is intended that when real people use the PRESENT service (for instance, related to WP5), the authentication will be based on facial recognition; thus, enrollment will imply that the face of the user will be stored somewhere. The security module stores this private data in a secure place, which is owned and controlled by the user itself (the user's "wallet"). It is worth mentioning that, in general, user private data, when necessary, will be stored safely through the use of a blockchain that manages the identities of the users. While this type of data has been explicitly excluded from the *content* this deliverable deals with, it seems convenient to include here a short mention to security for completeness.

Thus, data related to PRESENT is expected to be covered by these data security policies.

6. Ethical aspects

Except when evaluation involves users, the data discussed in this deliverable does not contain ethical issues concerns. The eventual tests with users will be handled with appropriate ethical procedures, which are discussed in separate deliverables.

7. Other issues

At the moment of writing this deliverable, the project does not envisage the use of further national, sectoral or departmental procedures for data management. The project documentation is stored in a shared Drive folder under the umbrella of UPF as coordinator, ensuring appropriate data privacy and security, which is accessible to all project partners.

This Data Management Plan has been inspired by the UPF Data Management Plan template, and has considered other plans which are linked through the supporting website[1]. Furthermore, through the coordinator, it draws on the experience of the Plans prepared in the framework of two H2020 projects whose aims are close to those of PRESENT, namely, HDR4EU and SAUCE, and their implementation during more than two years of the life of each project.

8. Conclusions and further steps

This deliverable has presented the data which the project intends to make accessible to the wider community, indicating the strategy to manage the different types of data both internally and to make data accessible.

The plan will be be mainly implemented through dissemination. Within dissemination reports, updates and change in accordance to the evolution of the project will be indicated, allowing to keep track of both the plan and its implementation.

[1] https://guiesbibtic.upf.edu/data/en/dmp