DREAM: Deferred Restructuring of Experience in Autonomous Machines
H2020-FETPROACT-2014

Deliverable D7.2
Data Management Plan (M6)

Due date of deliverable: 1st, July, 2015
Actual submission date: 26th, June, 2015
Partner responsible: UPMC
Name of participants:

- Nicolas Bredeche (UPMC, main author)
- Stéphane Doncieux (UPMC)
- Zoitsa Siaplaoura-Karagkouni (UPMC)

Grant Agreement Number: 640891
Contract Start Date: January, 1st, 2015
Duration: 48 months
Project coordinator: UPMC
Partners: UDC, QML, ARMINES, VU/VUmc

Document validated by the Management Committee the 22nd, June, 2015.

CONTENTS

PREAMBLE

This document summarizes the work done in the first six months with respect to the data management plan. It provides a general classification of data content (Chapter 1) and an overview of data types and storage solution (Chapter 2). It also provides a description of the hardware solution chosen (Chapter 3), including plans to address fault-tolerance and possible failures of the system (hardware failure, data corruption (intentional or not), physical theft). The last Chapter lists the contact information for general public and media.

1 GENERAL INFORMATION ON DATA MANAGEMENT POLICY IN DREAM

1.1 TYPES OF DATA CONTENT

Three types of data are considered:

- **Type 1.** web site, storage
- **Type 2.** large data, storage
- **Type 3.** structured data for real-time exchange (tentative)

Storage means data is available for download (e.g.: data related to a particular paper or experiment) or viewing (e.g.: public pages on the project's web site). Storage data can be public or private.

Exchange means data that are produced on the fly and are used for ongoing experiment, possibly being conducted between sites (e.g.: knowledge sharing between two robots located in two different countries). In the current data management plan, this part will not be developed, and is presented here as a tentative proposal which will be consolidated or discarded before the next update of the current document (M30), depending on the experiments conducted and requirement for such a storage option.

Two different methods are proposed for accessing data: web-based (public or medium-security privacy) and SFTP-based (public or strong privacy).

1.2 PUBLIC/PRIVATE POLICIES

We define a two-level access policy: (1) public ; (2) private (consortium members only). The access policy for each data type is summarized on the following table:

	type 1	type 2	type 3
public	x	x	
private		x	x

Data type 1 can be read by anyone, but can be modified only by WP7 leader (with content given from all partners). For data types 2, only con-

sortium member's are able to read/write. Data type 3 is private and may be accessed and modified by any members of the consortium.

As general guidelines, all published papers will be available on the web site, each with a link to a large data set including both results and code to allow the reproduction of experiments by third parties interested. Other data will be public or private depending on the nature of the data. In particular, most data for on-going works will not be made publicly available until publication is completed.

2 DATA TYPES, STORAGE AND ACCESS POLICIES

2.1 DATA TYPE 1 : WEB SITE

We use a php-based Zite+ CMS, which is lighter though less powerful than other well-known CMS such as Drupal, but which also provides the advantage of being less sensitive to security issues. Beyond security issues, this CMS corresponds to our requirement for static data storage (text description, links to papers and data) and minimal amount of extensions beyond a Twitter feed and embedded videos.

The official web address for the project is Robot-that-dreams.eu. However, we have secured the following address names for the five following years (up to March 2020) in order to cope for misspellings and such:

- Robotsthatdream.eu (default)
- Robotsthatdream.net
- Robotsthatdream.org
- Robotthatdream.com
- Robotthatdreams.eu
- Robotthatdreams.net
- Robotthatdreams.org
- Robotthatdreams.com
- Robot-that-dreams.net
- Robot-that-dreams.org
- Robot-that-dreams.com
- Robots-that-dream.eu
- Robots-that-dream.net
- Robots-that-dream.org
- Robots-that-dream.com

The website is on-line since April 22nd (this information was twitted on April 30th). It can be accessed using the domain names specified earlier, with or without *www* as prefix. It contains several pages: home (incl. twitter feed), project description, meet the robots (description of the robots used), partners (description of each partner), publications (incl. PDF download and link to public experimental data), and contact. Please refer to deliverable D7.1 for snapshots and details.

2.2 DATA TYPE 2 : LARGE DATASETS

We use a homemade storage systems developed at UPMC, which is already used internally for other projects, and is managed by the laboratory system administrator. Two alternatives are possible for accessing the stored data:

- Through a web-based interface (authentification-based if data is private). It is possible to browse through different directories, each corresponding to a particular upload, organized by institutions. Within one directory, it is possible browse through the directory hierarchy of the uploaded package, but not to access specific content. All content is stored in an archive (tar (not recommended), tgz or zip, exclusively), and content display is virtual, meaning that the directory structure can be seen, but the files cannot be accessed. This makes it possible to reduce storage space required, while still being able to view the content structure (at the cost, of course, of indirect access to actual content). Accessing is filtered through a .htaccess file.
- Through SFTP access for uploading an archive (tar (not recommended), tgz or zip, exclusively), along with a text-format description file which will then be automatically processed by a dedicated script in order to build the directory structure, description (ie. name of experiment, date, and any other relevant information) and access policy. The default access policy is set to private. Access policy can be switched, if explicited, from private to public on per case basis, and will be defined by the partner uploading content. Accessing is filtered through system authentification.

Though each partner can allow several members to upload/manage content, each partner also names a contact person which is responsible for its institution content (e.g. re-organizing and/or checking for consistency). Contact persons for each institution are:

- UPMC: Carlos Maestre (maestre@isir.upmc.fr)
- ARMINES: David Filliat (david.filliat@ensta-paristech.fr)

- VU/VUmc: Jacqueline Heinerman (jacqueline.heinerman@gmail.com)
- U. da Coruña: Pilar Caamaño (pcsobrino@udc.es)
- Queen Mary Univ.: Timothy Hospedales (t.hospedales@qmul.ac.uk)

3 HARDWARE SOLUTION AND DATA PRESERVATION

We are in the process of acquiring a computer solution for handling all data described in the data management plan. We term this machine the *data machine* in the following. The order has already been placed (since beginning of May), and the machine should be delivered shortly (expected: end of June, early July).

To address the tradeoff between cost, availability and storage security, we have defined the expected maximum downtime allowed to 72 hours (worst case), with an expected maximum number of occurrence to once per year. We have defined a requirement for 12To of data (4*4To), with possible extension before the end of the project. The server runs with an Intel Xeon processor E5-2620v3, and 32 Go DDR4. The guarantee runs for 5 years, and the (theoretical) delay to intervention is 24h (i.e. less than the downtime worst case tolerance we established).

We have also defined a strategy for data replication: to ensure secured storage (and to avoid data loss), we use a RAID 6 scheme (data written on the master is also duplicated on two other disks, changing disks during run-time is possible). In addition, we use two processes of rolling backup: 2 disks are used for weekly backups and 4 disks are used for monthly backups. These backup disks won't be located in the same room as the data machine.

The machine to be acquired is presented a server rack (2U), to be included in the existing cluster structure at UPMC. The cost is 4529 euros (best out of three competing suppliers).

Long-term plan to preserve the data will be achieved at UPMC, and team budget will be used to maintain the data machine as long as it is relevant to. Note that technical support is already ensured by the system administrator at ISIR/UPMC, which is technically not included in the project PM budget, but is provided as lab support for projects.

4 CONTACT AND PUBLIC RELATION

We have defined two mail addresses for contact purpose:

- dream-contact@isir.upmc.fr is given as contact name on the web site and on other support for communication. It forwards messages to Stephane Doncieux (scientific coordinator), Nicolas Bredeche (WP7 leader);
- dream-admin@isir.upmc.fr is used for technical issues, including contact address for managing web addresses registering. It forwards messages to Stephane Doncieux (scientific coordinator), Nicolas Bredeche (WP7 leader) and Ludovic Billard (system administrator at ISIR/UPMC).