Deliverable D7.4
DATA MANAGEMENT PLAN (M48)

Due date of deliverable: 31st, December, 2018
Actual submission date: 17th, December, 2018
Partner responsible: SU
Name of participants:

- Nicolas Bredeche (SU, main author)

- Stéphane Doncieux (SU)

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

CONTENTS

PREAMBLE

As a follow-up to two previous deliverables on the data management plan, this deliverable summarizes the work done between month 30 and 48. Chapter 1 summarizes the content of deliverables D7.2 (month 6) and D7.3 (month 30) as well as work done since then. Chapter 2 provides an overview of the current state of the data management plan and Chapter 3 discusses perspective for the short- and medium-term future of the data management plan following the official end of the project.

There were no significant deviation with respect to the original data management plan and directions given in D7.3.

As a reminder: the partner UPMC changed name to SU ("Sorbonne Université") on January 1st 2018. While previous deliverables D7.2 and D7.3 used the former acronym (and name), we use the latter in this deliverable. There were no other change except for the name.

1 SUMMARY OF WORK DONE SINCE M30

The main contribution of Deliverable D7.2 (M6) was the identification of data types produced by the project, whether we considered web-related content for a general audience[1], production of large (public or private) data sets or setting up a dedicated private server for data diffusion within the consortium[2].

In Deliverable D7.3, we described the implementation of the data server from which large data sets can be made accessible (ie. both public *and* private data sets). In particular, we described the Data Management Server in terms of Hardware and Software settings, as well as the first batches of actual data uploaded to the server.

The data server is up and running since February 2017 (early deployment and test started in mid-2016). The implementation plan described and used at the time of D7.3 are still valid and functional to this day. The quality of service has remained within the range of the initial specs (i.e. no downtime, no loss of data, robust backup strategy). The reader if referred to D7.3 for a comprehensive description of the specifications.

In addition to deploying the data server, the project's web site[3] was launch in April 2015 and has been on-line since then. It was updated by month 30 with a new Section on the **Robobo** robotic educational platform developed by Coruna's University and local industrial partners. Regular updates are available through an embedded twitter feed[4] (97 tweets and 331 followers as of December 18th, 2018), and the publication page[5] has been regularly updated (last update: November, 30th 2018). By month 30, the web site is hosted by the data server (previously hosted on a non-dedicated server at SU).

---

1 http://robotsthatdream.eu
2 https://dream.isir.upmc.fr/databases/, data with restricted access
3 http://robotsthatdream.eu
4 https://twitter.com/robotsthatdream
5 http://www.robotsthatdream.eu/Publications.html

2 DATA MANAGEMENT SERVER: CONTENT AND USAGE

Since its deployment, the data server has been used by several partners (but not all, as needs varied from one partner to another), with a special use from SU, ENSTA, UDC and ED. The data server is hosted at SU, with read/write accesses for all the consortium (account-based, managed by SU). Data is accessible either through terminal-based connection and web-based application.

Since the last deliverable 18 months ago, the data stored on the server has grown from 9GB to 28GB, spread over 8 entries (4 entries on M30), listed here as an illustration:

- *babbling db 1, 2017/02/13*: dataset produced by a sensorimotor babbling. This corresponds to the wave 1-2 and 3. Dataset produced with only one object on a table, with the real robot.

- *crustcrawler pushing, 2017/06/05*: this dataset contains crustcrawler pushing motions obtained by executing Quality-Diversity search.

- *baxter pushing, 2017/06/06*: raw data from preliminary experiment with the Baxter robot.

- *data archive sim 1, 2017/06/07*: dataset produced by a sensorimotor babbling. This corresponds to the wave 1-2 and 3. This dataset was produced with only one object on a table, in simulation.

- *ball throwing baxter, 2017/07/06*: The dataset contains ball pushing motions for baxter obtained by executing QD search (ver.1).

- *baxter throwing qd, 2017/10/09*: The dataset contains ball pushing motions for baxter obtained by executing QD search (ver.2).

- *data envo 07-12-17, 2017/12/07*: The dataset contains arm motion w.r.t. moveable object for reinforcement learning.

- *baxter SRL, 2018/11/27*: Dataset for State Representation Learning on real Baxter data[1]

---

1 code available at: https://github.com/araffin/srl-zoo

The data server is used for storing very large data sets, that cannot be sent by e-mail and/or require long-term backups and/or that required making data available in-between teams. It has been used in parallel with other techniques, such as using local storage for punctual exchange and represent but a part of data generated and exchanged within the project. However, it has been clear that the availability of a large data storage capacity is critical on specific occasions, and was beneficial to the project. Another indirect benefit is that the project's web server is hosted on the same hardware, thus benefitting from the same quality of service in terms of back-ups and guaranteed maximal downtime.

3 A NOTE FOR THE FUTURE

With respect to the data management plan, the obvious question is: what of data preservation and availability beyond the termination of the project? In our particular case, the hardware has been installed and integrated within the computer room at the Institute of Intelligent Systems and Robotics at SU. It is managed by the technical staff at SU, which means that maintenance cost is assumed by the lab, and will remain so after EU funding stops. Hardware maintenance, if necessary, will be funded by SU using local funding from the lab. We expect the server, and thus data hosted on the server, to remain available for the next 5 years (ie. until 2024). During that time, previous partners will be able to access existing data, and possibly to add new data would there be any continuation of the project in one form or another involving SU. This longevity plan also holds for the project's website, which is hosted on the same hardware.