



Co-funded by the Horizon 2020  
Framework Programme of the European Union

HORIZON2020 FRAMEWORK PROGRAMME

ICT – 21 -2014

Advanced digital gaming/gamification technologies



Gamification of Prosocial Learning

for Increased Youth Inclusion and Academic Achievement

## **D9.1 Data Management Action Plan**

action plan



### Document Control Page

<b>WP/Task</b>	WP9 / T9.1
<b>Title</b>	D9.1 Data Management Action Plan
<b>Due date</b>	30/06/2016
<b>Submission date</b>	29/07/2016
<b>Abstract</b>	This deliverable provides an analysis of the main elements of the data management policy that will be used by the consortium with regard to all the datasets that will be generated by the project.
<b>Author(s)</b>	Pilar Pérez (ATOS)
<b>Contributor(s)</b>	Antonio Díaz (ATOS), José Miguel Garrido (ATOS), Anna Zoakou (EA), Stefano Cobello (POLO), Michael Boniface (IT Innov), Kosmas Apostolakis (CERTH)
<b>Reviewer(s)</b>	EA, POLO, ITINNOV, CERTH, KTH, RK, PG, UCAM
<b>Dissemination level</b>	<input type="checkbox"/> internal <input checked="" type="checkbox"/> public <input type="checkbox"/> confidential

### Document Control Page

Version	Date	Modified by	Comments
0.1	01/06/2016	Pilar Pérez (ATOS)	ToC and preliminary draft
0.2	08/07/2016	Pilar Pérez (ATOS)	Definion of data management strategy and type of data
0.3	28/07/2016	Pilar Pérez (ATOS)	Integration of contributions of EA and POLO
0.4	29/07/2016	Pilar Pérez (ATOS)	Integration of contributions of CERTH and ITINNOV
1	29/07/2016	Pilar Pérez (ATOS)	Final and format review



---

### List of Abbreviations

Abbreviation	Description
EC	European Commission
DMP	Data Management Plan
PsL	ProsocialLearn

---



## Executive summary

This document is the Deliverable D9.1 Data Management Action Plan within the WP9 Project Management and is consequence of the joining of ProsocialLearn project to the Horizon 2020 Open Research Data Pilot. This report is part of management activities and it is key to highlight the importance of the data management strategy in support of moving from prototype to market.

The purpose of the Data Management Plan (DMP) is to provide an analysis of the main elements of the data management policy that will be used by the consortium with regard to all the datasets that will be generated by the project. This document can evolve along the project and it will be updated, if needed, within D9.6 Final Management Report due to M36. The update could include new set of data and information related to the discoverability, accessibility, intelligibility, usability and interoperability of the research data gathered.

The ProsocialLearn Data Management Action Plan addresses on a dataset by dataset basis the following aspects:

- Data set reference and name
- Data set description
- Standards and metadata
- Data Sharing
- Archiving and preservation (including storage and backup)



## Index

1	Introduction.....	6
1.1	Purpose of the document.....	6
1.2	Scope and Audience of the document .....	6
1.3	Structure of the document.....	6
2	Data Management Strategy .....	7
2.1	Types of data to be generated/collected .....	7
2.1.1	Login information and scoring.....	8
2.1.2	Audio data .....	8
2.1.3	Video data .....	9
2.2	Standards and metadata to be used .....	10
2.3	Exploitation, availability of data and re-use.....	10
2.4	Archiving and preservation .....	10
3	Information to be included in data tables.....	12
4	ProsocialLearn Data.....	13



## 1 Introduction

This section provides detailed information about the purpose, scope and structure of the document as well as the intended audience of the document.

### 1.1 Purpose of the document

ProsocialLearn consortium participates in the Pilot on Open Research Data in Horizon 2020 to offer open access to its scientific results reported in publications, to the relevant scientific data and to data generated throughout the project lifetime. The present document aims to define the data management strategy and detailing the types of data to be gathered, standards and metadata to be used and exploitation and preservation of data.

### 1.2 Scope and Audience of the document

The dissemination level of this document is public. This document will be shared within the consortium, EC services and external researchers.

### 1.3 Structure of the document

The document contains the following sections:

**Section 1:** Introduction – an introductory section, i.e. this present section, which describes the main purpose of the document

**Section 2:** Data Management Strategy – this section explains the different types of data gathered during the project, the standards and metadata to be used, the availability of data for its re-using and finally its archiving and preservation.

**Section 3:** It is a guide to complete the tables in section 4.

**Section 4:** This section presents the data gathered in the project detailing the reference and name of the data set, a description, the standards and metadata applicable, how it will be shared and preserved.



## 2 Data Management Strategy

The general strategy for data management is based on the identification and classification of data generated and collected, standards and metadata to be used, exploitation and availability of data as well as how the data will be shared and archiving and how the information will be preserved.

### 2.1 Types of data to be generated/collected

ProsocialLearn aims to improve and maximise access to and re-use of scientific data generated by the project i.e., face tracking, sound data and gameplay monitoring data among other.

To this end, this deliverable identifies datasets to be made discoverable, accessible, assessable and intelligible, useable and interoperable to specific quality standards.

Data can be classified according to levels of sensitivity:

- No personal data: this information is not affected by Data Protection legislation
- Personal data: data which relate to an individual who can be identified<sup>1</sup>
  - (a) from those data, or
  - (b) from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller, and includes any expression of opinion about the individual and any indication of the intentions of the data controller or any other person in respect of the individual.
- Sensitive personal data: data consisting of information of special protection as:
  - (a) the racial or ethnic origin of the data subject,
  - (b) his political opinions,
  - (c) his religious beliefs or other beliefs of a similar nature,
  - (d) whether he is a member of a trade union (within the meaning of the Trade Union and Labour Relations (Consolidation) Act 1992),
  - (e) his physical or mental health or condition,
  - (f) his sexual life,
  - (g) the commission or alleged commission by him of any offence, or
  - (h) any proceedings for any offence committed or alleged to have been committed by him, the disposal of such proceedings or the sentence of any court in such proceedings.

Next, the main categories of data, according to sensitivity, have been identified, the strategies for handling these data, outlined and their legitimate use modalities within the project, described. The system level architectural design of these data management provisions are mainly implemented in Task 2.3 with the rest of the system architecture design.

Specifically, there are three (3) categories of data identified in the ProsocialLearn platform according to their management requirements and sensitivity:

- a. **School information:** Information gathered during the small scale studies when teachers register to ProsocialLearn online community. The questionnaire includes information related to school conditions: demographic information, data related to infrastructure of the school, percentage of migrant pupils, phenomena of bullying, exclusion, aggressive behavior etc.

---

<sup>1</sup> Key definitions of the Data Protection Act, ico (Information Commissioner's Office), <https://ico.org.uk/>



Pupils on the other hand fill in a questionnaire with regards to their experience from playing prosocial games.

- b. Visual sensing, affect-related and monitoring data from the interaction of the user with the system** using each game engine's input devices, and
- c. Data derived from the knowledge management and ProsocialLearn data analysis.**

The outcomes of the project will gather the following data:

- From the **PsL Market Place**:
  - o Login information of: gaming providers, teachers, schools' directors, schools' administrative: country, name, address, phone, email.
  - o Children login information (id gamer): animal, shape, colour, password and idschool.
- From the **PsL Teachers' Community Space**:
  - o School information: information related to socio-economic aspects and technical aspects.
- From the **games**:
  - o Audio data, video data, scoring, sensor data.

In any case, ProsocialLearn follows the European directives:

- Data Protection Directive<sup>2</sup>
- Directive on privacy and electronic communications.<sup>3</sup>

The activities focussed on the ethical and privacy issues and compliance to legislation in ProsocialLearn are carried out in WP7, T7.1 Ethical Oversight and is reported periodically in D7.5 Year 1 Ethics report, D7.6 Year 2 Ethics Report and D7.7 Year 3 Ethics Report.

The type of data that will be collected in ProsocialLearn is described in details in the following paragraphs.

### **2.1.1 Login information and scoring**

All communications that involve transfer of this data category shall employ the necessary security precautions (e.g., HTTPS connection). Special provisions shall be taken, when these data leave the confines of the project and are archived for reasons other than academic or personal needs, for them to be anonymized and identified by means that cannot lead to the revelation of the child's identity and personal information.

### **2.1.2 Audio data**

The following is an exhaustive and complete list of all the audio data collected (through questionnaires, interviews, extraction from records, etc)

- Audio recordings of game participants during the collaborative game
- Audio recordings of participants and their teacher before and after the game

The audio data will be processed to extract acoustic cues for emotion.

The data is relevant to the study purposes because the natural interaction between gamers during collaborative game playing provides ecologically valid speech for to quantify prosocial behaviours

---

<sup>2</sup> <http://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX:31995L0046>

<sup>3</sup> <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2002:201:0037:0047:en:PDF>





and responses in this context. The data is adequate because the study seeks only to identify emotional response (affect) during prosocial activity; and the data is not excessive because it is restricted to the behaviours solely within the confines of the game itself, and to the audio associated with those responses. The 5 minute teacher-led discussion data provide a further validation for affective response in the context of the games.

The data will be processed fairly because the data will only be processed in accordance with the purposes explicitly set out in the consent forms. Anonymisation is assured initial through the arbitrary assignment of a gaming name (an animal name); as well as additional processing to extract acoustic correlates of affect, not speaker characteristics.

The data's accuracy is ensured because it is recorded directly from the data subjects during game playing and immediately after.

Data will be stored on a secure, UoS laptop, which will be hand-carried back to the UK. The data will be held in accordance with University of Southampton policy on data retention.

Data files and laptop where they are stored will be password protected and kept on University of Southampton premises.

The original audio data will be destroyed as soon as processed, and no later than 31st Nov, 2016 without further consent. All derived data will be held by University of Southampton, IT Innovation, for a period of 10 years for research purposes only. There is no current plan to make the data available outside IT Innovation.

The data will be processed in accordance with the rights of the participants because they may withdraw at any time during the game itself; and requires that their processed data be destroyed. Once the original audio data is destroyed, and no other data are held, it will not be possible for data to be modified or removed other than that: there is no linked anonymity, beyond the random animal name identifier: the researchers will have no reference list, linking name and data subject.

The data shall always be transferred and stored anonymised. Identification and correlation of such data shall always be made by means that will not allow the identification of personal information. All communications that involve transfer of this data category shall employ the necessary security precautions (e.g., HTTPS connection, etc). Additional measures shall be taken to ensure that all sensor and behaviour monitoring data that can provide personal information are not susceptible to signal interception.

### **2.1.3 Video data**

Video data captured by a web-camera, or other visual sensing devices, such as the Microsoft Kinect sensor, will be processed locally on the game client for the extraction of facial and/or skeletal features, such as Emotions, Facial Animation Units and Joint data. These anonymized features will be sent then to the platform to be used as input to the data fusion module. The transmission of raw video data will not be taking place.

A) These data shall always be transferred and stored anonymised. Identification and correlation of such data shall always be made by means that will not allow the identification of personal information. All communications that involve transfer of this data category shall employ the necessary security precautions (e.g., HTTPS connection, etc). Additional measures shall be taken to ensure that all sensor and behaviour monitoring data that can provide personal information are not susceptible to signal interception.



B) Visual sensing data will only be captured and processed on local terminals and they will not be stored. Transmission of video data will not be taking place. This was an early provision since the preliminary design phase provided in Section 1. Moreover, inclusion of visual (video or static image) data and personal information will, by no means, be used as part or example of any publicly available dataset that will derive from ProsocialLearn project.

## **2.2 Standards and metadata to be used**

Data regarding users will be stored on a MySQL database for internal use. Specific data published for scientific community will be published using commons standards such as JSON or XML.

## **2.3 Exploitation, availability of data and re-use**

A. both video and audio files are going to be used and analyzed only within the framework of the project research

B. both video and audio files are going to be used only for research purposes

C. no video and/or audio files or parts of them will be presented in public

D. both video and audio files will be deleted, as soon as analysis has been completed and the project has finished

Regarding the data sharing method the following questions must be considered:

- How will the data be available to others?
- With whom will the data be shared, and under what conditions?

Considerations about where, how, and to whom the data should be made available must be taken into account (e.g. share data via a data repository, handle data request directly etc).

The methods used to share data will be dependent on a number of factors such as the type, size, complexity and sensitivity of data. Mention earlier examples to show a track record of effective data sharing.

## **2.4 Archiving and preservation**

ProsocialLearn persist it information on database server taking into account privacy concerns. Some of the data can be used for scientific purposes but due the special sensitivity of the some information we will choose only very specific information and we will anonymize the data.

Each participant responsible for gathering, analysis of the data, will adhere to the overall Data Gathering & Production section of the Data Management Plan, describing the ways, policies to gather data in a statistically significant and academically correct way. Preservation will be taken care of as said before - through project's web site, partners' websites and suggested storage locations. Further specifics for data curation and preservation were provided in GA and CA.

Regarding the elaboration of a Preservation plan some questions should be considered:

- What is the long-term preservation plan for the dataset? e.g. deposit in a data repository
- Will additional resources be needed to prepare data for deposit or meet charges from data repositories?



How datasets that have long-term value will be preserved and curated beyond the lifetime of the grant must be specified as well as the plans for preparing and documenting data for sharing and archiving.

Regarding resourcing, some questions should be considered:

- What additional resources are needed to deliver your plan?
- Is additional specialist expertise (or training for existing staff) required?
- Do you have sufficient storage and equipment or do you need to cost in more?
- Will charges be applied by data repositories?
- Have you costed in time and effort to prepare the data for sharing / preservation?

All resources needed to deliver the plan should be carefully considered and also outlined and justified. Any relevant technical expertise, support and training required and the process to acquire this kind of expertise should be justified in detail as well as any hardware or software or additional storage and backup costs incurred by IT services.

Funding should be included to cover any charges applied by data repositories, for example to handle data of exceptional size or complexity.

The cost in time and effort to prepare data for deposit and the appropriate documentation to enable reuse should be taken into account. The cost data management (in a data repository or in own resources) must be also considered and detailed.

Also remember to cost in time and effort to prepare data for deposit and ensure it is adequately documented to enable reuse. If you are not depositing in a data repository, ensure you have appropriate resources and systems in place to share and preserve the data.<sup>4</sup>

---

<sup>4</sup> UKDS (UK Data Service) guidance on costing data management.  
<https://www.ukdataservice.ac.uk/manage-data/plan/costing>



### 3 Information to be included in data tables

According to the Guidelines on Data Management in Horizon 2020<sup>5</sup>, the purpose of the Data Management Plan (DMP) is to provide an analysis of the main elements of the data management policy that will be used by the applicants with regard to all the datasets that will be generated by the project.

The DMP is not a fixed document, but evolves during the lifespan of the project.

The DMP should address the points below on a dataset by dataset basis and should reflect the current status of reflection within the consortium about the data that will be produced.

#### Data set reference and name:

- Identifier for the data set to be produced.

#### Data set description:

- Description of the data that will be generated or collected, its origin (in case it is collected), nature and scale and to whom it could be useful, and whether it underpins a scientific publication. Information on the existence (or not) of similar data and the possibilities for integration and reuse.

#### Standards and metadata:

- It includes reference to existing suitable standards of the discipline. If these do not exist, an outline on how and what metadata will be created.

#### Data sharing:

- A detailed description of how data will be shared, including access procedures, embargo periods (if any), outlines of technical mechanisms for dissemination and necessary software and other tools for enabling re-use, and definition of whether access will be widely open or restricted to specific groups. Identification of the repository where data will be stored, if already existing and identified, indicating in particular the type of repository (institutional, standard repository for the discipline, etc.).

In case the dataset cannot be shared, the reasons for this should be mentioned (e.g. ethical, rules of personal data, intellectual property, commercial, privacy-related, security-related).

#### Archiving and preservation:

- Description of the procedures that will be put in place for long-term preservation of the data. Indication of how long the data should be preserved, what is its approximated end volume, what the associated costs are and how these are planned to be covered.

---

<sup>5</sup> Guidelines on Data Management in Horizon 2020, European Commission (2016) [https://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)



## 4 ProsocialLearn Data

1	<b>Data set reference and name:</b>
	Data set 1 – User information
2	<b>Data set description:</b>
	<p>This dataset includes every user involved on the system such as teachers, students, budget holders, game developers.</p> <p>Dataset will be collected on every user registration form on the market place. User details stored depends on the profile of the use. Relevant data for every user are: first name, last name, country, phone, email, state, a zip code. The data set will be stored in a database.</p>
3	<b>Standards and metadata:</b>
	There is not a defined standard for this task. We will use standards widely utilized on the industry as json, xml, ...
4	<b>Data sharing:</b>
	There will not be publicly accessible information of the user details due to the character of the data, taking into account the privacy and ethics considerations.
5	<b>Archiving and preservation (including storage and backup):</b>
	Every dataset will be preserved the minimal time necessary for the project.

1	<b>Data set reference and name:</b>
	(DOI REF to be added when published) – Voice of School Children Playing a Cooperative Digital Game annotated with Valance Arousal classification
2	<b>Data set description:</b>
	Audio recordings and valance arousal classification of 49 8-10 year old Italian children playing a cooperative digital game together. The recordings capture voice of individual children playing a tablet based game in groups of 4 around a table, with each child using their own tablet. The data set is from two schools (11 girls, 13 boys [age 10]) and (7 girls, 18 boys [age 8]) playing 12 and 10 games respectively. A total of 315 minutes of audio data is available.
3	<b>Standards and metadata:</b>
	<p>Audio Data is stored in raw Pulse Code Modulation (PCM)</p> <p>Metadata is stored as Experience API (<a href="https://www.adlnet.gov/adl-research/performance-tracking-analysis/experience-api/">https://www.adlnet.gov/adl-research/performance-tracking-analysis/experience-api/</a>)</p>
4	<b>Data sharing:</b>
	Green Open Access, published as soon as paper is accepted



5	<b>Archiving and preservation (including storage and backup):</b>
	Planned: University of Southampton ePrints ( <a href="http://eprints.soton.ac.uk/">http://eprints.soton.ac.uk/</a> ) and referencing through issued Digital Object Identifier (DOI)

1	<b>Data set reference and name:</b>
	(DOI REF to be added when published) – Prosocial skill performance data for children learning cooperative skills in a digital game
2	<b>Data set description:</b>
	Prosocial skill performance data for 8-10 year old children learning cooperative skills in a digital game. The data captures children playing a tablet based game in groups of 4 around a table, with each child using their own tablet. The data includes children’s interaction logs and a score providing an assessment of their cooperative behaviours. The data set is from two schools (11 girls, 13 boys [age 10]) and (7 girls, 18 boys [age 8]) playing 12 and 10 games respectively. A total of 315 minutes of audio data is available.
3	<b>Standards and metadata:</b>
	Experience API ( <a href="https://www.adlnet.gov/adl-research/performance-tracking-analysis/experience-api/">https://www.adlnet.gov/adl-research/performance-tracking-analysis/experience-api/</a> )
4	<b>Data sharing:</b>
	Green Open Access, published as soon as paper is accepted
5	<b>Archiving and preservation (including storage and backup):</b>
	Planned: University of Southampton ePrints ( <a href="http://eprints.soton.ac.uk/">http://eprints.soton.ac.uk/</a> ) and referencing through issued Digital Object Identifier (DOI)

1	<b>Data set reference and name:</b>
	Data set 2 – User video analysis
2	<b>Data set description:</b>
	This dataset will contain facial expression analysis as well as skeletal joint features.
3	<b>Standards and metadata:</b>
	We will use standards widely utilized on industry, such as XML.
4	<b>Data sharing:</b>
	There will not be external data sharing, only internally among the partners and for project purposes, e.g., evaluation and optimization of algorithms.
5	<b>Archiving and preservation (including storage and backup):</b>
	Every dataset will be preserved the minimal time necessary for the project.



1	<b>Data set reference and name:</b>
	Data set 3 – Teacher surveys and reports
2	<b>Data set description:</b>
	This dataset will contain the reports with results an data anonymous of the overcome.
3	<b>Standards and metadata:</b>
	There is not a defined standard for this task. We will use standards widely utilized on the industry as JSON, xml, ...
4	<b>Data sharing:</b>
	There will not be publicly accessible information of the user details due to the character of the data, taking into account the privacy and ethics considerations.
5	<b>Archiving and preservation (including storage and backup):</b>
	Every dataset will be preserved the minimal time necessary for the project.

1	<b>Data set reference and name:</b>
	Data set 4 - Pilot outcomes report –
2	<b>Data set description:</b>
	This database will contain the reports with results of the piloting in the schools. It will also contain the suggestions provided by the users and teachers
3	<b>Standards and metadata:</b>
4	<b>Data sharing:</b>
	There will not be external data sharing, only internally among the partners and for project purposes
5	<b>Archiving and preservation (including storage and backup):</b>
	Every dataset will be preserved the minimal time necessary for the project.