



This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement N° 709515



|                  |   |
|------------------|---|
| Project Title    | Science Technology Innovation Mathematics Engineering for the Young |
| Project Acronym  | STIMEY  |
| Project No       | 709515  |
| Instrument       | Research and Innovation Action                                      |
| Start Date       | 01/09/2016  |
| Project Duration | 36 Months   |
| Project Website  | <a href="http://www.stimey.eu">http://www.stimey.eu</a>             |



This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement N° 709515

|  |                              |
|--|------------------------------|
| <b>WP N°</b>   | WP10                         |
| <b>Delivery Rel. N°</b>  | D51                          |
| <b>Delivery N°</b>   | D10.7                        |
| <b>Title of Deliverable</b>  | A Data Management Plan       |
| <b>Lead Beneficiary</b>  | University of Cadiz (UCA)    |
| <b>Main Author:</b>  | José Luis Aparicio Rodríguez |
| <b>Contributing Authors</b>  | Mireya López Mesa            |
| <b>Reviewed By</b>   | Arturo Morgado               |
| <b>Approved By</b>   | Carlos Rioja del Rio         |
| <b>Nature</b>  | ORDP                         |
| <b>Dissemination Level</b>   | Public                       |
| <b>Est. Del. Date (annex I)</b>  | M6                           |
| <b>Due Date</b>  | February 28, 2017            |
| <b>Sent Date</b>   |                              |
| <b>Approval Date</b>   |                              |
| <b>Deliverable Description</b>   |                              |
| A Data Management Plan (DMP)   |                              |
|   <small>This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement N° 709515</small>         |                              |



This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement N° 709515



### Versioning and contribution history

| Version | Date       | Author                             | Notes  |
|---------|------------|------------------------------------|--|
| 1.1     | 03/11/2016 | José Luis Aparicio                 | Initial Draft Version  |
| 1.2     | 15/12/2016 | José Luis Aparicio                 | Document completed   |
| 1.3     | 16/01/2017 | José Luis Aparicio                 | First Revision   |
| 1.4     | 17/01/2017 | José Luis Aparicio                 | Second Revision  |
| 1.5     | 18/01/2017 | Mireya López                       | Third Revision   |
| 1.6     | 23/01/2017 | Mireya López                       | Fourth Revision  |
| 2.0     | 14/03/2018 | José Luis Aparicio<br>Mireya López | On date 09/03/2018, the officer "Simona MISITI" has requested the revision of the submission for the deliverable with title "A Data Management Plan" and number "7" for the project "STIMEY (709515)". |



This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement N° 709515

## Table of Contents

---

|   |    |
|---|----|
| 1. Document Purpose   | 5  |
| Executive Summary   | 5  |
| 2. Data Summary   | 6  |
| 2.1 Making data findable, including provisions for metadata [Fair Data] | 8  |
| 2.2 Making data openly accessible [Fair data]                           | 9  |
| 2.3 Making data interoperable [Fair data]                               | 10 |
| 2.4 Increase data re-use (through clarifying licenses) [Fair data]      | 10 |
| 3. Allocation of Resources  | 11 |
| 4. Data Security  | 12 |
| 5. Ethical Aspects  | 12 |
| Definition of Terminology (Glossary)                                    | 13 |



This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement N° 709515



## 1. Document Purpose

---

### Executive Summary

This document is part of the deliverable of Work Package 10, D10.7 A Data Management Plan. Data Management Plan (DMP) will be implemented due to our participation in the PILOT on Open Research Data in Horizon 2020. This deliverable describes all the data that will be collected and generated during the STIMEY project, how it will be created, stored and backed-up, who owns it and who is responsible for the different data and which data will be preserved and shared.



This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement N° 709515



## 2. Data Summary

---

STIMEY will both collect data from partners and third parties, and will generate new data within the project. These Data will be collected and generated with the only purpose of developing research activities in STIMEY.

The main goal of the project is to engage Society to Science, Technology, Engineering and Mathematics (STEM), awakening and supporting scientific and technical careers. In order to reach this goal, STIMEY will develop a platform which will contain personal and academic information as part of the student e-profile. The students have the opportunity of playing some serious games, and as they play. All of them will develop their cognitive and knowledge profile and build their personal creativity curve. The platform will have material developed by teachers, schools, universities, research centres and corporations they will be able to create personal learning environments combining existing resources and updating existing information. Every educational centre and company registered in the platform need to provide information related to them in order to make them clearly identifiable for further references.

Other objective within the project is to develop a socially assistive robotic artefact, this robot will be able to communicate with others, and send back information to the platform. This social media communication will be useful for both scientific and educational studies.

A variety of different methods of collection are used, but all adhere to high international standards. Personal data (e.g. age, gender, languages...) will be collected via questionnaires or register forms.

Different Stakeholders will actively participate and collaborate in the STIMEY project by giving specific information related to STEM areas. Educational centres and teachers will take part in the platform by creating new specific content or evaluating the student progress. Students will be assessed at different stages of the year using a variety of activities, serious games, robotic artefacts and other approaches at school and home. The task will have been designed as a way of engaging society to science and awakening and supporting scientific and technical careers. A complete list of datasets to be collected and created is shown in Table 1.



This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement N° 709515



Science Technology Innovation Mathematics  
Engineering for the Young

| Dataset                        | Purpose   | Type/ Format  | Origin of the data   |
|--------------------------------|---|---|--|
| Personal Data                  | <ul style="list-style-type: none"> <li>- User profile in the platform</li> <li>- Socioeconomic/cultural studies</li> </ul>  | <ul style="list-style-type: none"> <li>- Text</li> <li>- Images</li> <li>- Voice</li> </ul> | <ul style="list-style-type: none"> <li>- Questionnaires</li> <li>- Stakeholder &amp; Data collection</li> </ul>                          |
| Academic Data                  | <ul style="list-style-type: none"> <li>- Courses list</li> <li>- Courses' scores/awards on and out of the platform</li> </ul>   | <ul style="list-style-type: none"> <li>- Text</li> <li>- Images</li> </ul>                  | <ul style="list-style-type: none"> <li>- User input in the platform module</li> <li>- Interconnectedness with other platforms</li> </ul> |
| Family information             | <ul style="list-style-type: none"> <li>- Parent/Child (under 13) connection</li> </ul>  | <ul style="list-style-type: none"> <li>- Text</li> <li>- Images</li> </ul>                  | <ul style="list-style-type: none"> <li>- User input in the platform module</li> </ul>  |
| Professional Data              | <ul style="list-style-type: none"> <li>- Work history (job titles, company names, dates)</li> <li>- Achievements/awards</li> </ul>  | <ul style="list-style-type: none"> <li>- Text</li> <li>- Images</li> </ul>                  | <ul style="list-style-type: none"> <li>- User input in the platform module</li> </ul>  |
| Stakeholder Information        | <ul style="list-style-type: none"> <li>- List of collaborators in STIMEY and information about them</li> </ul>  | <ul style="list-style-type: none"> <li>- Text</li> <li>- Images</li> </ul>                  | <ul style="list-style-type: none"> <li>- Questionnaires</li> <li>- Stakeholder &amp; Data collection</li> </ul>                          |
| Educational Center Information | <ul style="list-style-type: none"> <li>- List of schools that collaborate in STIMEY and information about them</li> </ul>   | <ul style="list-style-type: none"> <li>- Text</li> <li>- Images</li> </ul>                  | <ul style="list-style-type: none"> <li>- Questionnaires</li> <li>- Stakeholder &amp; Data collection</li> </ul>                          |
| Cognitive profile              | <ul style="list-style-type: none"> <li>- Information about how the students process new information or activities</li> <li>- Learning potentials</li> <li>- Personal strengths</li> </ul> | <ul style="list-style-type: none"> <li>- Text</li> <li>- Spreadsheet</li> </ul>             | <ul style="list-style-type: none"> <li>- Questionnaires</li> <li>- Platform activities</li> </ul>  |
| Emotional profile              | Information about: <ul style="list-style-type: none"> <li>- Self-control</li> <li>- Self-motivation</li> <li>- Adaptability</li> <li>- Leadership</li> </ul>                              | <ul style="list-style-type: none"> <li>- Text</li> <li>- Spreadsheet</li> </ul>             | <ul style="list-style-type: none"> <li>- Questionnaires</li> <li>- Platform activities</li> </ul>  |
| Knowledge profile              | <ul style="list-style-type: none"> <li>- Skills and knowledge acquired</li> <li>- Indicator related to student characteristics</li> <li>- Learning improvements</li> </ul>                | <ul style="list-style-type: none"> <li>- Text</li> <li>- Spreadsheet</li> </ul>             | <ul style="list-style-type: none"> <li>- Questionnaires</li> <li>- Platform activities</li> </ul>  |



This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement N° 709515



|                              |  |   |   |
|------------------------------|--|---|---|
| Companies working for STIMEY | <ul style="list-style-type: none"> <li>- List of companies outsourced by STIMEY</li> <li>- Personal contacts of the company</li> <li>- Solvency index</li> </ul> | <ul style="list-style-type: none"> <li>- Text</li> <li>- Spreadsheet</li> </ul> | <ul style="list-style-type: none"> <li>- Questionnaires</li> <li>- Platform activities</li> </ul> |
| Statistical reports          | <ul style="list-style-type: none"> <li>- Data used as measures for indicators</li> </ul>   | <ul style="list-style-type: none"> <li>- Text</li> <li>- Spreadsheet</li> </ul> | <ul style="list-style-type: none"> <li>- Questionnaires</li> <li>- Platform activities</li> </ul> |

Table 1 Datasets

In order to improve the accessibility and the digital preservation in the long term it is recommended to use:

- Complete and open documentation
- Non-proprietary software
- No key-protection
- No total/partial encoding
- Open formats like RTF, TIFF, JPG or well-known proprietary formats

### **2.1 Making data findable, including provisions for metadata [Fair Data]**

Metadata facilitates exchange of data by making them Findable, Accessible, Interoperable and Re-Usable (F.A.I.R.). Metadata will be used to identify and locate the data through catalogues or search engines.

All the data produced in STIMEY project will be identified by using a unique and persistent identifier: HANDLE, guaranteeing a permanent access and allowing you to reference your data in a safety way.

The structure of HANDLE uri is: producer prefix + "/" + document suffix. i.e.: <http://rodin.uca.es/handle/10498/14617>

The metadata will be based on a generalised metadata scheme used in the RODIN platform. This is an institutional repository located in the University of Cadiz. Its goal is to create a digital deposit in order to store, preserve and disseminate all the documentation related to research activities.

RODIN repository can store data and documents related to Horizon 2020 projects. These data will be collected afterwards by OpenAIRE. Following the generalised metadata scheme used in RODIN, we will have the elements listed below:





This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement N° 709515



- Title.
- Creator/s: Last name, First name
- Contributor: Information provided by the EU or the STIMEY project itself.
- Subject: List of keywords
- Description: Text explaining the content of the data.
- Date.
- Type: type of document: i.e.: “info:eu-repo/semantics/workingPape”
- Identifier: HANDLE uri
- Language: document/data language
- Relation
- Rights: license, and access. I.e.: info:eu-repo/semantics/openAccess
- Format: details about the file format

A readme.txt file could be included to provide information on field methods and procedures.

## **2.2 Making data openly accessible [Fair data]**

Every data collected and generated (non-personal data) with the main aim of developing research activities within the STIMEY project will be openly available.

Due to the participation of underage people in activities related to the project, with the aim of protecting and guaranteeing their privacy, all their data will be collected and treated without being identified.

All the data related to the platform will be located in a server in the University of Emden/Leer and will be available by logging in the site. Previously, you must have signed up using your email or social media account. On the other hand, every personal or restricted data will be located in a computer server in the University of Cadiz.

Data will be processed for the limited purposes of the studies; therefore, only relevant data will be collected. Data will be used for the analysis of the proposed studies results.

Every data generated in this project will be:

- Collected in a fair, faithful and transparent way.
- Collected with the only purpose of develop research activities in STIMEY
- Suitable, appropriate and limited to necessary for the project.
- Stored no longer than necessary
- Protected against non-authorized treatment and guarantee their security

Regarding to personal data, every person involved in research studies will be assigned a unique key. In order to generate this key/code we will have a system with a deterministic hash function, this type of function receive a string and always return the same value. This string will be form



This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement N° 709515



by the full name and birthdate of the person. It is important to know that it is a one-way function, it is not possible to obtain a name with a given key.

Only the researcher team of the project will have access to the codify data. In order to be on the list (authorized people) it will be necessary to have the authorization of the Data Controller and the STIMEY coordinator.

Servers and the storage in the University of Cadiz will be located in their Data Center of level TIER III following the standard classification ANSI/TIA-942.

Apart from these repositories, STIMEY will also use the centralised repository RODIN to ensure the maximum dissemination of the information generated in the project. This repository make use of the OAI-PMH protocol (Open Archives Initiative Protocol for Metadata Harvesting), what allows that the content can be properly found by means of the defined metadata.

### **2.3 Making data interoperable [Fair data]**

By using RODIN repository to store and disseminate the data and metadata generated (non-personal data) associated to the STIMEY project we are facilitating the interoperability of the data. All the data generated can be exchanged between researchers and partners in STIMEY.

RODIN follows the *Dublin Core Scheme*; it is a small set of vocabulary terms that is used to describe web resources such as videos or images. The complete list of terms can be found on the Dublin Core Metadata Initiative (DCMI) website. The DCMI Abstract Model was designed to bridge the paradigm unbounded, linked data graphs with the more familiar paradigm of validatable metadata records like those used in OAI-PMH.

The full list of fifteen-element of metadata terms (DCMI) are ratified in the following standards: IETF RFC 5013, ANSI/NISO Standard Z39.85-2007, and ISO Standard 15836:2009.

All the Metadata used in the STIMEY project will be “mappable” on standard vocabularies by following the *Dublin Core Scheme*.

### **2.4 Increase data re-use (through clarifying licenses) [Fair data]**

Data generated (non-personal data) and associated software will be deposited in RODIN. To facilitate the re-use of the data, these will be made available under a Creative Commons BY-NC-SA License, this licence permit to others to copy, distribute, display and perform the work for not commercial purposes only, also permits other to create and distribute derivative works, but only under the same or a compatible license.



This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement N° 709515



This project will use “Open access publishing, also called ‘Gold’ open access, this means that an article is immediately provided in open access mode by the scientific publisher. The associated costs are shifted away from readers, and instead to (for example) the university or research institute to which the researcher is affiliated, or to the funding agency supporting the research.

All the data produced (non-personal data) in the project will be made available for reuse at the end of project. These data will be also usable by third parties the project but private or restricted data which won't be available in any case. The length of time for storing data will be, at least until the end of the project.

The data quality is ensured by different measures. These include validation of the sample, replication, comparison with results of similar studies and control of systematic distortion and statistical reports based on several indicators.

The procedures that are fundamental to effective data quality assurance could include:

- Document data quality requirements and define rules for measuring quality
- Assess new data to create a quality baseline
- Implement semantic metadata management processes
- Keep on top of data quality problems

### 3. Allocation of Resources

---

We will consider the costs of data for this project related to the price of keeping the information in the different repositories both for public data and for private data.

University of Cadiz will use RODIN repository and this service is for free. HS EL University, which will work with data of the STIMEY platform, uses BitBucket repository also for free if the number of users in the repository is less than five.

Associated costs for dataset preparation and data management during the project will be covered by the project itself.

Inmaculada Medina Bulo as General Director of Information Systems at University of Cádiz, has been assigned as Data Controller of the STIMEY project and has been added to the Project Advisory Board and the consultant committee for ethics aspects and personal data protection.

Every partner has assigned one person (Data Processor) responsible for following up the procedures designed by the Data Controller, specially collecting and storing procedures including informed consent process and transparency process.



This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement N° 709515



## 4. Data Security

---

During the project, data will be automatically saved on an institutional server with backup on a separate offsite server. Backup will be checked and validated manually.

The key generator system will be located in a computer server in the University of Cadiz. This computer will be turned off and with no connection, excepting one hour a week. In that period, the teacher in charge of the students will be able to access the system and generate new keys for them. Every access will be authenticated and stored in a logging file. Every IP address will be validated and previously known, minimizing every possible risk when it is generated a key. It is important to highlight that every name of the participants in STIMEY activities will be never stored.

In the same way, we can affirm that every access to the information located in Spain will be exclusive to the personal in charge of the project, following the recommendations of the standard ISO/IEC 27002:2005.

## 5. Ethical Aspects

---

According to the principle of “data minimization” based on article 5 section c) of the general data protection regulation, it set out that it will not be collected any data that contains ethnic or racial information, union or political affiliation, religious beliefs, as in genetic data, biometrical data, data health data, data related to the life or sexual orientation of a person.<sup>1</sup>

---

<sup>1</sup> Deliverable 11.3- STIMEY Data Protection, Section “Procedures implemented for data collection, storage, protection, retention and destruction”.



This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement N° 709515



## Definition of Terminology (Glossary)

---

| Acronym  | Description   |
|----------|---|
| DMP      | Data Management Plan  |
| EU       | European Union  |
| WP       | Work Package  |
| STIMEY   | Science Technology Innovation Mathematics Engineering for the Young |
| STEM     | Science Technology Engineering and Mathematics                      |
| FAIR     | Findable, Accessible, Interoperable and Re-Usable                   |
| RODIN    | Repositorio de Objetos de Docencia e Investigación                  |
| OpenAIRE | Open Access Infrastructure for Research in Europe                   |
| DCMI     | Dublin Core Metadata Initiative                                     |