EDUCATIONAL ROBOTICS FOR SCIENCE, TECHNOLOGY, ENGINEERING AND MATHEMATICS

DATA MANAGEMENT PLAN

Deliverable 8.1 v1.0

Dissemination Level: PU (Public)

Date of Publication: 31.03.2016

The ER4STEM project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 665972

CONTRIBUTORS

Name	Beneficiary	Contact
Lara Lammer	TUW	lammer@acin.tuwien.ac.at
Carina Girvan	CU	girvanc@cardiff.ac.uk

DOCUMENT REVISION HISTORY

Version Number	Date	Description	Author
V0.0	18.3.2016	First version	Lara Lammer
V0.1	23.3.2016	Review	Carina Girvan
V1.0	31.3.2016	Submitted first version	L. Lammer, C. Girvan
V1.1			
V2.0	M17	Submitted 2nd version	
V3.0	M35	Last version	

DISCLAIMER

The ER4STEM project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 665972

TABLE OF CONTENTS

The ER4STEM project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 665972

TABLE OF REFERENCES

The ER4STEM project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 665972

1 EXECUTIVE SUMMARY

1.1 ROLE/PURPOSE/OBJECTIVE OF THE DELIVERABLE

This Data Management Plan outlines how the research data collected will be handled during and after the ER4STEM project. It describes the data set, how it will be archived and preserved as well as how it will be shared. It is a "living document" with a first version in project month six (M6), an update before the mid-term review in month 17 (M17) and a final version before the final review in month 35 (M35).

1.2 RELATIONSHIP TO OTHER ER4STEM DELIVERABLES

Research data will be collected in work packages WP2 (Workshops and Curricula) and WP3 (Conferences and Competitions). The evaluation process is described in D6.1 (Evaluation Pre-Kit) and will be refined in D6.2. Deliverables D6.3, D6.4 and D6.5 will report on the evaluation results with the data collected. The dissemination deliverables D8.2, D8.3 and D8.4, especially the report on scientific dissemination, are also closely linked to the data management plan.

1.3 STRUCTURE OF THE DOCUMENT

The Introduction states the purpose of the document and explains the type of information it contains. In section 3, the data collected during the project is described in detail. Section 4 deals with standards and metadata. Section 5 elaborates on which data will be shared with whom and how and Section 6 gives an overview on how the data will be archived and stored during the project and afterwards.

The ER4STEM project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 665972

2 INTRODUCTION

The ER4STEM Data Management Plan (DMP) outlines how the research data collected during the project will be handled during and after the project.

It is structured as suggested in [1] and describes:

- the data set
- standards and metadata
- data sharing
- archiving and preservation

Throughout these sections, reference is made to data protection, ethics, the evaluation (WP6), publications and other forms of dissemination (WP7) as well as to the two main activities of data collection (WP2 and WP3).

The DMP is not a fixed document; it will evolve and gain more precision and substance during the lifespan of the project. The first version of the DMP is delivered in project month 6 (M6). It will be updated once in project month 17 (M17), before mid-term review, and finalized in project month 35 (M35), before the final project review, to fine-tune it to the data generated and the uses identified by the project consortium. The data management has already been discussed in milestone MS2 in project month M4, where important decisions have been taken for the first version of the DMP document. The DMP will continue to be part of the milestone reviews, next one being due in project month 7 (M7) for MS3, after submission of the DMP Version 1.0.

3 DATA SET

In the first year of the project research data will be collected during the workshops and conferences. Therefore, the first version of the DMP is mainly concerned with the management of this research data. As the project develops, so will the DMP to include the evaluation of the framework, the repository, the industrial perspective and the studies of the main stakeholders of the project, as required.

As part of work packages 2, 3 and 6 (WP2, WP3 & WP6), data is collected by partners from multiple sources at multiple sites. This data is quantitative and qualitative in nature and will be analysed from different perspectives for project development and scientific evaluation with results published in scientific conferences and journals.

Data will only be collected following the informed consent, and in the case of minors, their parent or guardian.

3.1 DATA SET REFERENCE AND NAME

In this document data regarding the workshops conducted during the project by each of the partners will be referenced to as **workshops data set** and will include data of over 4000 children by the end of the project. A similar approach will be followed for the conference data, but with smaller numbers. This data set will be referenced to as **conferences data set**. Collected data will be anonymized by using participant numbers (a randomly assigned number with partner code and project year). The participant key, which connects

The ER4STEM project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 665972

participant information to participant numbers, is the only document that contains personally sensitive material (name of the participant, age, parent or school name and contact information) and will not be shared outside of the partner organisation or with people in the partner organisation who do not require direct access to this information. The participant key will be stored securely according to Data Protection Laws and will not be removed from the partner organisations.

3.1.1 WORKSHOPS DATA SET

Since the data comes from multiple sources, the workshops data set will have its own folder structure and following documents collected: Workshop information, pre-questionnaire, post-questionnaire, observations, interviews, artefacts of learning, tutor reflections, and encrypted sensitive data like videos and audio files. These files will be named using the following convention:

For documents and templates created by the partner responsible for evaluation, Cardiff University, the data will be named after the organisation conducting the workshop or conference, a six-digit date of the workshop or conference, and the original file name. For example, PRIA_160416_ObservationSchedule.doc

If multiple documents from the same organisation on the same date exist, identifiers will be added as appropriate to the data, i.e. TutorName or GroupName. For example, TUW_250616_Lara_TutorReflection.doc or TUW_250616_Julian_TutorReflection.doc

If no Cardiff University template exists (typically for artefacts, audio and video), the name will state the organisation, the six-digit date, then the group name and data type. For example, AL_040216_RobotAddicts_AudioInterview

The files will be stored in a folder structure as in Figure 1. The detailed process of planning and conducting the workshops is part of work package 2, their evaluation is part of work package 6 and already described in D6.1.

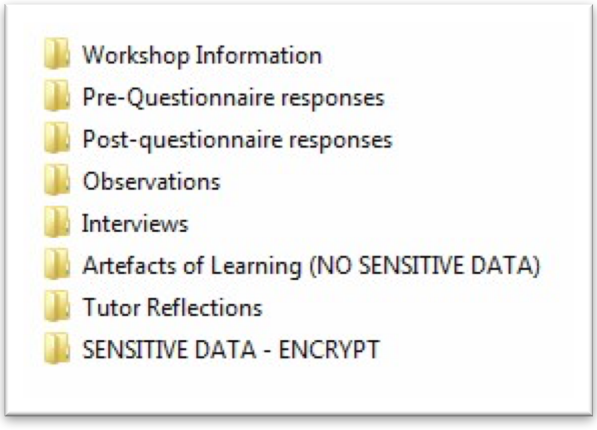

Figure 1: Workshop data folder structure

The ER4STEM project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 665972

3.1.2 CONFERENCES DATA SET

The conferences data set is a compact and slightly adapted form of the workshops data set with fewer numbers of children and follows the same naming conventions.

In ECER2016, roughly 300 students are expected to be present, many of whom will have participated in preparatory workshops and thus will have already contributed to the workshops data set. Further data will be collected from these participants, specifically focusing on the conference.

3.1.3 OTHER DATA SETS

Other data sets will need to be defined as the project progresses. Naming conventions will follow the workshops data set. An appropriate folder structure needs to be created depending on the data set and its purposes.

3.2 DATA SET DESCRIPTION

3.2.1 WORKSHOPS DATA SET

A detailed description of the evaluation method and the rationale behind it as well as detailed information on the collected data is provided in D6.1. In this document, the data set will be described as a summary.

The workshops data set includes following documents and information:

- Workshop Session Information (.doc)
  - Partner name
  - Dates (to-from)
  - Number of sessions
  - Location
  - Lead by
  - Other tutors/mentors
  - Age of students
  - Total number of students
  - Male/Female numbers
  - Group sizes
  - Total number of groups
  - How were the groups formed? Why?
  - Robotics kit
  - Programming languages
  - Domain
  - Aims of workshop
  - Please include all relevant lesson materials (e.g. activity plan, each session/lesson plan, handouts, etc.) in the folder with this document
- Draw a scientist (writings translated into English, anonymised and digitalised .pdf)

The ER4STEM project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 665972

    - Filled by the participants
    - To answer the question “are popular gender stereotypes about STEM held?”
- Questionnaires (anonymised and online or .xls)
    - Filled by the participants
    - Pre- and post-workshop questionnaires are used to collect largely quantitative data
    - Questions are split into personal information (age, gender and school), past experience and existing attitudes to STEM subjects and careers
    - The post-workshop questionnaire also includes questions about the activities to help understand learners’ experiences of the workshop as a whole, what participants feel they have learned and what their future intentions are
    - Paper questionnaire answers to be entered into the online system
    - Free-text responses translated into English and entered into excel files
- Observations (translated into English, anonymised and digitalised .xls)
    - Observation protocol filled by the workshop facilitators or other observers
    - Video observation where possible to verify and expand upon observation notes, as well as sensitising data analysts to the context.
- Interviews (anonymised, transcribed and translated into English, .xls)
    - With focus groups, audio-recorded
    - Conducted to understand the experience of participants and their reasons for particular actions
- Artefacts of learning (translated into English where applicable, anonymised and digitalised .pdf)
    - Created by the participants
    - Identified as group work
- Participant reflections (translated into English, anonymised if applicable and digitalised if applicable, .xls)
    - Created individually and as a team
    - Either as a blog which acts as a reflection tool and a living artefact of the learning process, or as a guided reflection document
- Tutor reflections (translated into English, anonymised and digitalised, .xls)
    - Done by each of the tutors, mentors or workshop facilitators
    - The purpose is two-fold: 1) To document changes to workshop plans and the reasons for these; and 2) to document the evolution of activity plans between workshops.
- Sensitive data (audio and video recordings)
    - Audio recordings of interviews and any video recordings are encrypted and stored by the partner organisation and only encrypted video files are shared with evaluation partner Cardiff University for the purpose of analysis and archiving.

The data will be collected during the workshops by partner organizations mostly from the workshop participants, children ages 7 to 18. Over 4000 participants in five European countries are planned for the whole duration of the project. The first year is regarded as a pilot year with approximately 1000 students participating in the pilot evaluation. Collected data will be used to improve processes regarding the workshops as well as their evaluation. The data will also inform the development of the framework. It is planned that the results of the data analysis will be used in scientific publications, along with illustrative, fully anonymised extracts from the data set. Parts of the data set will be made available via open access (details in section 5).

The ER4STEM project has received funding from the European Union’s Horizon 2020 research and innovation program under grant agreement No. 665972

3.2.2 CONFERENCES DATA SET

The conferences data set is a compact form of the workshops data set with fewer participants (in ECER2016 roughly 300 students are expected to be present). Collected data will be used to improve processes regarding the conferences as well as their evaluation. The data will also inform the development of the framework. It is planned that the data will be used in scientific publications and parts of it made available via open access (details in section 5).

- Conference Session Information (.doc)
  - Partner name
  - Dates (to-from)
  - Number of sessions
  - Location
  - Lead by
  - Other tutors/mentors
  - Age of students
  - Total number of students
  - Male/Female numbers
  - Group sizes
  - Total number of groups
  - How were the groups formed? Why?
  - Robotics kit
  - Programming languages
  - Domain
  - Aims of conference
  - Please include all relevant materials (e.g. activity plan, each session/lesson plan, handouts, etc.) in the folder with this document
- Questionnaires (anonymised and online or .xls)
  - Filled by the participants
  - Used to collect largely quantitative data
  - Questions are split into personal information, existing attitudes and conference experience.
  - Paper questionnaire answers to be entered into the online system
  - Free-text responses translated into English and entered into excel files
- Observations (translated into English, anonymised and digitalised .xls)
  - Observation protocol filled by the conference facilitators or other observers
  - Video observation where possible to verify and expand upon observation notes, as well as sensitising data analysts to the context.
- Interviews (anonymised, transcribed and translated into English, .xls)
  - With focus groups, audio-recorded
  - Conducted to understand the experience of participants and their reasons for particular actions
- Artefacts of learning (translated into English where applicable, anonymised and digitalised .pdf)
  - Created by the participants
  - Identified as group work
- Participant reflections (translated into English, anonymised if applicable and digitalised if applicable, .xls)
  - Created individually and as a team

The ER4STEM project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 665972

    - Either as a blog which acts as a reflection tool and a living artefact of the learning process, or as a guided reflection document
- Tutor reflections (translated into English, anonymised and digitalised, .xls)
    - Done by each of the tutors, mentors or workshop facilitators
    - The purpose is two-fold: 1) To document changes to workshop plans and the reasons for these; and 2) to document the evolution of activity plans between workshops.
- Sensitive data (audio and video recordings)
    - Audio recordings of interviews and any video recordings are encrypted and stored by the partner organisation and only encrypted video files are shared with evaluation partner Cardiff University for the purpose of analysis and archiving.

It is planned that the results of the data analysis will be used in scientific publications, along with illustrative, fully anonymised extracts from the data set. Parts of the data set will be made available via open access (details in section 5).

4 STANDARDS AND METADATA

The ER4STEM project follows the Ethical standards of the Cardiff School of Social Sciences and has ethical approval from the School of Social Sciences Research Ethics Committee. Besides following a rigorous evaluation protocol that includes informed consent of children participating in ER4STEM activities as well as their legal guardian, the project will comply with national and EU legislation on Data Protection, particularly the European Data Protection Legislation (Directive 95/45/EC). All ER4STEM project collected data will be anonymised before research analysis and any data that might make personal identification possible will be protected with adequate measures. For details, please see section 6.

The main purpose of the data collection is the evaluation of the impact of the framework tools and activities on young people. The findings will be made available via the project deliverables and scientific publications. Workshops are an important part of the data collection, therefore metadata needed for the workshops has been defined through the cooperation of work packages 2 and 6 (see section 4.1.1). This metadata or parts of it can be used as search parameters in an open access research repository that provides access to anonymised and processed research data (details of data sharing is in section 5). The metadata can also be made available in an appropriate form in the ER4STEM repository (work package 5). At this point, the metadata for these uses has not been defined but will be in subsequent versions of the DMP when relevant.

4.1.1 WORKSHOPS METADATA

The workshops metadata includes the following information for each session:

- Partner name
- Dates (to-from)
- Number of sessions
- Location
- Lead by
- Other tutors/mentors
- Age of students

The ER4STEM project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 665972

- Total number of students
- Male/Female numbers
- Group sizes
- Total number of groups
- How the groups were formed
- Robotics kit
- Programming languages
- Domain
- Aims of workshop

4.1.2 CONFERENCES METADATA

The conference metadata includes the following information:

- Partner name
- Dates (to-from)
- Number of sessions
- Location
- Lead by
- Other tutors/mentors
- Age of students
- Total number of students
- Male/Female numbers
- Group sizes
- Total number of groups
- How the groups were formed
- Robotics kit
- Programming languages
- Domain
- Aims of conference

4.1.3 OTHER METADATA

It is important to point out that research data collected from participants of workshops and conferences to evaluate the impact will not be the only data collected or generated during the project. Other data collected from or generated by the main stakeholders (teachers, ER activities organizers, and educational researchers), e.g. to understand requirements on the repository or framework, could be used for research purposes. Therefore, other metadata sets, their purpose and accessibility will need to be defined as the project progresses.

The ER4STEM project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 665972

5 DATA SHARING

At this point in time, all collected and anonymised data from the workshops and conferences as outlined in the sections before, will be disseminated in one form or another. So far, these datasets do not include any information that the consortium considers worth protection for exploitation. All collected data will be used for scientific evaluation and findings will be published via scientific channels. Open access to these publications will be made available depending on the form and cost of the open access. However, not all of the raw data can be made accessible to everyone for ethical reasons. Figure 2 outlines the decisions made by the consortium on how to handle data sharing at this point in time. In the following sections the decisions will be explained in further detail.

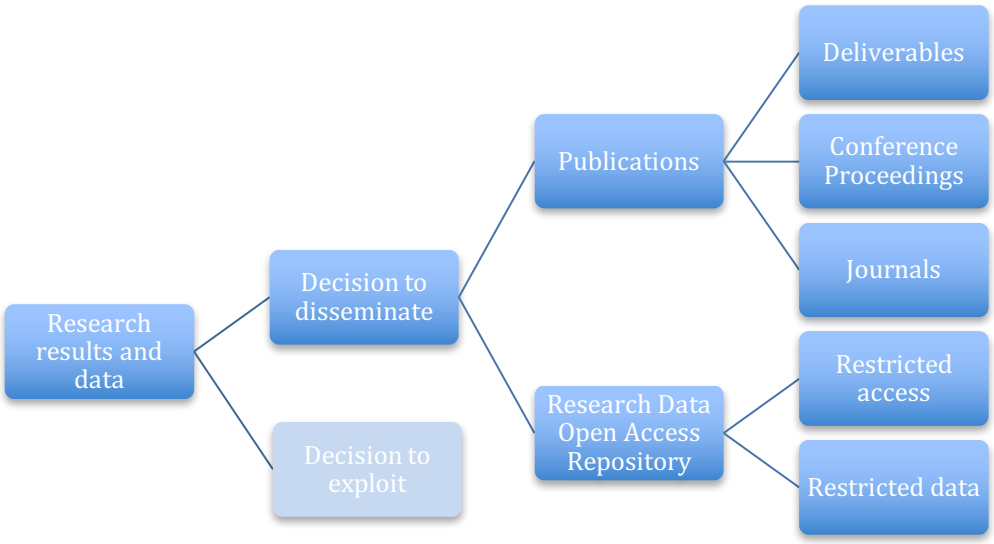

Figure 2: Data sharing plan

5.1.1 RESEARCH RESULTS AND DATA

As a European project under Horizon2020, the ER4STEM project consortium has declared its willingness to make all knowledge generated from the project publically available and provide open access to its scientific publications and research data [2]. Therefore, all deliverables of the project are open to public and accessible via the project's web page **er4stem.com**. As of this moment, research data is collected during workshops and conferences. The data sharing decisions have been taken regarding these data sets only and need to be revised when other data sets are added. It is expected that research data collected from the conferences will be treated similarly as the workshops data set.

5.1.2 DECISION TO EXPLOIT

The consortium has agreed that the workshops data set does not include any information that should be protected for exploitation reasons. However, the project aims to create an educational robotics repository which shall be sustainable after the project. Therefore, some data or knowledge generated or collected in the

The ER4STEM project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 665972

project might be identified as a unique selling point worthwhile of protection. The consortium needs to decide this on a case by case basis as the project progresses.

5.1.3 DECISION TO DISSEMINATE

The consortium has decided to disseminate findings from the research data in scientific publications. At this time any relevant data may be included, however the final decision can only be taken when all data is completely collected and processed. So, the research data dissemination will be linked to each scientific publication and will be decided on a case by case basis as the project progresses.

The consortium has also decided to use other non-scientific means to promote the project and its tools as well as the scientific findings. Scientix has already been proven to be a very competent collaboration partner in reaching one of the main stakeholders – STEM teachers – all over Europe.

The decision about dissemination channels is also affected by the cost factor (such as open access for scientific publications) and, although some budget is foreseen for open access publications, the consortium will prefer routes that minimize costs in order to make the research and knowledge generated by the project as diversely public as possible.

5.1.4 PUBLICATIONS

As of this moment, research data is being collected from the workshops and ECER 2016 conference as a pilot until month 10 (M10) of the project. The data will be processed, analysed and findings will be published in D6.3 in project month 12 (M12). As stated in 5.1.3 above, only once all data has been analysed will there be a decision on what will be disseminated via scientific journals or conferences.

5.1.4.1 DELIVERABLES

Findings from the first project year's research data will be published in deliverable D6.3 Evaluation and Analysis of First Project Year and publicly available via the project website. The content of the deliverable and its publishing date on the website depend on the consortium's strategy of the scientific publications and their submission dates. What part of D6.3 can be disseminated in which scientific publication will be decided in project month 12 (M12), MS4 milestone meeting in Malta.

5.1.4.2 CONFERENCE PRESENTATIONS AND PROCEEDINGS

Conference proceedings and books are very expensive for open access. Yet, there are also conferences who require only abstracts and dissemination is aural. The consortium will look for conferences that minimize the open access costs or make findings otherwise accessible. E.g., all presentation materials will be made available on the project website, as the conference will not hold the copyright for these.

The ER4STEM project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 665972

5.1.4.3 JOURNALS

Journal publications will be open access and publicly available, linked through via the project website. It is also a common route to publish first findings in conferences, and then enhance them together with further findings in a journal paper which then can be open access. It has been decided that the final scientific results that are going to be published in a journal paper can be gold access. Other journal papers will have green access for financial reasons (minimised cost, maximised dissemination).

5.1.5 RESEARCH DATA OPEN ACCESS REPOSITORY

The consortium is committed to make all research data that is going to be scientifically published accessible. However, the consortium has also decided that the data and the persons having access to the research data should be restricted. In the following subsections the restrictions and the rationale behind these are explained. As of this moment, the consortium is looking for research data open access repositories that fulfil these restrictions and will select one to be used. The research data will only be available in an open access repository after scientific publication, thus not before project month 12 (M12).

5.1.5.1 RESTRICTED DATA

The research data collected during workshops and conferences will be anonymised so that participants cannot be identified from the data. However, there is always the potential that individuals can be identified in audio, video and still images, even though they have been anonymised. Sharing this data with third parties would infringe data and child protection laws of the consortium countries. The consortium is not equipped with the competencies and time to take measures against this kind of identification, and even if it did so, it cannot guarantee that others could not apply countermeasures once in the possession of these materials. Thus the consortium has decided not to share with third parties audio, video or still images which include any participant.

Even within the consortium, only the evaluation lead partner and partners who originally collected the data will have access to this data which will be securely stored as described in section 6 of this document. The transcribed interviews and observation protocols may be made available via open access but only if they contain no identifying information.

The consortium has decided that all research data needs to be “cleaned” and processed, for example, school names or other identifying information needs to be removed, and brought into a form that is useful for other researchers to validate or replicate research results, and also fitting into the open access repository metadata and search options. There needs to be a compromise between the chosen repository and its data formats and the research data processed for the repository. The consortium has concluded that excel and word formats will be best formats in an open access repository, however, the decision can only be taken together with the selection of the right repository.

As part of the process of gaining informed consent to collect data from participants, they must be informed of the storage, protection and use of the data. If the data is made accessible to others, the consent is not fully

The ER4STEM project has received funding from the European Union’s Horizon 2020 research and innovation program under grant agreement No. 665972

informed, as the consortium has limited means to identify for what purposes the data will be used and by whom. Therefore, the open access pilot is explained to participants and they are given the opportunity to opt out of open access, thus data from participants who have opted out of open access pilot will not be made accessible. In addition, the consortium will limit the access to data to researchers who confirm their compliance with the same ethical obligations stated in the informed consent process.

5.1.5.2 RESTRICTED ACCESS

In order to ensure that the research data is used by third parties as explained to the participants with the informed consents, access to the data needs to be restricted. The consortium needs to know who is accessing the database and for what purpose. Criteria for access will include membership of a research institution based in Europe and they must submit a plan outlining how they will use the data (research questions and analysis approach) which will be reviewed before any decision to grant access.

The time frame of access is also part of the restriction. The research data cannot be made accessible before scientific publication. It therefore needs to be decided how long after publication it could be useful for other researchers to have access to the research data, this could range from one year post-publication to five years after the project.

6 ARCHIVING AND PRESERVATION

All research data will be stored until 2023. Partners also will need to archive personal data about the participants (the participant key) in a separate location, so that participants are able to use their right of withdrawing from the research anytime. The project partners will comply with national and EU legislation on Data Protection, particularly the European Data Protection Legislation (Directive 95/45/EC).

Each project partner will store the research data that is collected by that partner anonymously on a password protected server. Videos and audio files (and files where participant can be recognized) will only be stored in an encrypted drive and shared encrypted. Cardiff University will store all the data in the same way for all partners to ensure archiving in two separate locations. TU Wien will save Cardiff University research data for the same reason. The consortium has agreed to use the software VeraCrypt for encryption. For the needs of ER4STEM VeraCrypt provides sufficient security. It is open source, can be used on different systems (Windows, OS, Linux), it originates in Europe, is maintained by a French company, and is free of charge.

Software: https://veracrypt.codeplex.com/

Tutorial: https://veracrypt.codeplex.com/wikipage?title=Beginner%27s%20Tutorial

The website er4stem.com, which stores all publications of the project for public access, and the ER4STEM repository, which will store different tools and plans developed in the project, will be available at least five years after the project. Details about the research data open access repository can be found in section 5.1.5.

The ER4STEM project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 665972

7 CONCLUSION / OUTLOOK

The ER4STEM DMP Version 1.0 outlines how the research data collected during the project will be handled during and after the project as of March 2016. The document will be reviewed at each milestone meeting and adapted as the project progresses. Following topics that need to be discussed are identified:

- Other data sets: Definition of research data, structure, and metadata
- Open access repository: identify repositories that could fit the project consortium's access policies, choose one, determine costs of maintenance
- M12: Decide what to do with the workshops and conferences data set of the first year

Next update of the DMP is due in project month 17 (M17).

8 GLOSSARY / ABBREVIATIONS

EC	European Commission
ER4STEM	Educational Robotics for STEM
DMP	Data Management Plan
REA	Research Executive Agency
STEM	Science, Technology, Engineering, and Mathematics

9 BIBLIOGRAPHY

[1] European Commision. Guidelines on Data Management in Horizon 2020. Version 2.0. 30 October 2015

[2] European Commision. Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020. Version 2.0. 30 October 2015

The ER4STEM project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 665972