SCReLProg (Self and Co-regulation in e-Learning of Computer Programming)

DATA MANAGEMENT PLAN

Creators: Yulia Karimova, Daniela Pedrosa, Jose Cravino, Leonel Morgado

Affiliation: University of Porto (Universidade do Porto)

ORCID iD:
Yulia Karimova https://orcid.org/0000-0002-1015-6709,
Daniela Pedrosa https://orcid.org/0000-0001-9536-4234,
José Cravino https://orcid.org/0000-0002-5376-6128,
Leonel Morgado https://orcid.org/0000-0001-5517-644X.

Project abstract:
There is a shortage and a growing need for skilled professionals in computer science. Students show difficulties in learning computer programming in higher education, particularly in the transition from initial programming to advanced programming, so it's necessary to develop effective teaching strategies. The challenges are even greater in distance education, because it requires students to have greater discipline, autonomy, and self-regulation and co-regulation skills to successfully complete the course. We develop a pedagogical approach (SimProgramming) that showed promising results in the face-to-face context, helping students to overcome programming difficulties and contributing to develop effective strategies for self and co-regulation of learning. This project aims to adapt and develop this approach in the context of e-learning, to study how it influences the learning of programming and the strategies of self-regulation and co-regulation of learning used by the students.

ID: 36851

Grant number / URL: 30040

Version 1: May 2019

Version 2: March 2020

Version 3: 17 November 2021

1. DATA COLLECTION

What data will you collect or create?

The SCReLProg Project is related to the shortage of and growing need for skilled professionals in software development, seeking to develop effective teaching strategies in support of the learning of computer programming in higher education, particularly in its distance education format, because it requires students to have greater discipline, autonomy, and self-regulation and co-regulation skills to successfully complete the courses. During the project, the pedagogical approach SimProgramming [**refs.: 1, 2, 3, 4, 5, 6**] that was developed to help students overcome programming learning difficulties, will be adapted for the context of e-learning, to study how it influences the learning of programming and the strategies of self-regulation and co-regulation of learning used by the students. Moreover, the types of self-regulation and co-regulation strategies adopted by students will be identified, a new learning model for online teaching proposed, for the specific context of programming education, and students' self-regulation and co-regulation of learning skills will be developed.

Various types of data will be collected throughout the project, namely: paginated documents with analysis grids; answers to questionnaires; reflections, written by students; recordings of interviews (audio and video); activity reports with timestamps (obtained through the Moodle platform) such as: items visualized, materials downloaded, posts published, etc.

References:

1. Pedrosa, D. (2017). Autorregulação e corregulação das aprendizagens no ensino superior: estratégias adotadas por alunos de programação de computadores. PhD Thesis. Universidade de Trás-os-Montes e Alto Douro, Vila Real. Pp. 317. URL: http://hdl.handle.net/10348/7651

2. Pedrosa, D., Cravino, J., Morgado, L., & Barreira, C. (2017). Self-regulated learning in higher education: strategies adopted by computer programming students when supported by the SimProgramming approach. *Production*, *27* (SPE). e20162255. DOI: http://dx.doi.org/10.1590/0103-6513.225516

3.   Pedrosa D., Cravino J., Morgado L., Barreira C. (2016) Self-regulated Learning in Computer Programming: Strategies Students Adopted During an Assignment. In: Allison C., Morgado L., Pirker J., Beck D., Richter J., Gütl C. (eds) *Immersive Learning Research Network*. iLRN 2016. Communications in Computer and Information Science, vol 621. pp. 87-101. Springer, Cham. DOI: https://doi.org/10.1007/978-3-319-41769-1_7

4.   Pedrosa, D., Cravino, J., Morgado, L., & Barreira, C. (2016). Self-regulated learning in higher education: strategies adopted by computer programming students. In *Proceedings of the PAEE/ALE'2016, 8th International Symposium on Project Approaches in Engineering Education (PAEE) and 14th Active Learning in Engineering Education Workshop(ALE).* pp. 588-595. PAEE–Project Approaches in Engineering Education Association/Universidade do Minho. ISBN: 978□989□20□6829□9. Guimarães, Portugal. URL: http://hdl.handle.net/10400.2/5453

5. Pedrosa, D., Cravino, J., Morgado, L., Barreira, C., Nunes, R. R., Martins, P., & Paredes, H. (2016). Simprogramming: the development of an integrated teaching approach for computer programming in higher education. In *Proceedings 10th annual International Technology, Education and Development Conference (INTED 2016).* pp. 7162-7172. IATED Academy. Valencia, Spain. doi: 10.21125/inted.2016.0699. URL: http://hdl.handle.net/10400.2/5162

6. Nunes, R. R., Pedrosa, D., Morgado, L., Martins, P., Paredes, H., Cravino, J., & Barreira, C. (2017, October). SimProgramming: uma abordagem motivacional para a aprendizagem de alunos intermediários de programação. In *Anais dos Workshops do Congresso Brasileiro de Informática na Educação*, 6 (1), pp. 1099-1110. DOI: http://dx.doi.org/10.5753/cbie.wcbie.2017.1099

How will the data be collected or created?

The data will be collected through LMS Moodle, semi-structured interviews, ERICA questionnaire (Lime Survey), biweekly reflections.

a) LMS Moodle: E-learning platform of Universidade Aberta:

1. Interaction of the participants in the forums for clarification of the doubts and forums of the teams. After data collection, the data will be processed, anonymized, and codified. It is a coding of names. We need this coding to later understand the evolution over time (changes in self-regulation and co-regulation strategies.

2. Activity reports and information related to the platform (platform-access reports, material consultation reports, posting, statistics, etc.). In this case, the information includes the names of the participants and sometimes the photo, so this data also will be processed, anonymized, and codified. In case the photos and videos, we let's hide the faces. In other words, to cover them

b) Semi-structured interviews:
During the interviews, we will be using Colibri/Zoom software or another communication tool, either for video recording or just for audio recording. The interviews will be held with all students of teams that agreed to participate in the research at the end of the course. However, throughout the course biweekly sessions of coaching will be carried out (15-minute meetings) with leaders of each team and biweekly sessions of coaching with all members of team. The purpose of these interviews is to understand the evolution of learning. Even if there was no detailed presentation of the participant during the interviews, the following personal information may be collected: names, surnames, identification of colleagues, name of the course, teachers, notes, etc. After the interview is done, a transcript will be made, where all information related to personal data will be codified and anonymized. We need this information because we need to know which geographical area they are (EU; USA; etc) and the Name to be replaced later with a code to then compare.
Also, final interviews may be performed with some students participating in the research, about their strategies for self-regulation of learning.
c) ERICA questionnaire: The questionnaires will be carried out in diverse courses through the Lime Survey and will collect the following information: name, surname (not mandatory), individuals' gender, age, residence: city and country, and academic habilitations. The questionnaire results will be processed and all information related to personal data will be codified and anonymized. Moreover, the data will not be interconnected.
d) Biweekly reflections: The reflections are made using Lime Survey, related to the course "Software Development Laboratory" and contain three questions about the accomplishment of tasks by students. From a students' point of view, these reflections help understand what can be improved and where. From a project point of view, it helps track students' progress and learning, see what difficulties they face, what motivates them the most. In general, reflections will collect the following information: E-mail; Name; Name of Team and chosen path. All the information related to personal data will be codified and anonymized.
The data will be treated using the following Software:

- For content analysis: Nvivo; or MatLab; or Word;

- For ERICA questionnaire (pre and post-test): SPSS or Excel, Lime Survey.

And using both methodologies: quantitative and qualitative.
Moreover, the students will read the abstract of the project, understand the goals of the research and sign informed consent statements agree to participate in the research. The informed consent will contain checkboxes, which must be signalled by students, in the case of understanding of all project, collecting and processing purposes before the start of any activity. All personal information that students provide will be collected during interviews, and during their interaction with the LMS Moodle, and by filing and answering questionnaire and writing reflections. After collection, all the information will be processed, during which personal information will be anonymized and codified. The RAW data will be destructed after processing the data or preserved according to the storage rules indicated in the "Storage and Backup" section. In general, the RAW data will be destructed 3 years after project completion.

According to DMP monitoring on 10.03.2020 new forms of data collection, namely "Metacognitive tasks in moodle test format" are going to be added, which in turn do not bring major changes to the plan, continue to be in accordance with what was described in the 2d version of DMP, with the same conditions and rules already established.
Also, the amount of data collected will be increased, but also following the rules established in the 2d version of DMP.
The Biweekly reflections also were changed to Moodle.
The second research cycle has started, so some changes may appear that will be reviewed in the third monitoring of the DMP, which is scheduled for the end of July.

2. DOCUMENTATION AND METADATA

What documentation and metadata will accompany the data?

- Metadata will be created using Dublin Core, and specific descriptors created for specific needs during the description process.
- Documentation will include: message exchanges, such as entities emails, Gmail emails, Skype and paper documents such as notes about the teams (name of participants and names of the teams), some notes in the meetings or print documents of the interviews to analyze (grids, etc).
- Moreover, there are PIA (Privacy Impact Assessment), Assigned the informed consent, and other documents relevant to the project.

3. ETHICS AND LEGAL COMPLIANCE

How will you manage any ethical issues?

The SCReLProg Project is financed by Norte 2020 (Nº 30040), and responds to all existing requirements related to the Research Data Management and Protection of Personal Data.
UTAD follows the Regulation of the UTAD Ethics Committee ("art.3º - Competências") which define a set of the rules to guide projects, including this one, which must be read and applied by all project collaborators.
The organization process of the research study SCReLPRog includes the following:

1. During the initial 3 weeks of a given course (curricular unit), the students enrolled in the curricular unit must choose the type of assessment: (a) by continuous evaluation or b) by examination. This procedure is described in the instructions of the Technological Support Services to Teaching - Universidade Aberta.

2. For continuous evaluation there are 3 options: a) Individual path with agreement to participate in the research study and with the research use of the data, which is provided on a voluntary basis; b) Team path, with agreement to participate in the research study and with the research use of the data, which is provided on a voluntary basis (currently we have 2 teams following this path); c) Individual path, which does not accept to participate in this study nor in the research use of data about the individuals' activity.

3. For evaluation by examination, there is no option for participation in the research study. They are not included in the study.

In case of participation in the research study, the students receive the documents with the detailed information about the project, such as the abstract, its goals, participating partners, and previous published research works. Moreover, they sign acknowledgements of understanding the finality of the treatment of their data and provide informed consent for the use of their data.
This DMP has been verified by the INESC TEC DPO and is in compliance with the General Data Protection Regulation.
In addition, confidentiality agreement protocols were created for researchers, a contract draft of a subcontracting protocol related to GDPR for an external researcher (from the University of Aveiro) and a protocol between institutions.
Both the DPIA and the interinstitutional protocol were verified and validated by the University of Aveiro and the Universidade Aberta, and are being signed by all institutions. In case of doubts, the person responsible for collection, processing and preservation will clarify all those and answer all questions related to the research project.

How will you manage copyright and Intellectual Property Rights (IPR) issues?

The RAW data is not open and will be protected according to this DMP.
The processed data will be deposited and open without any restriction with a defined license.

4. STORAGE AND BACKUP

How will the data be stored and backed up during the research?

The RAW data will be preserved on the working laptop or desktop computer of the collector and will be shared with others responsible through the https://drive.inesctec.pt. The questions related to the sharing of RAW data with partners is not defined yet. The RAW data will not have backups.
 The backups the data processed will be created weekly. The processed data will be preserved during 5 years on the working laptop or external disks, anonymized and codificaded.

How will you manage access and security?

The users don't have access to RAW data that is why they don't need logical access control. However, they have access to the Moodle Platform, with authentication by user and password created during the inscription on the CU (formal request with approval from the responsible entity Universidade Aberta).
The responsibles Professor José Paulo Cravino, Professor Leonel Morgado and Daniela Pedrosa also have access to the platform with authentication by user and password. The password is difficult, will be changed 3 in 3 months and may be blocked if repeatedly enter incorrectly.
Moreover, the all responsible of the research project have access to the https://drive.inesctec.pt, OSF platform, also only with authorization.
During the videoconference sessions only participants have the link to it access.
The paper documents will be preserved in the offices of the researchers responsible for the project with a padlock with access limited by supervisions and responsible by preservation Daniela Pedrosa. They will be destroyed in five years. Also, can be destroyed earlier, if they are no longer needed for the research.
Also, during the project will be used several operating systems, such as Windows 10 Pro and applications, such as Excel, Word, Adobe Reader. All software is licensed that reduces the risk of intruders. Moreover, all the computers have properly updated antivirus software.
The browsers have Adblock installed that permit block the pop-up pages and other insecure connections. All technical issues related to the software will be controlled by each member of the project and in case of necessary support, they contact the IT staff of the responsible entity.

5. SELECTION AND PRESERVATION

Which data are of long-term value and should be retained, shared, and/or preserved?

The RAW data will be preserved the shortest time possible (the maximum three year after the project conclusion), limited by the needs to process the data. After data processing, the RAW data will be erase or stored on the personal external disks, encrypted with a password, and with limitation of the access for this RAW data. This data can be shared with others responsible by request. The responsible for the preservation of the RAW data will be Daniela Pedrosa.
The processed data will be preserved on the working laptop of the collector and research data repository of the INESC TEC (https://rdm.inesctec.pt/) with the license for reuse Share-Alike. Moreover, the processed anonymized data will be open and share with project partners and other interested researchers. Data that cannot be anonymized will be deleted (recording videos/interviews, photos etc. all that can identify a person) right after processing.
Other information related to the project, such as bibliographic references will be preserved and shared with others, using the site of the project OSF (osf.io/ugehx). The information can be either open to all or closed and not yet defined.

What is the long-term preservation plan for the dataset?

All processed data should be preserved in the long-term. The supervisors Prof. José Paulo Cravino and Prof. Leonel Morgado, and Daniela Pedrosa will be responsible for any action related to the long-time preservation.

6. DATA SHARING

How will you share the data?

The processed data and their metadata will be open on the INESC TEC research data repository (https://rdm.inesctec.pt/).

Are any restrictions on data sharing required?

The processed data have no restrictions on data sharing.

7.   RESPONSIBILITIES AND RESOURCES

Who will be responsible for data management?

Responsible for the collection, processing and preservation of the data are:

- **Daniela Pedrosa** (PhD in Science and Technology Didactics) is a Post-Doc Researcher of the Research Centre on Didactics and Technology in the Education of Trainers (CIDTFF) and Universidade de Trás-os-Montes e Alto Douro (UTAD), in Portugal. ORCID: https://orcid.org/0000-0001-9536-4234

- **Post-Doc Researcher** at UTAD, to be hired in June.

- **MEng: Technology/Software** developer at INESC TEC.

Supervisors:

- **Prof. José Paulo Cravino** is Assistant Professor of Physics at the School of Science and Technology of the Universidade de Trás-os-Montes e Alto Douro. Director of the Physics Department. Researcher at the Research Centre "Didactics and Technology in Education of Trainers" (CIDTFF) at the Universidade de Aveiro. ORCID: http://orcid.org/0000-0002-5376-6128

- **Prof. Leonel Morgado** is Assistant Professor with Habilitation, at Universidade Aberta (Portuguese Open University), and a Senior Researcher at INESC TEC, ORCID: http://orcid.org/0000-0001-5517-644X

Consultants:

- **Prof. Jonathan Kaplan** is Associate Professor at the Institute of Sciences and Practices of Education and Training (ISPEF) at the University of Lyon; Website: http://recherche.univ-lyon2.fr/ecp/equipes/enseignants-chercheurs/jonathan-kaplan

- **Prof. Pedro Rosário** is Associate Professor in the School of Psychology at the Universidade do Minho. Dr. Rosário did his PhD in Educational Psychology at the Universidade do Minho. ORCID: https://orcid.org/0000-0002-3221-1916

Others Researchers of the Project:

- **Prof. Teresa Bettencourt** is an Assistant Professor at the Department of Education and Psychology of the Universidade de Aveiro and a Researcher at CIDTFF. ORCID: https://orcid.org/0000-0002-2913-952X

- **Prof. José Bidarra** is Assistant Professor with Habilitation at the Department of Science and Technology of Universidade Aberta. ORCID: https://orcid.org/0000-0002-2082-5996

- **Prof. António Coelho** is Assistant Professor with Habilitation at the Department of Informatics Engineering of the Faculty of Engineering (FEUP), Universidade do Porto (UP) and a Researcher at INESC TEC. ORCID: https://orcid.org/0000-0001-7949-2877

- **Prof. Elizabeth Carvalho** is Assistant Professor at Universidade Aberta, with the Department of Sciences and Technology. https://www2.uab.pt/departamentos/DCT/detaildocente.php?doc=129

- **Prof. Vitor Cardoso** is Assistant Professor at Universidade Aberta, with the Department of Sciences and Technology. https://www2.uab.pt/departamentos/DCT/detaildocente.php?doc=69

Subsequent hires during the project will be identified in the second version of this document.
Five responsible entities:

- **Principal**: Universidade de Trás-os-Montes e Alto Douro;

- **Partners**: Universidade Aberta; Universidade de Aveiro; Research Centre on Didactics and Technology in the Education of Trainers (CIDTFF); Institute for Systems and Computer Engineering, Technology and Science (INESC TEC).

According to DMP monitoring on 10.03.2020 there was the integration of collaborators in the research team, namely:
- a Post-Doc Brazilian researcher Mário Fontes;
- two students from the Universidade de Aveiro within the scope of PIC -Edu (scientific initiation) - Maria Castelhano and Eliana Curado.

Moreover, The UC professor is on the sabbatical license, so a new professor was hired to replace him in this academic year (2019/2010). The new professor Pedro Pestana aware of the research project, its rules and conditions, and agreed with its realization.

According to DMP monitoring on 26.10.2021, there was changing of collaborators in the research team, namely:
- Tania Araujo left the project;
- Eliana Curado left the project;
- Ceres Morais - was research fellow (project fellowship) and is currently PhD student and voluntary researcher;
- Rita Mota - Master's student and voluntary researcher of the project;
- Cristiane Bonfim - PhD student and voluntary researcher;
- Maria Castelhano - Master's student and voluntary researcher;
- Rodrigo Faria: PhD student and voluntary researcher;
- Fernando Cassola – Voluntary Researcher joined to the project.

What resources will you require to deliver your plan?

It is necessary to access to the INESC TEC research data repository (https://rdm.inesctec.pt/) and site of the project OSF (osf.io/ugehx)
Moreover, during the project the following assets will be used:

7. Hardware/devices: work desktop and laptop computers, personal laptop computers, USB flash drives, external disks, institutional servers, cameras, microphones, loudspeakers, smartphones, tablets.

8. Software: Windows, Linux, SPSS, Excel, Word.

9. Networks: entities' local networks and Wi-Fi networks with access to the Internet.

10. Cloud services: https://drive.inesctec.pt.

11. Sites/repositories: OSF (osf.io/ugehx), research data repository INESC TEC (https://rdm.inesctec.pt/) for only processed and anonymised data.

12. Message exchanges: entities' email services, Google "GMail" emails, Skype.

13. Paper transmission channels: notes related to teams, meetings, participants names, etc.